



**American Society of Human Genetics  
69<sup>th</sup> Annual Meeting Poster Abstracts**

Abstract Topic	Abstract/Poster Board Numbers	
	Start #	End #
Precision Medicine, Pharmacogenomics, and Genetic Therapies	372	505
Prenatal, Perinatal, Reproductive, and Developmental Genetics	506	620
Genetic Counseling, ELSI, Education, and Health Services Research	621	750
Cancer Genetics	751	1082
Mendelian Phenotypes	1083	1398
Bioinformatics and Computational Approaches	1399	1746
Molecular Phenotyping and Omics Technologies	1747	1886
Complex Traits and Polygenic Disorders	1887	2297
Evolution and Population Genetics	2298	2410
Molecular and Cytogenetic Diagnostics	2411	2617
Cardiovascular Phenotypes	2618	2793
Statistical Genetics and Genetic Epidemiology	2794	3063
Molecular Effects of Genetic Variation	3064	3212
Epigenetics and Gene Regulation	3213	3364

**Posters should remain on the board for all three days (Wednesday through Friday)**

The program and abstract/poster board number next to each listing is followed by an **W** (Wednesday), **T** (Thursday), or **E** (Friday) to indicate the day on which authors must be present at their poster boards.

## PgmNr 372: A mechanism and treatment strategy for the sexual dimorphism seen in vascular Ehlers-Danlos syndrome.

### Authors:

C.J. Bowen <sup>1,2</sup>; Z. Burger <sup>1,2</sup>; G. Rykiel <sup>1,2</sup>; J.F. Calderon Giadrosic <sup>1,2</sup>; N.K. Wilson <sup>1,2</sup>; H.C. Dietz <sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Genetic Medicine, Johns Hopkins Medicine, Baltimore, MD.; 2) Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Division of Pediatric Cardiology, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

---

We demonstrated that the PLC/IP<sub>3</sub>/PKC/ERK pathway is activated in vulnerable vascular segments in a mouse model (*Col3a1*<sup>G939D/+</sup>) of vascular Ehlers-Danlos syndrome (vEDS), a connective tissue disorder associated with spontaneous arterial rupture. Early postnatal attenuation of this signaling axis using either PKCβ (ruboxistaurin) or MEK/ERK (cobimetinib) inhibitors affords overt life-long protection in vEDS mice. Curiously, while potent inhibition of IP<sub>3</sub>-mediated PKCβ activation with hydralazine also prevented vascular rupture in early life (98% vs 50% survival at 45 days), protection was abruptly lost at sexual maturity, an event that was exaggerated in male mice (72% fatal aortic rupture between postnatal day 50 and 75 vs 15% in females). This is reminiscent of the elevated risk of vascular rupture in young men with vEDS at the end of puberty. We hypothesized that androgen signaling may contribute to puberty-associated aortic rupture. To test this, we treated vEDS mice with bicalutamide, an androgen receptor antagonist (ARA), and hydralazine. We found near-complete rescue from aortic rupture in both male and female mice (90% and 100% long-term survival, respectively). Notably, ARAs alone do not afford the same degree of protection as seen when used with hydralazine, suggesting therapeutic synergy. In anticipation of a reluctance to use potent ARAs in young men with vEDS, we considered the potential use of spironolactone, a diuretic that reduces androgen production, is a competitive antagonist of the androgen receptor and is used clinically for a variety of indications in adolescents of both sexes. Initiation of treatment with hydralazine and spironolactone at 21 days of age in vEDS mice of both sexes achieved 100% long-term survival, as compared to 45% in untreated mice. Risk correlated directly with ERK1/2 phosphorylation in the aortic wall. The vulnerability at puberty seen with hydralazine, but not ruboxistaurin or cobimetinib, positions androgen cross-talk downstream of IP<sub>3</sub> but proximal to PKCβ. Our results provide the first evidence that targetable signaling events contribute to the pathogenesis of vEDS and that puberty-associated vascular risk is androgen dependent, highlighting unanticipated therapeutic opportunities. This study may inform the mechanism for a more generalized increase in disease incidence and/or severity in males with other aortopathies including Marfan syndrome, bicuspid aortic valve with aneurysm and abdominal aortic aneurysm.

## **PgmNr 373: Elucidation of mechanism for accentuation of aortic aneurysm predisposition in males reveals a novel therapeutic strategy for heritable aortopathies.**

### **Authors:**

N.K. Wilson <sup>1</sup>; J.J. Doyle <sup>1</sup>; E. Gallo MacFarlane <sup>1</sup>; R. Bagirzadeh <sup>1</sup>; D. Bedja <sup>2</sup>; C.J. Bowen <sup>1</sup>; K.A. Toomer <sup>1</sup>; T.J. Creamer <sup>1</sup>; S.K. Cooke <sup>1</sup>; H.C. Dietz <sup>1</sup>; MIBAVA Leducq Consortium

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute of Genetic Medicine, Johns Hopkins Medical Institutions, Baltimore, Maryland.; 2) Division of Cardiology, Johns Hopkins Medical Institutions, Baltimore, Maryland.

---

Bicuspid aortic valve (BAV) with distal ascending aortic aneurysm (DAscAoA) is the most common inherited aneurysm condition. Despite clear autosomal dominant inheritance with incomplete penetrance, very little is known regarding the etiology or pathogenesis of BAV/DAscAoA. Obstacles to progress include extreme locus heterogeneity, with currently defined genes accounting for less than 5% of disease, and the lack of robust animal models. There is a clear male bias with regard to disease incidence, penetrance, and severity that has not been mechanistically characterized. Unlike most heritable aortopathies that show restricted dilatation of the aortic root such as Marfan syndrome (MFS), Loeys-Dietz syndrome (LDS), and isolated familial thoracic aortic aneurysm (iFTAA), the aortopathy associated with BAV typically shows predominant enlargement of the ascending aortic segment distal to the sinotubular junction. We have now shown that this segmental predisposition is recapitulated in MFS mice treated with calcium channel blockers (CCBs). ERK activation and angiotensin II type 1-receptor signaling are drivers of aortic disease in this model, and Notch signaling is overtly protective as Notch antagonists dramatically accentuate the DAscAoA phenotype. Importantly, this model shows the male gender bias with regard to incidence and severity of aneurysm observed in people with BAV/DAscAoA, providing a unique opportunity to elucidate the mechanism underlying this sexual dimorphism. We performed RNA-sequencing on DAscAo specimens from male MFS mice treated with CCBs with or without rescue with ERK inhibition (ERKi). Regulation of androgen receptor activity emerged as a top enriched pathway in the filtered dataset of expression changes induced by CCBs but abrogated by addition of ERKi. We hypothesized that androgen receptor signaling may be responsible for the increased severity of the aortic phenotype in males. To test this, we treated male MFS mice receiving CCBs and Notch inhibitor with the androgen receptor antagonist Enzalutamide. Informatively, this isolated intervention completely abrogated DAscAoA formation in male MFS mice treated with CCBs and Notch inhibitors; protection was also evident in corresponding female mice, presumably due to the inhibition of female androgens. The natural protection afforded in preadolescent or female MFS mice suggests that androgen antagonism may emerge as a viable therapeutic strategy in a broader clinical context.

## PgmNr 374: Multimodal analysis in Myhre syndrome unravels potential endpoints for monitoring disease severity and efficacy of losartan treatment.

### Authors:

G. Cappuccio <sup>1,2</sup>; M. Caiazza <sup>3</sup>; N. Bertolini <sup>4</sup>; A. Roca <sup>5</sup>; D. Apuzzo <sup>2</sup>; L. Pierri <sup>2</sup>; D. Melis <sup>2</sup>; A. Colucci <sup>6</sup>; G. Matyas <sup>7</sup>; C. Heneggeler <sup>7</sup>; A.M. Spinelli <sup>8</sup>; I. Bruno <sup>9</sup>; V. Bruner <sup>4</sup>; V. Bascherini <sup>4</sup>; M. Rubino <sup>3</sup>; J. Troisi <sup>10</sup>; G. Limongelli <sup>3</sup>; N. Brunetti-Pierri <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Telethon Institute of Genetics and Medicine, Pozzuoli, Naples, Italy; 2) Department of Translational Medicine, Section of Pediatrics, Federico II University, Naples, Italy.; 3) Paediatric Cardiology Unit, AORN Colli, Department of Cardiothoracic Sciences, Luigi Vanvitelli University, Naples, Italy.; 4) Rheumatology Unit Federico II University, Naples, Italy.; 5) Department of Neurological, Reproductive and Odontostomatological Sciences, Federico II University, Naples, Italy; 6) Department of Medicine, Surgery and Dentistry "Scuola Medica Salernitana" Pediatrics Section, University of Salerno, Baronissi (SA), Italy; 7) Center for Cardiovascular Genetics and Gene Diagnostics, Foundation for People with Rare Diseases, Schlieren-Zurich, Switzerland; 8) Private practice, Bari, Italy; 9) Department of Paediatrics, IRCCS Burlo Garofolo, Trieste, Italy; 10) Theoreo srl, Montecorvino Pugliano (SA), Italy

---

**Introduction:** Myhre syndrome (MS) is an ultra-rare autosomal dominant disorder due to recurrent mutations in the *SMAD4* gene encoding a member of TGF $\beta$ -SMAD4 pathway regulating extra-cellular matrix (ECM) homeostasis. The core clinical features of MS are related to progressive fibrosis and include thickening of skin and joint stiffness. Studies in fibroblasts have previously shown that losartan, an antihypertensive drug and TGF- $\beta$  antagonist, corrects the ECM deposition defect of MS.

**Materials and methods:** Five MS subjects (mean age: 22.4 years) with *SMAD4* mutations underwent clinical and laboratory evaluations including: 1. skin thickness by Rodnan score; 2. joint stiffness by goniometry; 3. speckle-tracking echocardiogram for myocardial fibrosis; 4. untargeted metabolomic GC/MS analysis in blood and urine to search for disease biomarkers.

**Results:** MS subjects were found to have increased skin thickness with an average Rodnan score of  $18 \pm 5.6$  [ranging from 0 (no thickening) to 51 (severe thickening)] and reduced joint range of motion, particularly at the ankle and knee by goniometry. Speckle-tracking echocardiogram revealed an increased strain (average  $12.5 \pm 3.2$  with normal values of  $19.7 \pm 0.28$ ) suggesting cardiac fibrosis that improved in one MS subject treated with losartan. The metabolomes of MS subjects were well separated from controls and butyric acid, ribitol and ribitol-related metabolites were found to be significantly higher in MS compared to controls.

**Conclusions:** MS patients showed multisystem involvement with fibrosis of skin, joints, and heart. A distinctive metabolomic signature involving fatty acids and polyols was detected. The present study unravels endpoints to monitor disease progression and for the evaluation of therapeutic efficacy. Preliminary data suggest that losartan can improve cardiac fibrosis in MS.



## PgmNr 375: Nanogel-mediated protein replacement therapy for Tgase-1-deficient congenital ichthyosis: Results from the proof of concept study.

### Authors:

K.M. Eckl<sup>1,2,3</sup>; R. Plank<sup>1,2,4</sup>; G. Yealland<sup>5</sup>; E. Miceli<sup>6</sup>; P. Graff<sup>1</sup>; M. Calderón<sup>6</sup>; S. Hedtrich<sup>5</sup>; H.C. Hennies<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biological and Geographical Sciences, University of Huddersfield, UK; 2) Division of Human Genetics, Medical University of Innsbruck, Austria; 3) Dept for Biology, Edge Hill University, Ormskirk, United Kingdom; 4) Cologne Center for Genomics, University Hospital Cologne, Germany; 5) Institute of Pharmaceutical Sciences, Freie Universität Berlin, Germany; 6) Institute of Chemistry & Biochemistry Freie Universität Berlin, Germany

---

Autosomal recessive congenital ichthyosis (ARCI) is a rare and severe disorder of disturbed cornification. Patients present with a more or less intense scaling on the entire body, often accompanied by secondary features including hair loss, ectropion, eclabion, and joint stiffness, especially in the very young. ARCI is both genetically and clinically heterogeneous. There are presently no curative therapies available, current treatment options are restricted to frequent and extended bathing, application of crèmes and scrubbing off of excess scales.

Approximately one third of ARCI cases are caused by mutations in *TGM1* making it the most common known cause. *TGM1* encodes transglutaminase 1 which is required to cross-link proteins in the outermost layer of the skin (horny layer). The disease disrupts normal keratinisation resulting in generalised scaling of the skin with notable detriments to barrier function, the latter of which leads to increased trans-epidermal water loss and a greater susceptibility to infections. In the present study, thermoresponsive nanogels (tNG) loaded with functional, recombinant transglutaminase 1 were applied to the surface of full thickness skin equivalents generated with the keratinocytes of transglutaminase 1 deficient ARCI patients. The skin equivalents demonstrated clear deficiencies in transglutaminase 1 activity and skin barrier function, in line with the pathophysiology of ARCI. Topical application of the loaded tNGs instated transglutaminase 1 activity within the viable epidermis of the mutant skin equivalents and improved skin barrier function in a manner dependent on transglutaminase 1 dose; application of the tNG alone or recombinant transglutaminase 1 alone had no impact. Toxicity studies in skin equivalents as well as keratinocyte and fibroblast monolayers confirmed the high biocompatibility of the treatment. Topical protein replacement therapy for transglutaminase 1 deficient ARCI patients appears highly promising and demonstrates a novel therapeutic approach for monogenic skin diseases.

## PgmNr 376: Protein O-GlcNAcylation is a novel mechanism regulating ureagenesis and ammonia detoxification.

### Authors:

L. Soria <sup>1</sup>; A. De Angelis <sup>1</sup>; D. Paris <sup>2</sup>; C. Braccia <sup>3</sup>; S. Attanasio <sup>1</sup>; E. Nusco <sup>1</sup>; A. Armirotti <sup>3</sup>; A. Motta <sup>2</sup>; N. Brunetti-Pierri <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy; 2) Institute of Biomolecular Chemistry, National Research Council, Pozzuoli, Italy; 3) Italian Institute of Technology, Genova, Italy; 4) Department of Translational Medicine, Federico II University, Naples, Italy

Life-threatening hyperammonemia results from inherited defects of the urea-cycle and acquired liver diseases. O-GlcNAcylation is a dynamic post-translational modification of intracellular proteins catalyzed by the enzyme O-GlcNAc transferase (OGT) whereas the O-GlcNAcase (OGA) removes the O-GlcNAc. O-GlcNAcylation regulates liver metabolism of glucose, lipids, and bile acids. Here, we investigated the role of hepatic O-GlcNAcylation in ammonia detoxification.

Pharmacological and genetic inhibition of OGA was investigated in C57BL/6 wild-type mice with acute hyperammonemia induced by intraperitoneal injection of <sup>15</sup>N-labeled ammonium chloride. Ammonia removal was monitored by measurements of serum ammonia whereas ureagenesis was evaluated by <sup>15</sup>N-NMR spectroscopy. O-GlcNAcylated protein levels were determined by Western blot and hepatic gene expression by qPCR. Whole-liver metabolome and proteome analyses were performed by <sup>1</sup>H-NMR and mass spectrometry, respectively.

Treatment with OGA inhibitor molecules (PUGNAc and Thiamet-G) increased O-GlcNAcylation of liver proteins. Both drugs reduced concentrations of serum ammonia and increased serum levels of <sup>15</sup>N-labeled urea. Similar results were observed in mice with siRNA-induced hepatic *knockdown* of OGA. Interestingly, whole-liver metabolome was well separated between mice with enhanced O-GlcNAcylation and controls, and metabolite set enrichment analysis unravelled urea-cycle among the main implicated pathways. Neither expression of genes encoding urea-cycle enzymes nor genes involved in ammonia detoxification were affected by OGA inhibition. The protein amount of these enzymes was also unaffected, suggesting that hepatic O-GlcNAcylation regulates ammonia clearance capacity by controlling the enzymatic activity of the urea-cycle.

In conclusion, hepatic O-GlcNAcylation plays an important role in ammonia detoxification. The mechanisms underlying increased ureagenesis by OGA inhibition is under investigation. Nevertheless, OGA is a novel therapeutic target for treatment of hyperammonemia of both urea cycle disorders and acquired hyperammonemia.

## **PgmNr 377: Inhibition of Grb14, a negative modulator of insulin signaling, ameliorates insulin resistance.**

**Authors:**

Ding; R. Rugmani; Y. Zhou; C. Novotny; D. Metzger; J. Vogel; G. Swaminath; S. Talukdar

[View Session](#) [Add to Schedule](#)

**Affiliation:** Merck Research Laboratories, South San Francisco, CA.

---

Insulin resistance increases a patient's risk of developing type 2 diabetes, non-alcoholic steatohepatitis (NASH) and a host of other comorbidities including cardiovascular disease and cancer. At the molecular level, insulin exerts its function through the insulin receptor (IR), a transmembrane receptor tyrosine kinase. Data from human genetic studies have shown that Grb14 functions as a negative modulator of IR activity, and germline Grb14-knockout mice have improved insulin signaling in liver and muscle tissues. Here, we confirm the insulin dependent interaction between IR and Grb14 and show that Grb14 knockdown in the liver and the heart with an AAV-shRNA (Grb14-shRNA) improved glucose homeostasis and insulin resistance in diet-induced obese (DIO) mice. A previous report has shown that germline deletion of Grb14 in mice resulted in cardiac hypertrophy and decreased systolic function; effects that could severely limit the therapeutic potential of targeting Grb14. In order to understand the direct role of Grb14 in the heart, we investigated the cardiac function longitudinally for a period of four months in DIO mice treated with Grb14-shRNA. There were no significant changes in hemodynamic function as measured by echocardiography between the treated animals with respect to its control. While additional studies are needed to further establish efficacy and to de-risk potential negative cardiac effects under heart failure conditions, our data support modulating Grb14 as a bona-fide strategy to treat insulin resistance and NASH.

## **PgmNr 378: The link between low back pain and diabetes: Gene-based association study in people with comorbid diabetes mellitus and low back pain.**

### **Authors:**

M. Schmitt <sup>1</sup>; G. Pathak <sup>2</sup>; N. Phillips <sup>2,3,4</sup>; J. Licciardone <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Texas College of Osteopathic Medicine, University of North Texas Health Science Center, Fort Worth, Texas.; 2) Department of Microbiology, Immunology and Genetics, University of North Texas Health Science Center, Fort Worth, Texas.; 3) Graduate School of Biomedical Sciences, University of North Texas Health Science Center, Fort Worth, Texas.; 4) Institute of Healthy Aging, University of North Texas Health Science Center, Fort Worth, Texas.; 5) Department of Family Medicine, University of North Texas Health Science Center, Fort Worth, Texas.

---

### **Purpose**

Diabetes mellitus type 2 is a disease associated with chronic pain. Mechanisms of this association are not fully understood. Few studies have explored the genetics of this association, and an independent genetic link between diabetes and low back pain (LBP) has not been completely investigated. We hypothesize that variants in candidate genes are associated with comorbid diabetes and LBP phenotypes.

### **Methods**

This cross-sectional study is a part of a larger study of subacute and chronic LBP within the PRECISION Pain Research Registry. Subjects were recruited from the Dallas-Fort Worth Metroplex and reported LBP for at least 2 months (subacute) or 6 months (chronic) for half or more of the days with the relevant time period.

Primary outcomes were assessed using a Numerical Rating Scale (NRS) for pain intensity, the Roland-Morris Disability Questionnaire for back-specific functioning, and the Patient-Reported Outcomes Measurement Information System for quality of life measures.

Study participants self-reported diabetes. Biological samples were collected to determine participants' genotypes using the Infinium Global Screening Array. Analyses were conducted using SPSS Statistics software among 488 study participants, 109 of whom reported diabetes.

Twenty-seven candidate genetic loci concurrently implicated in both pain phenotypes and diabetic phenotypes were identified from the literature. From these loci, 35 candidate genes were used to perform a gene-based association study (GBAS) to identify their potential association with diabetes and LBP. Genetic variants were mapped to candidate genes for association with each of the pain phenotypes after adjusting for age, sex, diabetes status, and ancestry components 1-10 using MAGMA computational software.

### **Results**

GBAS results point to variants in the PRKCA gene as being associated with both NRS of LBP intensity and diabetes. PRKCA is a family of protein kinases implicated in neuropathic pain and formation of advanced glycation end products.

### **Conclusions**

Of the 35 genes tested, only variants in PRKCA are significantly associated with LBP and diabetes in the PRECISION cohort. These findings build on a framework which suggests that genetic predisposition in PRKCA may underlie diabetes and LBP.

## PgmNr 379: Does 23&Me raw data accurately detect GCK-MODY?

### Authors:

H. Zhang; T.I. Pollin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Endocrinology, Diabetes and Nutrition, University of Maryland, Baltimore, Baltimore, Maryland.

### Introduction

Recent studies raise a concern about the accuracy and interpretation of direct-to-consumer genetic testing (DTC-GT) raw data for health information. We have found that 23andMe raw data contains 316 variants in the *GCK* (glucokinase) gene, including 273 reportedly pathogenic or likely pathogenic variants that have been implicated in the Maturity Onset Diabetes of the Young 2 (MODY2, or GCK-MODY), a lifelong form of mild hyperglycemia which can be misdiagnosed as type 1 or type 2 diabetes but requires no treatment. In this project, we used GCK-MODY as an example to discover how many consumers may find GCK-MODY causing variants in their raw data and how many are possibly real.

### Methods

23andMe DTC-GT raw data of 3040 anonymous consumers was downloaded from openSNP and *GCK* variant genotypes extracted for analysis. Call rate and Hardy-Weinberg equilibrium (HWE) p-value were calculated. Pathogenicity of exonic/ splicing variants found in openSNP contributors was classified using 2015 ACMG/AMP guidelines. We also evaluated how many *GCK* variants identified through sequencing in the Personalized Diabetes Medicine Program (PDMP), an implementation project to demonstrate the clinical utility of monogenic diabetes diagnosis, would have been detected through from the DTC-GT raw data.

### Results and Conclusion

488 (16.1%) consumers showed a total of 18 pathogenic/likely pathogenic variants, with only 2 having a call rate >99% and HWE  $p > 0.05$ . One pathogenic frameshift allele, c.174delG, was called in 239 consumers (call rate = 14%, HWE  $p < 0.01$ ), for a MAF of 8% compared to 0% in gnomAD, underscoring the value of population data for clarifying the likelihood of false positives (apparently 99.7% in this case). Finding 2 individuals with credible GCK-MODY variants in the 3040 samples (0.07%) is consistent with a recent population estimate 0.1% for this condition. The DTC-GT raw data could potentially have identified 11 of the 21 GCK-MODY patients from the PDMP, although only 5 of the 11 variants had call rates > 99%, with the other 6 having call rates of 80%-98%, suggesting a high risk for false negatives even for variants included on the array. While actual GCK-MODY cases may have been identified, these results also support an empirically high risk of false positive and negative results from DTC-GT raw data, indicating consumers should heed disclaimers about limitations of non-medical grade genetic testing accordingly and pursue medical grade testing when indicated.

## PgmNr 380: Genome-wide association study for serum alkaline phosphatase identifies novel associations influencing metabolic disease and diverse liver-related pathways.

### Authors:

Du <sup>1</sup>; V.L. Chen <sup>1,2</sup>; Y. Chen <sup>1</sup>; S.K. Handelman <sup>1,2</sup>; E.K. Speliotes <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Gastroenterology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109; 2) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109

---

**Introduction:** Serum alkaline phosphatase (ALP) is a commonly-used indicator of liver disease. ALP levels are genetically influenced but exact genetic and biochemical etiologies underlying ALP variation are not known.

**Methods:** We carried out genome-wide association study (GWAS) of inverse normally-transformed ALP with >20 million imputed SNPs using Scalable and Accurate Implementation of GEneralized mixed model, controlling for age, sex, and principal components, in >390,000 white-British individuals from UK BioBank. We then meta analyzed these results with ALP association results from >130,000 individuals of Japanese ancestry from BioBank Japan (5,951,600 imputed SNPs) with using METAL. Cross-trait analysis was performed using GWAS of quantitatively-measured nonalcoholic hepatic steatosis (GOLD consortium) and alcoholic cirrhosis (PMID 26482880) ( $P < 0.05$  reported), and 9 targeted metabolic UK BioBank traits/diagnoses (FDR < 0.05 reported). DEPICT analyses were carried out using SNPs with ALP association  $P < 1 \times 10^{-5}$ ; we reported tissues/pathways with FDR < 0.05.

**Results:** On meta analysis of UK BioBank and BioBank Japan, 316 SNPs associated with ALP at genome-wide significance ( $P < 5 \times 10^{-8}$ ), 301 of which were novel. Many ALP-increasing SNPs affected liver-related traits: SNPs in/near *GCKR* increased quantitatively-measured nonalcoholic hepatic steatosis, and others in/near *MARC1* increased risk of alcoholic cirrhosis. A range of metabolic patterns were seen: ALP-increasing SNPs in/near *MSL2* increased body mass index, triglycerides, and LDL-cholesterol while lowering serum HDL-cholesterol; other SNPs near *GCKR* produced a mixed phenotype with higher triglycerides and LDL-cholesterol but decreased diabetes, and yet others in/near *APOC1* caused decreased both triglycerides and LDL-cholesterol. Some ALP associated SNPs were in/near genes involved in glycoprotein biology (*ASGR1*), immunity (*TIRAP*), drug metabolism (*CYP2A6*), or monogenic liver disease (*HFE*). DEPICT showed enrichment in liver, followed by adipose tissue, small intestine, and pancreas. The top enriched gene sets were PPAR-alpha, nonspecific liver, and lipid localization pathways.

**Conclusions:** We identified 301 novel ALP-increasing SNPs in UK BioBank and BioBank Japan. These variants influence hepatic steatosis, alcoholic liver disease, and diverse metabolic traits/processes. Genes associated with ALP-increasing SNPs are primarily liver-expressed and highlight multiple pathways underlying liver disease.



## PgmNr 381: Genetic variants that associate with liver fibrosis and cirrhosis have pleiotropic effects on human traits.

### Authors:

E. Speliotes<sup>1,2</sup>; V.L. Chen<sup>1,2</sup>; Y. Chen<sup>1</sup>; X. Du<sup>1</sup>; S.K. Handelman<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Gastroenterology and Hepatology, University of Michigan Health System, Ann Arbor, MI, USA; 2) Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, USA

**Background:** Cirrhosis is characterized by extensive fibrosis of the liver and is a major cause of liver-related mortality. Cirrhosis is partially heritable but genetic contributions to cirrhosis have not been systemically explored. Here, we perform an unbiased genome-wide analysis of variants for association with cirrhosis in two large biobanks and determine the effects of cirrhosis associated variants on multiple human disease/traits.

**Methods:** We carried out a genome-wide association analysis of cirrhosis (based on diagnostic codes) as a diagnosis in UK BioBank (1,088 cases vs. 407,873 controls) and then tested top associating loci for replication with cirrhosis (based on diagnostic codes and a validated text search) in the hospital-based cohort Michigan Genomics Initiative (MGI; 875 cases of cirrhosis vs. 30,346 controls). For replicating single nucleotide polymorphisms (SNPs) or SNPs previously associated with cirrhosis, we determined their effects on other diseases (phenome-wide association study), common metabolic traits, and serum/plasma metabolites.

**Results:** SNPs near the genes *PNPLA3* and *HFE* associated with cirrhosis at genome-wide significance ( $p < 5 \times 10^{-8}$ ) in UK BioBank and replicated in MGI. We confirmed that previously-reported common SNPs in *HSD17B13*, *MBOAT7*, *TM6SF2*, *SERPINA1*, *STAT4*, and *IFNL4* associate with cirrhosis in UK BioBank or MGI. Most SNPs associated with liver enzyme abnormalities. Phenome-wide association study on these SNPs identified diverse and novel associations; the cirrhosis increasing alleles at *SERPINA1* increased height and cirrhosis promoting SNPs at *STAT4* increased respiratory infection. The cirrhosis increasing allele in *HFE* increased blood pressure but decreased serum low-density lipoprotein and total cholesterol while the cirrhosis promoting allele at *MBOAT7* decreased triglycerides and increased high-density lipoprotein. The cirrhosis promoting allele at *SERPINA1* increased low-/intermediate-density lipoproteins while the cirrhosis promoting allele at *TM6SF2* decreased these.

**Conclusions:** We identified loci associated with cirrhosis at genome-wide significant and validated previously-reported common variants for association with population based cirrhosis. These variants produce diverse effects on human diseases and traits, metabolic traits/disease, and circulating metabolites.

## PgmNr 382: HMI-102, an investigational gene therapy for phenylketoneuria (PKU).

### Authors:

D. Lamppu; S. Ahmed; L. Benard; J. Ellsworth; E. Faulkner; O. Francone; M. Lobikin; T. Wright; A. Seymour

[View Session](#) [Add to Schedule](#)

**Affiliation:** Homology Medicines, Inc, Bedford, Massachusetts.

---

HMI-102 is an investigational gene therapy for PKU due to phenylalanine hydroxylase (PAH) deficiency. The PAH enzyme converts phenylalanine (Phe) into tyrosine (Tyr). PAH deficiency is due to mutations in the *PAH* gene, resulting in excess Phe. It's inherited as an autosomal recessive, monogenic defect, making it suitable for potential AAV-based gene therapy.

PAH deficiency has a continuum of clinical phenotypes characterized by elevated blood Phe, ranging from mild hyperphenylalaninemia (HPA) (Phe 120-360  $\mu\text{mol/L}$ ) to classic PKU (Phe over 1200  $\mu\text{mol/L}$ ). Current US treatment guidelines indicate treatment is not required for mild HPA. Untreated PKU in children results in severe neurological impairment. Adults treated with Phe-restricted diets as children present with higher rates of neuropsychiatric comorbidities.

HMI-102 delivers the human *PAH* gene to the liver using AAVHSC15, a clade F vector isolated from human hematopoietic stem cells. The gene is transcribed and translated into active PAH. HMI-102 was tested in  $\text{PAH}^{\text{enu}2}$  mice, a murine model of PKU with several features consistent with the human classic PKU phenotype, including blood Phe over 1200  $\mu\text{M}$ . HMI-102 normalized blood Phe within one week of administration in mice on a normal chow diet. Phe reduction was sustained for 48 weeks (lifespan of the mouse model). HMI-102 normalized blood Tyr, 5-HIAA in the brain, and darkened the mouse coat color, demonstrating restoration of Phe metabolism.

In GLP studies in non-human primates and mice, there were no adverse test-article related findings. Durable vector genomes (vg) and mRNA expression were present in the livers; vg levels were comparable in both species at given weight-based doses. Dose modeling simulations for baseline Phe up to 2400  $\mu\text{mol/L}$  were done using  $\text{PAH}^{\text{enu}2}$  efficacy data to select doses for the first-in-human (FIH) study. The modeling predicted a robust response for the starting dose (>90% of subjects would achieve Phe <360  $\mu\text{mol/L}$ ).

A retrospective chart review was conducted to characterize blood Phe control in 152 patients with HPA. The data showed high Phe levels despite dietary restriction, demonstrating an unmet need for therapies to control blood Phe. The data also confirmed the FIH study design eligibility criteria and endpoints.

Collectively, the nonclinical data and retrospective study support initiation of the FIH study, which will assess the safety, tolerability, and efficacy of HMI-102 in adults with classic PKU.

## PgmNr 383: Host genetics and gut microbiome drive individual variation in blood metabolites in a large human cohort.

### Authors:

C. Dai <sup>1,2</sup>; S. Kornilov <sup>1,2</sup>; O. Manor <sup>2</sup>; J. Citronberg <sup>2</sup>; E. Sheng <sup>2</sup>; C. Bare <sup>1,2</sup>; B. Smith <sup>1,2</sup>; A. Magis <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Systems Biology, Seattle, WA; 2) Arivale, Seattle, WA

---

The human blood metabolome is comprised of thousands of small molecules, many of which are associated with human health and disease. Host genetics and environmental factors contribute to the variation in the levels of some of these metabolites. While certain genetic loci have been shown to be associated with peripheral blood metabolite concentrations, the extent to which the human gut microbiome influence these variations remains understudied. In particular, gene-environment interactions with the gut microbiome are uncharacterized. In this study, we investigated the associations between host genetics and gut microbiome features with the concentrations of 943 blood metabolites in a US-based cohort of 1,675 individuals. A genome-wide association analysis using whole-genome data identified 232 metabolites associated with 141 genomic loci (genome-wide significance  $P = 5.42 \times 10^{-12}$ ). We also identified microbiome-wide associations between 191 metabolites and 49 bacteria genera (microbiome-wide significance  $P = 5.36 \times 10^{-7}$ ) using a beta-binomial model. 40 of these metabolites were significantly associated with both host genetics and microbiome. Of these associated genetic loci and associated bacteria genera, 5 metabolites (2 amino acids, 2 xenobiotics, and 1 unidentified) had significant gene-environment interactions ( $P_{\text{FDR}} < 0.05$ ) between genetic variants and either *Faecalibacterium* or *Lachnospira*, two butyrate-producing bacteria genera. Taken together, we highlight several host-microbe interactions and demonstrate a joint genetic and microbial effect on blood metabolites. Understanding these relationships provides the foundation for future applications in precision medicine.

## PgmNr 384: Resilience to obesity in genetically at-risk individuals: A study on the underlying mechanisms.

### Authors:

R.A.J. Smit<sup>1,2</sup>; M. Graff<sup>3</sup>; R.W. Walker<sup>1,4</sup>; A.R. Wood<sup>5</sup>; M. Preuss<sup>1</sup>; K.L. Young<sup>3</sup>; E. Marouli<sup>6</sup>; C.M. Lindgren<sup>7,8</sup>; K.E. North<sup>3</sup>; T.M. Frayling<sup>5</sup>; J.N. Hirschhorn<sup>9,10,11</sup>; R.J.F. Loos<sup>1,12,13</sup>; on behalf of the Genetic Investigation of ANthropometric Traits (GIANT) Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 2) Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, NL; 3) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 4) Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 5) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Royal Devon and Exeter NHS Trust, Exeter, UK; 6) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 7) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 8) Li Ka Shing Centre for Health Information and Discovery, Big Data Institute, University of Oxford, Oxford, UK; 9) Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA; 10) Departments of Genetics and Pediatrics, Harvard Medical School, Boston, MA, USA; 11) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA; 12) The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 13) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA

---

Obesity is a major risk factor for chronic disease. Besides an obesogenic lifestyle, genetic factors also determine susceptibility to obesity. Using a polygenic risk score (PRS) for obesity, those in the top 10% were shown 25x more likely to be severely obese compared to the bottom 10%. Interestingly, a proportion (17%) of the top 10% were of normal weight. To gain insight into the mechanisms that determine resilience to genetic predisposition to obesity, we characterized normal weight and obese individuals with a high (top 5%) and low (bottom 5%) PRS within the UK Biobank.

Using PRSice, we built a PRS with BMI association summary statistics of 2.1 million variants (Locke et al. 2015) that we tested in 372,584 unrelated UKB participants of European ancestry. We defined 6 groups; by PRS [bottom 5%, middle 35-65%, top 5%) and BMI-category (under/normal weight [BMI: <25kg/m<sup>2</sup>, NW] and obese [BMI≥30, OB]), and compared demographics, body composition, clinical history, and measures of the obesogenic environment. All analyses were adjusted for age, sex, center, and the first ten PCs.

In the top 5% of the PRS (N=18,630), 44.5% were OB, whereas 17% were NW. In the bottom 5% PRS (N=18,628), the discrepancy was more pronounced; 53.9% were NW, and only 9% were OB. Among NW individuals, those in the top 5% PRS had ( $P<1e-3$ ) somewhat higher adiposity, lower energy intake, more sedentary time, were more often smokers, were less educated, and had more often diabetes, compared to those in the bottom 5%, though material deprivation and birth weight were

similar. Among OB individuals, we saw similar differences. In the high-PRS group, 84% of the NW reported to be thinner or of average body size at age 10, compared to 59% of the OB ( $P < 2e-16$ ). Conversely, in the low-PRS group, 27% of OB reported to be “plumper” at age 10, compared to 6% of NW ( $P < 2e-16$ ). Among the high-PRS group, NW were ( $P < 1e-3$ ) more physically active and educated, less sedentary and materially deprived, fewer had diabetes, but more were smokers, than OB. Inverse trends were observed in the low-PRS group.

Taken together, body size tracks across the life course, which seems to be largely driven by a genetic predisposition (PRS), but also by (non-)genetic factors not captured by the PRS. Moreover, material deprivation, education and lifestyle may counteract one’s genetic predisposition to obesity. Understanding the genetic and non-genetic contributors to body size can help define the scope and timing of prevention.

## PgmNr 385: Deconstructing polygenic scores for type 2 diabetes to capture etiological and clinical heterogeneity.

### Authors:

A. Mahajan; N. Robertson; M.I. McCarthy

[View Session](#) [Add to Schedule](#)

**Affiliation:** Wellcome Centre for Human Genetics, Univ Oxford, Oxford, United Kingdom

Over 400 robust signals of T2D association have been identified to date. Whilst “global” polygenic scores using all signals of T2D predisposition are appropriate for risk prediction, “partitioned” scores (pPS) that capture the etiological diversity of T2D may have more value for clinical decision-making related to the heterogeneity within T2D.

Here, we extend previous pPS analyses to an expanded set of 337 T2D signals, using clusters defined on the basis of multi-trait association patterns across 10 core traits. Fuzzy clustering defined six genetic clusters (GCs), capturing variants with primary effects on adiposity (GC1-BMI), lipids (GC2-lipids), insulin action (GC3-IA) and beta-cell function (GC4-BCF, GC5-BCF), plus a sixth with mixed features (GC6-MIX).

We considered the impact of individual T2D-pPS loadings on complication risk utilizing publicly available GWAS for CAD, atrial fibrillation (AF), hypertension, CKD, renal function (eGFR, albuminuria), stroke and fatty liver (liver fat%), and observed multiple GC-specific patterns of association with outcomes ( $p < 0.0007$ ). T2D-risk mediated by GC1-BMI was exclusively associated with AF (OR=1.3 per 1SD pPS increase), whereas GC2-Lipids associated with CKD (OR 1.4) and reduced eGFR (OR 0.95). Hypertension and CAD were most strongly associated with GC1-BMI and GC3-IA (OR>1.3). T2D-risk attributable to GC2-Lipid scores was associated with lower CAD risk (OR 0.60) but increased liver fat (OR 2.5), counter to the broader epidemiological association between them.

Next, we used available metabolomic and proteomic data to identify candidate biomarkers for each of these processes. We detected 56 protein, 83 metabolite, and 26 glycan associations ( $q$ -value<0.05 within each marker group). Many of these were associated with the GC2-Lipid pPS (including AKT2, APOB, APOE). Other candidates included IGFBP2 and APOF as markers of insulin resistance (GC3-IA). GC4-BCF and GC5-BCF pPS showed divergent associations with proinsulin (high vs low levels, respectively) pointing to distinct mechanisms of beta-cell dysfunction. Circulating levels of ATF6, a known marker of ER stress response, were associated exclusively with GC5-BCF.

Our findings demonstrate that partitioned polygenic scores capture individual differences in the etiological contributions to T2D development that can be related to clinically relevant outcomes, and which provide a mechanistic framework for understanding disease heterogeneity.

## PgmNr 386: The relationship between PCSK9 and sepsis: A genetic approach.

### Authors:

Q. Feng <sup>1</sup>; W.Q. Wei <sup>2</sup>; S. Chaugai <sup>1</sup>; B.G. Carranza Leon <sup>3</sup>; V. Kawai <sup>1</sup>; D.A. Carranza Leon <sup>1</sup>; L. Jiang <sup>1</sup>; X. Zhong <sup>4</sup>; G. Liu <sup>1</sup>; A. Ihegword <sup>1</sup>; C.M. Shaffer <sup>1</sup>; M.F. Linton <sup>5,6</sup>; C.P. Chung <sup>1,7</sup>; C.M. Stein <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee.; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 3) Division of Diabetes, Endocrinology and Metabolism, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN;; 4) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; Vanderbilt Genetic Institute, Vanderbilt University, Nashville, TN; 5) Department of Pharmacology, Vanderbilt University, Nashville, TN; 6) Division of Cardiovascular Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 7) Division of Rheumatology & Immunology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN

**Objective** Inhibition of proprotein convertase subtilisin kexin type 9 (PCSK9) decreases risk of sepsis in animal models. We tested whether functional *PCSK9* genetic variants, a *PCSK9* genetic risk score (GRS), or genetically-predicted *PCSK9* expression levels were associated with the risk of sepsis in patients admitted to the hospital with infection.

**Design, Setting and Participants** De-identified electronic health records were used to define a cohort of patients admitted to Vanderbilt University Medical Center, Nashville, Tennessee, with infection. Patients were white adults, had an International Classification of Diseases, Ninth Revision, Clinical Modification (ICD9) or ICD10 code indicating infection and received an antibiotic within 1 day of hospital admission (N=61502).

**Interventions** (1) 4 known *PCSK9* functional variants (rs11591147, rs11583680, rs562556 and rs505151, N=10922), (2) a GRS for *PCSK9* (N=7624), and (3) genetically-predicted *PCSK9* expression (N=6033).

**Main Outcomes** The primary outcome was sepsis; secondary outcomes included cardiovascular failure (use of a pressor agent) and in-hospital death.

**Results** None of the 4 functional *PCSK9* variants were significantly associated with sepsis, cardiovascular failure, or in-hospital death, with or without adjustment for (1) age and sex, or (2) age, sex, and Charlson/Deyo comorbidities. In the fully-adjusted model, the odds ratios [ORs] (95% CI) for rs11591147 were: sepsis, 0.84(0.64-1.09); cardiovascular failure, 1.24(0.82-1.85); and in-hospital death, 0.73(0.36-1.51). The ORs for rs11583680 were: sepsis, 1.01(0.93-1.10); cardiovascular failure, 0.94(0.81-1.09); and in-hospital death, 1.06(0.86-1.31). The ORs for rs562556 were: sepsis, 0.99(0.92-1.07); cardiovascular failure, 1.02(0.90-1.17); and in-hospital death, 0.99(0.81-1.21). The ORs for rs505151 were: sepsis, 0.92(0.78-1.08); cardiovascular failure, 1.05(0.80-1.37); and in-hospital death, 0.80(0.51-1.26). Similarly, the *PCSK9* GRS were not significantly associated with sepsis 1.01 (0.96-1.06), cardiovascular failure, 1.03 (0.95-1.12), and in-hospital death 1.05 (0.92-1.19). Genetically-predicted *PCSK9* expression was not associated with sepsis 1.01 (0.95-1.06), cardiovascular failure 0.96 (0.88-1.05), and in-hospital death 0.99 (0.87-1.14). All p-values>0.05.

**Conclusion** *PCSK9* genetic variants were not associated with risk of sepsis in patients hospitalized



with infection nor were they associated with the outcomes of sepsis.

## PgmNr 387: An African ancestry uterine fibroids polygenic risk score (PRS) identifies associations with other gynecologic conditions in the clinical phenome.

### Authors:

J.N. Hellwege<sup>1,2</sup>; J.A. Piekos<sup>2</sup>; Y. Zhang<sup>3</sup>; E.S. Torstenson<sup>2,4</sup>; G.P. Jarvik<sup>5</sup>; O. Dikilitas<sup>6</sup>; I.J. Kullo<sup>6</sup>; D.J. Schiada<sup>7</sup>; S.A. Pendergrass<sup>8</sup>; D.M. Roden<sup>2,9</sup>; J.C. Denny<sup>2,10</sup>; T.L. Edwards<sup>2,4</sup>; D.R. Velez Edwards<sup>2,10,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 3) Genomic Medicine Institute, Geisinger, Danville, PA; 4) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA; 6) Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN; 7) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 8) Biomedical and Translational Informatics, Geisinger, Danville, PA; 9) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 10) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 11) Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN

---

Uterine fibroids are the most common female pelvic tumor with prevalence up to 77% by menopause in African Americans. African American women have increased risk of fibroids than other continental populations, as well as larger and more numerous fibroids. The known genetic architecture of uterine fibroids currently includes ~30 loci, most discovered in European populations. We conducted a phenome-wide association study (PheWAS) of uterine fibroids polygenic risk scores (PRS) to understand the shared genetic contributions across clinical phenotypes. We constructed an African-ancestry uterine fibroids PRS of 317 SNPs using a p-value threshold of  $<1 \times 10^{-4}$ . Effect sizes were derived from imaging-confirmed uterine fibroids genome-wide association study (GWAS) of 1,272 cases and 1,379 controls from the Electronic Medical Records and Genomics (eMERGE) network. We then performed PheWAS across 6,061 independent women from eMERGE using the PRS as the predictor for clinical diagnoses adjusted for age, body mass index, and principal components. We identified 26 significant ( $p\text{-value} < 2.5 \times 10^{-5}$ ) results among 1,381 diagnoses. The top association was uterine fibroids ( $p\text{-value} = 6.42 \times 10^{-66}$ ), showing that the PRS is well-calibrated even when not restricted to imaging-confirmed cases and controls. Other top results included genitourinary conditions related to uterine fibroids, including six diagnoses related to menstrual dysregulation. There were also novel associations with other gynecological conditions including endometriosis ( $p\text{-value} = 2.51 \times 10^{-13}$ ), ovarian cysts ( $p\text{-value} = 1.26 \times 10^{-7}$ ), and abnormal Papanicolaou smear ( $p\text{-value} = 1.55 \times 10^{-6}$ ). Overall, results suggested that this imaging-confirmed uterine fibroids PRS is a valid correlate of the genetic risk underlying uterine fibroids development. We also observed associations with other gynecologic conditions, including neoplastic phenomena, suggesting shared genetic risks.

## PgmNr 388: Eculizumab therapy for protein-losing enteropathy due to CD55-deficiency: Insights from a two-years-long follow-up.

### Authors:

A. Kurolap <sup>1,2</sup>; O. Eshach Adiv <sup>2,3</sup>; T. Hershkovitz <sup>1</sup>; A. Tabib <sup>4</sup>; N. Karbian <sup>4</sup>; T. Paperna <sup>1</sup>; A. Mory <sup>1</sup>; Y. Zohar <sup>5</sup>; D. Mevorach <sup>4</sup>; H. Baris Feldman <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Genetics Institute, Rambam Health Care Campus, Haifa, Israel; 2) The Ruth and Bruce Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel; 3) Pediatric Gastroenterology Unit and Pediatrics B, Rambam Health Care Campus, Haifa, Israel; 4) Rheumatology Research Center, Hadassah Medical Center and the Hebrew University, Jerusalem, Israel; 5) Institute of Pathology, Rambam Health Care Campus, Haifa, Israel

---

**Introduction:** The *CD55* gene encodes the decay accelerating factor, a complement system regulator. Bi-allelic loss-of-function variants in *CD55* cause an early-onset protein-losing enteropathy (PLE) syndrome, associated with intestinal lymphangiectasia and susceptibility to large-vein thrombosis. A short-term three-months-trial of eculizumab (C5-inhibitor) therapy in three *CD55*-deficiency patients showed marked improvement in disease manifestations. Here we present the two-year treatment outcomes for these patients, highlighting treatment safety and sustained therapeutic response.

**Methods:** Off-label eculizumab treatment was administered to three *CD55*-deficiency patients. Clinical treatment outcomes included frequency and consistency of bowel movements, weight, patient/parent reports of overall well-being; laboratory outcomes were underscored by serum albumin and total protein levels. Membrane attack complex (MAC) deposition on leukocytes was tested by flow cytometry, before and throughout treatment.

**Results:** All three patients had marked clinical improvement in all components, including resolution of PLE manifestations (diarrhea, edema, malabsorption), hypercoagulability, overall well-being, growth and quality of life. Correspondingly, all laboratory outcome parameters improved and normalized, including albumin and total protein levels, and up to 80% reduction in MAC deposition was observed on leukocytes ( $p < 0.001$ ). There were no severe adverse events over two years of treatment.

**Conclusions:** Patients with *CD55*-deficiency present with early-onset diarrhea, edema, severe hypoalbuminemia, abdominal pain, malnutrition and susceptibility to venous thrombosis. Previously a life-threatening condition, *CD55*-deficiency is highly responsive to targeted therapy with the terminal complement inhibitor eculizumab, with positive clinical and laboratory outcomes. This study underscores the significant role of the complement system in the gut, and warrants further studies on the function of the complement system and *CD55* in other gastrointestinal disorders.

## PgmNr 389: New frontiers to cure Alport syndrome by a personalized CRISPR/Cas9 gene editing approach.

### Authors:

A.M. Pinto <sup>1</sup>; S. Daga <sup>2</sup>; C. Rivera <sup>7</sup>; S. Clark <sup>7</sup>; F. Donati <sup>3,6</sup>; K. Capitani <sup>3,6</sup>; S. Croci <sup>2</sup>; A. Giliberti <sup>2</sup>; F. Valentino <sup>2</sup>; F. Ariani <sup>1,2</sup>; R. Tita <sup>1</sup>; C. Fallerini <sup>2</sup>; M.A. Mencarelli <sup>1</sup>; M. Baldassarri <sup>1</sup>; E. Benetti <sup>1</sup>; S. Furini <sup>6</sup>; E. Frullanti <sup>2</sup>; F. Niccheri <sup>3</sup>; A. Auricchio <sup>4,5</sup>; S. Conticello <sup>3</sup>; M. Nabity <sup>7</sup>; A. Renieri <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, University of Siena, Siena, Siena, Italy; 2) 2.Medical Genetics, University of Siena, Siena, Italy; 3) Core Research Laboratory - ISPRO, Florence, Italy; 4) Telethon Institute of Genetics and Medicine (TIGEM), Pozzuoli, Italy; 5) .Department of Advanced Biomedicine, Federico II University, Naples, Italy; 6) Department of Medical Biotechnologies, University of Siena, Siena, Italy.; 7) Department of Veterinary Pathobiology, Texas A&M University College of Veterinary Medicine and Biomedical Sciences, College Station, Texas.

Alport syndrome (AS) is an inherited genetic disorder characterized by clinicopathologic alterations ranging from glomerular basement membrane abnormalities to end-stage renal disease. Mutations in the collagen  $\alpha 3$ ,  $\alpha 4$  and  $\alpha 5$  encoding genes are causative of both the autosomal and X-linked forms of AS. A curative therapy for AS is still lacking, and the podocytes' functional disruption can be only partially recovered by symptomatic treatments.

Podocytes, a key component of the glomerular structure, are the only kidney cells able to produce the COLIV $\alpha 3$ - $\alpha 4$ - $\alpha 5$  heterotrimer. We have previously demonstrated that it is possible to isolate podocyte-lineage cells from urine of patients. Building upon the availability of urine-derived podocytes as an easy source of disease-relevant cells, we have established the efficiency of CRISPR/Cas9 gene therapy to restore the wild-type genotype *in-vitro*. We have engineered a two-plasmid system to correct the causative mutation in two stable podocyte-lineage cell lines, harboring a mutation in the X-linked COL4A5 (p.(Gly624Asp)) and in the autosomal COL4A3 gene (p.(Gly856Glu)). We have achieved reversion of mutations greater than 40% with undesired insertions/deletions lower than 15%. CRISPR/Cas9 gene editing experiments on urine-derived podocyte-lineage cells from a naturally occurring dog model harboring a deletion of 10 bp in exon 9 of COL4A5 replicated the results in human cells with a rate of correction greater than 50%.

In vivo experiments using our two-plasmid system to correct the COL4A5 mutation in the AS dog model are currently ongoing. A single femoral artery catheterization will be used to deliver the AAV2-CRISPR/Cas9 correction system *in-vivo*, injecting into both renal arteries. The procedure will be performed at 8 weeks of age, which is typically before the occurrence of microalbuminuria, using one of two dosages (1011-1012 and 1013-1014) in two affected males. A third affected male dog will be treated with mut-AAV2-CRISPR/Cas9 as a negative control.

The described approach can provide proof-of-principle for a gene therapy strategy for AS treatment, opening up the possibility of in-vivo clinical trials on AS patients based on personalized transformative medicine tailored to the pathogenic mechanism and able to directly act on disease-relevant cells.

## PgmNr 390: Uterine leiomyomata Polygenic Risk Scores (PRS) confer novel relationships in the clinical phenome.

### Authors:

J.A. Piekos<sup>1</sup>; J.N. Hellwege<sup>1,2</sup>; Y. Zhang<sup>3</sup>; E.S. Torstenson<sup>1,4</sup>; G.P. Jarvik<sup>5</sup>; O. Dikilitas<sup>6</sup>; I.J. Kullo<sup>6</sup>; D.J. Schaid<sup>7</sup>; D.R. Crosslin<sup>5</sup>; S.A. Pendergrass<sup>8</sup>; D. Roden<sup>1,9</sup>; J.C. Denny<sup>1,10</sup>; T.L. Edwards<sup>1,4</sup>; D.R. Velez Edwards<sup>1,10,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 3) Genomic Medicine Institute, Geisinger, Danville, PA; 4) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA; 6) Department of Cardiovascular Medicine, Mayo Clinic, Rochester, MN; 7) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 8) Biomedical and Translational Informatics, Geisinger, Danville, PA;; 9) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Center, Nashville, TN; 10) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 11) Division of Quantitative Science, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN

Uterine leiomyomata (UL) are the most common cause of pelvic tumors in women with lifetime incidence of 70%. Little is known about the etiology of UL. Prior Genome-Wide Association Studies (GWAS) in predominately European Ancestry (EA) populations have identified ~30 loci that are associated with UL. We developed a UL polygenic risk score (PRS) and combined it with a Phenome-Wide Association Study (PheWAS) approach to gain understanding about the shared genetic contribution across many clinical phenotypes. We constructed the PRS with PRSice software using effect sizes derived from a non-Hispanic EA, imaging-confirmed UL GWAS set from the Electronic Medical Records and Genomics (eMERGE) Network (N=2,651 cases and 4,326 controls). It was optimized in an independent set of UL imaging-confirmed cases and controls (N= 429 cases, 4,750 controls) from BioVU. The PRS consisted of 4,448 linkage disequilibrium pruned ( $r^2 < 0.2$ ) variants ( $p < 0.001$ ) and the model was significant ( $r^2 = 0.0017$ ,  $p$ -value = 0.041). PheWAS analyses were performed in 53,116 non-Hispanic EA and 5,583 non-Hispanic African-American (AA) men and women from eMERGE, excluding samples used for model construction, using the PRS as the predictor for 1,738 clinical disease phenotypes adjusted for sex, age, body mass index, and 10 principal components. Our PRS PheWAS detected UL as the most significant phenotype for non-Hispanic EAs combined sexes and non-Hispanic EAs females-only test ( $p = 9.58 \times 10^{-167}$ ). We detected 40 and 47 significant phenotypes in non-Hispanic EAs and non-Hispanic EA females, respectively. Top phenotype results included dysmenorrhea, other benign neoplasms of the uterus, endometriosis, and malignant neoplasm of the uterus. Many associated phenotypes were in the genitourinary category; we also detected associations in the neoplasm and sense organ category. The sense organ phenotypes were eye diseases including astigmatism, optic atrophy, and myopia. In the non-Hispanic AA group, the PRS did not detect any phenotypes at genome wide significance, possibly due to power. Validation of the PRS is underway in the BioVU Repository (N=94,000) and Geisinger Health Systems (N=85,582). Our results indicate that UL shares biological etiology with many other diseases, most notably other

neoplasms and eye diseases. Determining the relationships between UL and these other diseases may reveal more about their molecular mechanisms and broaden our understanding about these diseases.

## PgmNr 391: Understanding intersectionality in health disparities through genetic models.

### Authors:

K. Actkins<sup>1,3</sup>; P. Straub<sup>2,3</sup>; N. Cox<sup>2,3</sup>; L. Davis<sup>2,3</sup>; on behalf of the Precision Medicine and Health Disparities Collaborative

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Microbiology, Immunology, and Physiology, Meharry Medical College, Nashville, Tennessee; 2) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; 3) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, Tennessee

---

It is widely known that health disparities are multifactorial with origins that can be biological, environmental, social, or political. Their combined impact on marginalized groups has significantly lowered their quality of life, increased their difficulty of successful treatment and care, and impeded any positive long-term standards of living. These realizations have ignited the precision health initiative and the genetic research field has made substantial efforts to understand the underlying genetic and biological effects of these diseases. However, many of the current genetic models used to identify and analyze these effects often lack the data and statistical methods necessary to capture non-biological effects leading to misrepresented models and misinterpreted results. Disentangling these factors is crucial to understanding the structure of these disparities and we created a genetic model that allows us to analyze these effects independently. Using the Vanderbilt University Medical Center biobank (BioVU), a repository of DNA linked to de-identified electronic medical records, we observed that a high proportion of diseases significantly associated with African ancestry are confounded with race. Although genetic ancestry represents your ancestral origin and race is a social construct, these two concepts are often used interchangeably or not fully accounted for in current models. We partitioned out genetic ancestry from self-reported race by including their residuals in our phenome-wide association study (PheWAS). After adjusting for genetic ancestry, race is significantly associated with "Hypertension" ( $p=7.57e-05$ ,  $OR=1.16$ ) and "Malaise and fatigue" ( $p=2.05e-04$ ,  $OR=1.19$ ) in African Americans. After adjusting for race, African genetic ancestry is significantly associated with "Other anemias" ( $p=5.62e-07$ ,  $OR=1.32$ ) and "Disorders resulting from impaired renal function" ( $p=6.31e-06$ ,  $OR=1.70$ ). This method allows us to examine the effects of genetic ancestry and race separately which will enable us to parse out our resources and efforts to each contributing cause more effectively.



## PgmNr 392: TLA & long-read sequencing: Efficient targeted sequencing and phasing of the *CFTR* gene.

### Authors:

C. Heiner<sup>1</sup>; E. Splinter<sup>2</sup>; V. Boersma<sup>2</sup>; M. Simonis<sup>2</sup>; K. Hajo<sup>2</sup>; J. Harting<sup>1</sup>; L. Aro<sup>1</sup>; J. Ziegle<sup>1</sup>; M. van Min<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pacific Biosciences, Menlo Park, California.; 2) Cergentis, Utrecht, The Netherlands

---

### Background

The sequencing and haplotype phasing of entire gene sequences improves the understanding of the genetic basis of disease and drug response. One example is cystic fibrosis (CF). CFTR (cystic fibrosis transmembrane conductance regulator) modulator therapy has revolutionized CF treatment, but only in a minority of CF subjects.

Observed heterogeneity in CFTR modulator efficacy is related to the range of CFTR mutations; revertant mutations can modify the response to CFTR modulators, and other intronic variations in the ~200kb CFTR gene have been linked to disease severity. Heterogeneity in the CFTR gene may also be linked to differential responses to CFTR modulators.

The Targeted Locus Amplification (TLA) technology from Cergentis can be used to selectively amplify, sequence and phase the entire CFTR gene. With PacBio long-read SMRT sequencing TLA amplicons are sequenced intact and long-range proximity information of all fragments in entire amplicons is retrieved.

### Experimental Design and Methods

TLA was performed on cell line and genomic DNA from Coriell GM12878, which has few heterozygous SNVs in CFTR, and the IB3 cell line, with known haplotypes but heterozygous for the delta508 mutation. All sample types were prepared with high and low density TLA primer sets, targeting coverage of >100kb of the (~200kb) CFTR gene. The TLA process produced amplicons consisting of 5-10 proximity ligated DNA fragments.

A multiplexed SMRTbell library of 10 TLA samples was sequenced on the PacBio Sequel System (1M SMRT cell), generating hundreds of thousands of high-fidelity single molecule reads using CCS (Circular Consensus Sequencing). CCS sequences were mapped, resulting in phased SNPs across the entire gene. Phased reads were clustered and compared between DNA samples. Additionally, we evaluated the coverage and data quality from the different sample types and TLA primer sets.

### Conclusion

We demonstrate the power and utility of TLA with long-read SMRT sequencing as a valuable research tool in sequencing and phasing across very long regions of the human genome. This process can be

done in an efficient manner, multiplexing multiple samples per SMRT cell in a process amenable to high-throughput sequencing.

## PgmNr 393: Pharmacogenomics and opioid use patterns in patients with chronic spine pain.

### Authors:

K. Fisch <sup>1</sup>; R. Gabriel <sup>2,3</sup>; B. Shahidi <sup>4</sup>; E. Said <sup>2</sup>; R. Waterman <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Center for Computational Biology & Bioinformatics, University of California, San Diego, La Jolla, CA; 2) Department of Anesthesiology, University of California, San Diego, La Jolla, CA; 3) Department of Medicine, Division of Biomedical Informatics, University of California, San Diego, La Jolla, CA; 4) Department of Orthopaedic Surgery & Radiology, University of California, San Diego, La Jolla, CA

---

Patients undergoing spinal surgery have a higher risk of persistent opioid use compared to other surgical subtypes, with 23-52% of patients developing long term use patterns and dependence after surgery. Interindividual differences in opioid analgesia associated with genetic variability have been well-documented in surgical populations. This study aims to assess the genetic variability in a postsurgical spine surgery population through pharmacogenomics (PGx) testing on pain and opioid use outcomes. PGx testing of perioperative patients at UC San Diego Health was performed through CQuentia NGS, LLC. Under an IRB-approved protocol, diplotypes for 53 spine surgery patients were retrospectively extracted from PGx reports. PharmGKB and CPIC gene-specific information tables and EHR priority result notifications were used to distinguish Normal from Abnormal drug metabolism phenotypes for each of these diplotypes. Outcomes included baseline and three month postoperative pain intensity ratings and postoperative opioid consumption in intravenous morphine equivalents at 24, 48 and 72 hours. A Wilcoxon rank sum test was used to test for differences in outcomes in patients with clinically actionable CYP2D6 variants. Multivariate linear regression analysis was performed on each PGx gene to determine association with increased opioid use postsurgically at 48 hours controlling for sex, age and body mass index. Clinically actionable PGx variants were observed in the perioperative population (N= 3,527) at the following frequencies: CYP2D6-14%, CYP2C19-28.9%, CYP2C9-3.1%, CYP3A4-7.9%, CYP3A5-4.4%, DPYD-1.8%, TPMT-8.5%. In patients with an actionable CYP2D6 variant for opioid metabolism, 247 patients that were prescribed opioids may have benefited from a change in clinical care based on the PGx results (Hydrocodone: 115 patients; Oxycodone: 56 patients; Tramadol: 76 patients). Of 53 spine surgery patients, 11 patients (20.7%) had actionable PGx variants in CYP2D6 (2 ultra; 9 poor metabolizers), indicating an altered opioid regimen is warranted. Baseline and postoperative numeric pain rating scale and opioid consumption did not significantly differ in these CYP2D6 patients ( $p > 0.05$ ) and there were no statistically significant differences between cohorts at any PGx variant for 48 hour opioid consumption ( $p > 0.05$ ). PGx testing may benefit a clinically relevant subset of patients undergoing spine surgery to improve postsurgical pain and opioid use outcomes.

## PgmNr 394: Low dose, daily or intermittent administration of infigratinib (BGJ398), a selective FGFR inhibitor, as treatment for achondroplasia in a preclinical mouse model.

### Authors:

L. Legeai-Mallet <sup>1</sup>; G. Li <sup>2</sup>; C. Dambkowski <sup>2</sup>; B. Demuynck <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Imagine Institute, Inserm U1163, Paris Descartes University, Paris, France; 2) QED Therapeutics, Inc. 75 Federal Street, San Francisco, CA, USA

---

Fibroblast Growth Factor Receptor 3 (FGFR3) plays a crucial role in the process of bone elongation as genetically demonstrated by the identification of FGFR3 gain-of-function mutations in people exhibiting the most common forms of dwarfism, namely achondroplasia (ACH) and hypochondroplasia. Various therapeutic strategies have been considered, among which the most advanced treatment is to employ an analog of C-type Natriuretic Peptide (CNP) which antagonizes the MAP kinase pathway. Here, considering that FGFR3 activates many downstream signaling pathways (e.g., STAT1), we evaluated a therapeutic strategy aimed to directly target the activation of FGFR3 and all signaling pathways, not just MAP kinase. The oral tyrosine kinase inhibitor (TKI) infigratinib (BGJ398) is a potent, selective FGFR1-3 inhibitor with well characterized activity against FGFR3. We hypothesized that a very low dose of infigratinib would be able to improve the defective bone elongation. To answer this question, we treated a dwarf mouse model (*Fgfr3*<sup>Y367C/+</sup>) mimicking ACH, with subcutaneous injections of infigratinib either daily (0.2 mg/kg/day or 0.5 mg/kg/day) or intermittently (1 mg/kg, every three days) for a total treatment duration of 15 days (PND1-PND15). Results were compared to vehicle treated mutant mice. We observed a significant improvement of the upper (humerus +7%, ulna +11%) and lower (femur +11%, tibia +16%) limbs for a dose of 0.5 mg/kg, as well as improvement in the foramen magnum. The effect of the bone elongation was reduced with a lower dose (0.2 mg/kg), thus confirming a dose-response relationship. To test the hypothesis of whether daily treatment was needed, we performed intermittent injections of infigratinib (1 mg/kg, every three days). The gain of growth compared to vehicle treated mice was significant for all the long bones (+7%) and the size of the foramen magnum was increased. In addition to the gain of growth, we observed a modification of the growth plate structure, displaying a better organization of the hypertrophic zone, among other improvements. In conclusion, these data demonstrated that low, as well as intermittent dose, of infigratinib promotes growth in this ACH mouse model. No apparent toxicity of infigratinib was observed, thus suggesting that TKI therapy, as with infigratinib, has the potential to be a valuable and relevant option for children with achondroplasia.

## PgmNr 395: TGF-beta inhibition restores the responsiveness to osteoanabolic PTH treatment in the *Crtpap*<sup>-/-</sup> model of recessive osteogenesis imperfecta.

### Authors:

I. Grafe; J. Zieba; Y. Chen; E. Munivez; M.M. Jiang; B. Dawson; A. Castellon; J. Sliepka; C. Jiang; S. Nagamani; B. Lee

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TEXAS.

Osteogenesis Imperfecta (OI) is characterized by low bone mass and fractures. Most cases are caused by dominant pathogenic variants in the genes encoding type I collagen. Pathogenic variants in *CRTAP*, involved in post-translational collagen modifications, cause severe recessive OI. Previously, we have shown that excessive TGFb signaling is an important contributor to the OI phenotype in mouse models of both dominant and recessive OI, including *Crtpap*<sup>-/-</sup> mice.

Intermittent PTH treatment increases bone mineral density (BMD) in patients with osteoporosis. In OI however, a randomized clinical trial found that PTH treatment did not increase BMD in patients with severe forms of the disorder. Others have demonstrated interactions between the TGFb and PTH signaling pathways, e.g. that TGFb signaling can reduce PTH receptor cell surface expression and thereby decrease PTH signaling. Our preliminary studies suggest that BMSCs of *Crtpap*<sup>-/-</sup> mice, cultured in osteogenic conditions, exhibit a reduced pCREB/CREB ratio, suggesting impaired PTH signaling.

Hence, we hypothesized that inhibition of the excessive TGFb signaling in *Crtpap*<sup>-/-</sup> mice restores the osteoanabolic effects of PTH treatment. To test this hypothesis, we treated 8-week-old *Crtpap*<sup>-/-</sup> and wild-type (WT) mice with PTH (40 µg/kg, 5x/week, s.c.), low-dose of the TGFb neutralizing antibody 1D11 (Genzyme/Sanofi; 1mg/kg, 3x/week IP), or control treatment (vehicle/control antibody) for 8 weeks.

MicroCT of femurs of control *Crtpap*<sup>-/-</sup> mice confirmed reduced bone mass as measured by trabecular bone volume/total volume (BV/TV; 38% of WT control; p<0.05) and cortical thickness (Cort.Th; 90% of WT; p<0.05). PTH treatment increased trabecular and cortical bone mass in WT mice, but was not effective in *Crtpap*<sup>-/-</sup> mice, consistent with the findings of the clinical PTH treatment trial in patients with more severe forms of OI. Low-dose TGFb inhibition in *Crtpap*<sup>-/-</sup> mice only moderately increased BV/TV (to 57% of WT; n.s.) and did not improve Cort.Th. Interestingly, combined PTH/anti-TGFb treatment of *Crtpap*<sup>-/-</sup> mice synergistically restored BV/TV and Cort.Th to 95% and 98% of control WT mice, respectively (p<0.05 for both).

These findings suggest that low-dose TGFb inhibition restores responsiveness to PTH treatment in the *Crtpap*<sup>-/-</sup> model of severe recessive OI. Because increased TGFb signaling also contributes to the bone phenotype in dominant OI, combined anti-TGFb and PTH treatment might be a new treatment option for patients with OI.

## PgmNr 396: Fine mapping of human leukocyte antigen complex to study asthma in African Americans.

### Authors:

H. Gui<sup>1,2</sup>; A. Mak<sup>3</sup>; P. Sleiman<sup>4</sup>; A.M. Levin<sup>5</sup>; S. Xiao<sup>1,2</sup>; M. Yang<sup>1,2</sup>; J. Yang<sup>6</sup>; S. Hochstadt<sup>1,2</sup>; K. Whitehouse<sup>1,2</sup>; D. Rynkowski<sup>1,2</sup>; D. Hu<sup>7</sup>; S. Huntsma<sup>3</sup>; F. Gilliland<sup>8</sup>; W.J. Gauderman<sup>8</sup>; R. Kumar<sup>9</sup>; F.D. Martinez<sup>10,11</sup>; D.J. Erle<sup>3</sup>; H. Hakonarson<sup>4,12</sup>; E.G. Burchard<sup>3,7</sup>; L.K. Williams<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Individualized and Genomic Medicine Research, Henry Ford Health System, Detroit, MI; 2) Department of Internal Medicine, Henry Ford Health System, Detroit, MI; 3) Department of Medicine, University of California San Francisco, San Francisco, CA; 4) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Public Health Sciences, Henry Ford Health System, Detroit, MI; 6) School of Nursing, University of Michigan, Ann Arbor, MI; 7) Department of Bioengineering & Therapeutic Sciences, University of California San Francisco, San Francisco, CA; 8) Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; 9) Ann & Robert H. Lurie Children's Hospital of Chicago, Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL; 10) Asthma and Airway Disease Research Center, University of Arizona, Tucson, AZ; 11) Department of Pediatrics, University of Arizona College of Medicine, Tucson, AZ; 12) Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, PA

Asthma is an inflammatory condition involving the adaptive immune system. Human leukocyte antigen (HLA) genes play a central role in inflammatory responses and in the recognition of self and foreign antigens. This region has not been extensively studied in African Americans (AA). This analysis included 4,317 AA participants (3,444 with and 873 without an asthma diagnosis) from the Study of Asthma Phenotypes and Pharmacogenomic Interactions by Race-ethnicity (SAPPHIRE) cohort. Whole genome sequence (WGS) were generated as part of the NHLBI TOPMed Program and the Asthma Translational Genomics Collaborative (ATGC). RNA-seq data were also available for 408 asthma cases and 405 controls. Single nucleotide variations (SNV) within extended major histocompatibility complex (MHC) region (chr6:27~34Mb, hg38) were generated using the WGS data, and high resolution HLA alleles were derived from both the WGS and RNA-seq data. HLA gene expression abundance was estimated from RNA-seq data. Amino acids for each MHC protein were translated from 4-digit HLA alleles. We assessed for associations between asthma and variants defined by alleles, SNVs, and amino acids. We performed an expression quantitative trait locus (eQTL) analysis for those variants significantly associated with asthma. A few SNVs were found with  $P$ -value below  $10^{-5}$  including the top one rs12215482 (OR=0.79,  $P=8.9 \times 10^{-6}$ ). We also examined two SNVs (rs9469220 and rs9273349) previously associated with IgE levels and asthma. Only rs9273349 was nominally associated (OR=1.13;  $P=0.05$ ) with asthma in our data. But both two candidate SNVs were significant eQTL (FDR  $<1.0 \times 10^{-5}$ ) for HLA-DQ genes HLA-DQB1, HLA-DQA2 and HLA-DQB2. Top associations from HLA alleles and amino acids included HLA-DQB1\*05 and a histidine at DQB1 residue 30 ( $P < 0.001$ ). HLA-DQB1 was also differentially expressed when comparing asthma cases and controls (FDR  $< 0.05$ ). In summary, this study replicated a previously described variant rs9273349 in the HLA region and verified that it was associated with HLA-DQ expression. We also identified an amino acid residue associated with asthma status in HLA-DQB1, a gene found to be differentially expressed by

asthma status. These findings require additional validation in other ATGC cohorts, as well as studies to understand the functional consequence on amino acid substitutions at position 30 in HLA-DQB1.



## **PgmNr 397: Development of a 36-gene expression panel using Nanostring nCounter to profile type I interferon signature in juvenile myositis potentially for disease stratification and guiding treatment.**

### **Authors:**

C. Kao <sup>1</sup>; A. Thomas <sup>1</sup>; F. Salley <sup>1</sup>; R. Pellegrino <sup>1</sup>; F.-X. Wang <sup>1</sup>; H. Hakonarson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Some autoimmune diseases are notable for displaying prominent signs of type I interferon (IFN) activation. Originally, this type I IFN 'signature' was definitively characterized and confirmed in systemic lupus erythematosus based on whole transcriptome profiling of patient blood, where ~300-400 IFN responsive genes were found to be markedly upregulated. Subsequently, this signature was also observed in inflammatory myopathies (IM), including dermatomyositis and polymyositis (although only minimally so in inclusion body myositis). The signature in IM is especially prominent in skeletal muscle, which is usually (and by definition) the primary site of pathology; however, it is also observable in blood. Importantly, there is a general correlation between the magnitude of the signature with disease activity (although there is heterogeneity), and therapies targeting type I IFN signaling are either in development (e.g. anti-IFN- $\alpha/\beta$  or IFNAR antibodies) or already approved (e.g. JAK inhibitors), which show promise as treatment in IM. Thus, the development of a clinical assay measuring type I IFN signature could be useful for disease stratification, monitoring disease activity, and/or as a companion diagnostic for drugs targeting type I IFN signaling.

We use the Nanostring nCounter platform as the basis for development of a pilot type I IFN signature assay. Genes were selected by literature review from previous transcriptome studies in blood in IM; in general, those that were the most consistently and highly expressed, especially in juvenile myositis, were chosen. Our initial panel contains 32 type I IFN-inducible genes, along with 4 housekeeping genes. We present results here from 19 subjects with juvenile myositis that were in remission or had mild to moderate disease activity at time of collection along with 5 control subjects (without apparent inflammatory or autoimmune diseases). Blood from all participants were collected in PaxGene tubes, then extracted for RNA using the Promega Maxwell instrument. Oligos for nCounter Elements chemistry were designed by Nanostring and synthesized by IDT.

We observed that the magnitude of the type I IFN signature tended to correlate with disease activity. Notably, 3 subjects experiencing significant morbidity who are on multiple immunomodulatory drugs had among the highest IFN signatures observed. Translational implications of these findings as well as potential areas for refinement of the assay are discussed.

## PgmNr 398: Rare germline variant contributions to myeloid malignancy susceptibility.

### Authors:

S. Li <sup>1</sup>; J. Wang <sup>1</sup>; R. Wei <sup>1</sup>; R. Shi <sup>1</sup>; V. Adema <sup>2</sup>; Y. Nagata <sup>2</sup>; C. Kerr <sup>2</sup>; T. Kuzmanovic <sup>2</sup>; P. Bartlomiej <sup>2</sup>; F. Sole <sup>3</sup>; J. Maciejewski <sup>2</sup>; T. LaFramboise <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Genome Sciences, Case Western Reserve Univ, Cleveland, Ohio.; 2) 2Department of Translational Hematology & Oncology Research, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio; 3) 3Myelodysplastic Syndrome Research Group, Josep Carreras Leukaemia Research Institute, Institut Català d'Oncologia-Hospital Germans Trias i Pujol, Universitat Autònoma de Barcelona, Badalona, Spain

---

Myeloid malignancy susceptibility alleles reported to date have largely been rare variants identified in families. With few exceptions, common variant associations have been lacking, and it is likely that as-yet undiscovered susceptibility variants are rare alleles. We analyzed whole-exome sequence from 690 individuals with myeloid malignancies to elucidate the contribution of rare (<1% population allele frequency) protein-altering alleles to disease susceptibility. We estimate that 3-4% of patients have rare variants in genes that would classify the patients as having germline predisposition under current guidelines. We also identify *MPO* as a new candidate predisposition gene, with 19 patients (2.8%) harboring rare pathogenic *MPO* variants, and significantly higher pathogenic allele frequencies in cases as compared to population controls. An elevated rare-variant burden is observed in autosomal recessive genes, particularly Fanconi anemia (FA) genes ( $P < 0.002$ ). Even accounting for high levels of rare variants, pairs of such variants in the same patient and FA gene (presumably affecting both parental homologs) occur at a rate significantly higher than would be expected by chance ( $P < 0.002$ ), suggesting that biparental inheritance of rare germline alleles in FA genes is a risk factor for myeloid malignancy. The FA gene *BRCA2* is particularly enriched for biparental inheritance, and rare germline variants in *BRCA2* are associated with poor overall survival in myeloid malignancy patients (age-adjusted HR 2.62, 95% CI 1.26-5.44).

## **PgmNr 399: The beliefs and values of patients with hemophilia and their caregivers about gene therapy and gene editing.**

### **Authors:**

T. Vasquez <sup>1</sup>; T. Lucas <sup>2</sup>; J. Harris-Wai <sup>1</sup>; D. Bowen <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) University of Washington, Seattle, WA.; 2) University of California, San Francisco

---

**Introduction:** Hemophilia is a lifelong debilitating inherited disease that relies on factor replacement therapy to prevent bleeding. However it is not curative and there is a negative impact on emotional health and quality of life. Gene therapy and gene editing are potential curative therapies that might address current unmet needs. However, we need to know the patient's beliefs and values about these therapies before they are implemented.

**Methods:** This qualitative study uses semi structured interviews about beliefs of gene therapy/editing in patients with hemophilia A or B and their caregivers. Grounded theory was used to analyze the data.

**Results:** About gene therapy, values differ among participants. Even it is a potential curative therapy, they expect to have clearer data about efficacy, however small sample size in trials is a barrier. They expect an improvement in the factor levels and in the duration of the effect, so there is a lower risk of bleeding and factor infusions are needed only for trauma events or surgery. They expect that the decrease in their risk of bleeding will have a positive impact in their mental health and in their quality of life. Given that these therapies might be expensive, they expect that insurance companies are willing to cover for it. The fact there is a prior history of HIV and HVC blood born infections in this community makes them consider the "safety net". Their concern about liver cancer relies on the DNA change and the use of viral vectors. To accept gene therapy, they would like to know about successful cases in the short and long term. They are less optimistic about gene editing. There is a moral concern. They think that DNA defines them as humans and changing it during early life is "scary". They won't give it to their preborn child and they would prefer letting their affected child make the decision as an adult. There is a concern about safety, because CRISPR will affect other regions of DNA that might lead to cancer, autism, Down syndrome, developmental delay, etc. Finally, knowing that this therapy is delivered during pregnancy, they are concerned about the high risk of miscarriage and losing their baby.

**Conclusion:** Preliminary results indicate patients and caregivers would accept gene therapy if its efficacy and long-term safety is known. Patients are more optimistic about gene therapy than their caregivers. Overall, they are more reluctant to accept gene editing.

## PgmNr 400: Developing a genetic risk index for peanut allergy.

### Authors:

D. Daley <sup>1,2</sup>; S. Lu <sup>1</sup>; Y. Yu <sup>1</sup>; A. Sandford <sup>1,2</sup>; Y. Asai <sup>3</sup>; A.E. Clarke <sup>4</sup>; A. Becker <sup>5</sup>; E.S. Chan <sup>6</sup>; A. Eslami <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Respiratory Medicine, University British Columbia, Vancouver, British Columbia, Canada; 2) Centre for Heart Lung Innovation, St. Paul's Hospital, Vancouver, British Columbia, Canada; 3) Division of Dermatology, Department of Medicine Queen's University, Kingston, Canada; 4) Division of Rheumatology, Department of Medicine, Cumming School of Medicine, University of Calgary, Alberta, Canada; 5) Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Canada; 6) Department of Pediatrics, University of British Columbia, Vancouver, BC

Over 200 single nucleotide polymorphisms (SNPs) have been found to be associated with food allergy (FA) in genome-wide association studies (GWAS). A genetic risk score (GRS), can be derived from GWAS to summarize genetic risk, risk stratification and/or prediction. Our objective was to use information from the Canadian Peanut Allergy Registry (CanPAR) GWAS study to develop a GRS using the weighted sum of the number of risk alleles (with values 0/1/2) by the natural log of their respective odds ratio (OR).

The positive predictive value of the GRS was then evaluated in an independent cohort, the Canadian Asthma Primary Prevention Study (CAPPS). The CAPPS study was initiated in 1995 and recruited 549 children at high risk for developing asthma, defined as 1 first degree (fd) relative with asthma or 2 fd with allergic disease, mothers were recruited during the second and third trimester of pregnancy, and the children have been followed until 15 years of age. Food allergy cases at the 15 year time point were determined by a pediatric allergist and controls were defined by the absence of food allergy, asthma, atopy and airway hyperresponsiveness at all time-points (12 and 24 months, 7 and 15 years of age). The CAPPS study has contributed to several asthma GWAS studies and genotypes and principal components (PCs) were extracted from the existing GWAS study.

### Methods

Models were fit with and without PC (10 for CanPAR and 5 for CAPPS). Using a p-value threshold of 1.49E-06 and LD ( $r^2 > .80$ ) pruning we identified 25 independent CanPAR SNPs for use in the GRS, of these 13 SNPs were identified (either directly genotyped, imputed, or a proxy SNP in high LD  $> .80$ ) in the CAPPS study. We then evaluated the area under the curve (AUC) which is used to determine the effectiveness of the classification and the positive predictive value (PPV).

### Results

Table 1: Summary of GRS risk models

p-value threshold	Study	Genotyped	Imputed	Total	AUC (95% CI)	PPV
1.49E-06	CanPAR	6	19	25	0.653 (0.628-0.678)	0.59

1.49E-06*	CanPAR	6	19	25	0.878 (0.862-0.894)	0.83
1.49E-06	CanPAR	5	8	13	0.638 (0.613-0.664)	0.61
1.49E-06*	CanPAR	5	8	13	0.878 (0.862-0.894)	0.84
1.49E-06	CAPPS	10	3	13	0.579 (0.404-0.754)	0.67

## Conclusions

Only minimal differences were observed between models with 13 vs. 25 SNPs and model fit was greatly improved when including PCs. Preliminary findings indicate excellent replication of the model in the CAPPS study using hyper-controls. Next steps include evaluating model fit when using asthmatic and atopic controls.

## PgmNr 401: Association between clonal hematopoiesis and non-relapse mortality after autologous hematopoietic cell transplantation.

### Authors:

T.P. Slavin<sup>1,2,3</sup>; J.B. Teh<sup>3</sup>; J.N. Weitzel<sup>1,2,3</sup>; K. Peng<sup>3</sup>; F.L. Wong<sup>3</sup>; H. Qin<sup>4</sup>; J. Wang<sup>4</sup>; X. Wu<sup>4</sup>; M. Mei<sup>5</sup>; R. Pilai<sup>6</sup>; Y. Wang<sup>6</sup>; K.W.K. Tsang<sup>1</sup>; A. Pozhitkov<sup>7</sup>; A. Krishnan<sup>5</sup>; S.J. Forman<sup>5</sup>; S.H. Armenian<sup>3,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Clinical Cancer Genomics, City of Hope, Duarte, CA; 2) Department of Medical Oncology and Therapeutics Research, City of Hope, Duarte, CA; 3) Department of Population Sciences, City of Hope, Duarte, CA; 4) Integrative Genomics Core, City of Hope, Duarte, CA; 5) Department of Hematology & Hematopoietic Cell Transplantation, City of Hope, Duarte, CA; 6) Molecular Pathology Core, City of Hope, Duarte, CA; 7) Division of Research Informatics, City of Hope, Duarte, CA; 8) Department of Pediatrics, City of Hope, Duarte, CA

**BACKGROUND:** Clonal hematopoiesis (CH) is an age-related condition that is characterized by somatic mutations in the blood cells of otherwise healthy individuals. In non-cancer populations, CH is associated with a higher risk of aging-related diseases such as acute myeloid leukemia (AML) and cardiovascular disease (CVD) and may be a biomarker of biologic aging. Patients who undergo hematopoietic cell transplantation (HCT) are at high risk of premature onset of aging-related conditions (subsequent neoplasms [SN], CVD) after HCT. There is a paucity of information on the impact of pre-treatment CH on aging-related conditions after HCT, including non-relapse related mortality.

**METHODS:** We used a case-control study design, and identified 10 patients who died due to non-relapse mortality (NRM; e.g., SN, CVD, organ failure) after autologous HCT at City of Hope (cases). Cases were matched (1:3) to controls (underwent HCT, alive or died of relapse-related mortality, N=29) on age at HCT  $\pm 2$  years, sex, diagnosis, and length of follow-up after HCT. We evaluated mobilized hematopoietic stem cell DNA, collected prior to HCT, using a custom 79-gene-myeloid-CH-coding-exon-amplicon-based QIAseq panel. Pathogenic and likely pathogenic CH variants (PV) above an allele fraction of 2% were used for analyses.

**RESULTS:** Patient median age at HCT was 64 years (range: 38 to 71 years), 79.5% were male, 61.6% underwent HCT for non-Hodgkin lymphoma, 17.9% for Hodgkin lymphoma, 20.5% for multiple myeloma. Among cases, causes of death were: SN (60%), CVD (10%), respiratory failure (10%), and other (20%). Due to matching, there were no differences in age, sex, or diagnosis among cases and controls. Cases were significantly more likely to have CH compared to controls (70% vs. 24.1%,  $p=0.009$ ). Cases were also significantly more likely to have multiple ( $\geq 2$ ) unique PVs (e.g. *DNMT3* and *TET2*) compared to controls (60% vs. 6.9%,  $p<0.001$ ) and significantly more likely to have PVs with high ( $\geq 10\%$ ) allelic fraction (40% vs. 3.4%,  $p=0.003$ ).

**CONCLUSION:** Our study provides preliminary evidence of an association between pre-HCT CH and adverse outcomes after HCT, irrespective of chronologic age. Integration of CH analyses into existing pre-HCT risk-prediction models may improve the accuracy of these models, setting the stage for the development of personalized risk assessment strategies and targeted treatments that optimally prevent or manage complications associated with HCT.

## PgmNr 402: Candidate gene study of adverse liver effects of methotrexate in patients with rheumatoid arthritis.

### Authors:

M. Wadelius<sup>1</sup>; J. Karlsson Sundbaum<sup>1,2</sup>; E. Baecklund<sup>1</sup>; N. Eriksson<sup>1,3</sup>; H. Kohnke<sup>1</sup>; P. Hallberg<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Sciences and Science for Life Laboratory, Uppsala University, Sweden; 2) Department of Health Sciences, Luleå University of Technology, Luleå, Sweden; 3) Uppsala Clinical Research center, Uppsala, Sweden

---

**Background:** Low-dose methotrexate (MTX) is the anchor drug for the treatment of rheumatoid arthritis (RA), either in monotherapy or in combination with other drugs. MTX has proven effects on both inflammatory symptoms and signs of radiographic destruction, and early initiation improves treatment outcome. However, adverse effects on for example the liver may lead to discontinuation of MTX therapy. It would be advantageous to be able to identify patients that risk elevation of liver enzymes in order to modify treatment.

**Objective:** To investigate whether genetic variants of the candidate genes *MTHFR*, *TYMS*, *SLCO1B1* and *PNPLA3* are associated with elevation of the liver enzyme alanine aminotransferase (ALT).

**Method:** RA patients starting MTX treatment at the Rheumatology clinic, Uppsala University Hospital, Sweden in 2005-2013 were included. Clinical and laboratory data from the onset of RA until MTX was stopped or to the end of the study period were attained from medical records and telephone interviews. A blood sample was obtained, and genotyping was performed with PCR-based methods. Linear regression was used for statistical analysis of the continuous variable log transformed maximum ALT. Logistic regression was used for statistical analysis of the binary variable elevation of ALT to at least 1.5 times the upper limit of normal (ULN).

**Results:** The study comprised 207 patients with RA: 138 females (67 %) and 69 males (33%). Elevation of ALT to  $\geq 1.5 \times \text{ULN}$  was observed in 20% of the patients. *MTHFR* 677T was significantly associated with a 9.8% decrease in maximum ALT per minor allele of the variant (beta -0.0970, 95% CI -0.179, -0.0165,  $p = 0.0193$ ). No genetic variant of *MTHFR*, *TYMS*, *SLCO1B1* or *PNPLA3* was significantly associated when patients with ALT  $\geq 1.5 \times \text{ULN}$  vs.  $< 1.5 \times \text{ULN}$  were compared.

**Conclusion:** The effect of MTX is largely explained by its interactions with enzymes in the folate pathway. *MTHFR* is one of the key enzymes in the folate pathway. It is therefore of interest to study inherited variants of *MTHFR* in relation to MTX-induced toxicity. According to our results, the *MTHFR* 677T allele appears to protect against an elevation of ALT in patients with RA. We aim to verify our findings in a larger study, and to use clinical and genetic factors to develop a prediction model for elevation of ALT during MTX treatment. A robust prediction model could be a useful tool, and would be a step towards precision medicine in the field of rheumatology.

## PgmNr 403: Efficient gene editing in Rett syndrome and related disorders.

### Authors:

I. Meloni <sup>1</sup>; S. Croci <sup>1</sup>; K. Capitani <sup>1,2</sup>; M.L. Carriero <sup>1</sup>; S. Daga <sup>1</sup>; F. Donati <sup>2</sup>; F.T. Papa <sup>1</sup>; E. Frullanti <sup>1</sup>; D. Lopergolo <sup>1,3</sup>; V. Lamacchia <sup>1,3</sup>; R. Tita <sup>3</sup>; A. Giliberti <sup>1</sup>; E. Benetti <sup>1,4</sup>; C. Lo Rizzo <sup>3</sup>; S. Furini <sup>4</sup>; S.G. Conticello <sup>2</sup>; A. Renieri <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, University of Siena, Siena, Italy; 2) Molecular Mechanisms of Oncogenesis, ISPRO Core Research Laboratory (CRL), Firenze, Italy; 3) Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 4) Department of Medical Biotechnologies, University of Siena, Siena, Italy

---

Mutations in MECP2 and FOXP1 genes respectively cause the classic form and the congenital variant of Rett syndrome, one of the most common genetic causes of intellectual disability in girls. Both genes are transcriptional regulators and both under- and over-expression cause disease in humans. We thus reasoned that adding a gene under a not native regulator would be a risky therapeutic strategy while gene editing would be much more effective. We present here the successful application of CRISPR/Cas9 gene editing technology in patient-specific human cellular models, namely fibroblasts, induced Pluripotent Stem Cells (iPSCs) and iPSC-derived neurons. We have engineered a two-plasmid system to correct FOXP1 (c.688C>T-p.Arg230Cys) and MECP2 (c.473C>T-p.Thr158Met) mutations. Mutation-specific sgRNAs and donor DNAs have been selected and cloned together with an mCherry/GFP reporter system. Cas9 flanked by sgRNA recognition sequences for auto-cleaving has been cloned in a second plasmid. The system has been designed to be ready for in vivo delivery via Adeno-Associated Viral (AAV) vectors. We demonstrated that different serotypes have different efficiency in different target cells, the best being either AAV9 in fibroblasts and iPSC-derived neurons or AAV2 in iPSCs. mCherry + /EGFP + fibroblasts and neurons isolated by Fluorescent Activated Cell Sorting and analyzed by Next Generation Sequencing have shown efficient gene editing via Homology-Directed Repair for both genes, with up to 80% of mutated alleles correctly reverted to the wild type sequence, outlining the relevant potentiality of the approach for Rett syndrome therapy.



## **PgmNr 404: Development of targeted genetic therapeutics and cellular models for Smith-Magenis syndrome: A proof of concept for the therapy of haploinsufficiency-based disorders.**

### **Authors:**

L. Castan Chibas <sup>1</sup>; S. Sridhar <sup>1</sup>; T. Paretkar <sup>1</sup>; M. Ramirez <sup>1</sup>; N. Yew <sup>2</sup>; T. Ezashi <sup>3</sup>; S. Elsea <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Sanofi Genzyme.; 3) Bond Life Sciences Center and Division of Animal Sciences, University of Missouri, Columbia, MO.

---

Haploinsufficiency is the primary mechanism for several genetic conditions like Smith-Magenis syndrome (SMS) and MBD5-associated neurodevelopmental disorder. In SMS, the condition is caused by either heterozygous deletion or mutation of the dosage-sensitive gene *RAI1*, proven to be the main cause of the disease. Some features of this condition include intellectual disability, early onset obesity, sleep disorders, self-injurious behaviors, and craniofacial anomalies. Currently, there is not an effective treatment for SMS or any other disorder of haploinsufficiency, and patients receive only supportive therapy. Our targeted therapeutic strategies are based in restoring the function of *RAI1* by either providing the missing copy of the gene or increasing the expression of the remaining functional allele. In the first approach, SMS patient fibroblasts were transduced in a controlled manner with lentiviral vectors containing the sequence of *RAI1*. Our results revealed that the protein *RAI1* is properly localized in the nucleus, and the expression can be increased proportionally by controlling the multiplicity of infection (MOI) to restore normal levels of *RAI1*. The second approach is based on a novel methodology using a deactivated version of the protein Cas9 fused with the viral activator VP64 (CRISPR-dCas9-VP64). NIH3T3 fibroblasts were transduced with several constructs containing different guide (g)RNAs targeting the murine *Rai1* promoter region. Our results reveal that specific constructs are able to increase *Rai1* expression, which subsequently increases the expression of the *RAI1*-regulated downstream gene *Bdnf*, which is typically reduced in expression in SMS. Finally, we developed patient-derived induced-pluripotent stem cells (iPSC) using episomal plasmids and generated neural progenitor cells (NPC) derived from these iPSC using the monolayer protocol. Immunostaining revealed that the pluripotency marker OCT4 is present in iPSC but neural differentiation marker PAX6 is absent, and upon neural induction, OCT4 is not expressed, while PAX6 is present. The combination of our targeted therapeutic strategies with the developed iPSC-NPC models will provide valuable tools with potential applications in the development of therapies for gene haploinsufficiency-based disorders.

## PgmNr 405: Specification of ACMG/AMP guidelines for standardized variant interpretation in familial hypercholesterolemia: On behalf of the ClinGen FH Variant Curation Expert Panel.

### Authors:

M.A. Iacocca<sup>1</sup>; J.R. Chora<sup>2</sup>; T. Freiburger<sup>3</sup>; A. Carrie<sup>4</sup>; C.L. Kurtz<sup>5</sup>; S.E. Leigh<sup>6</sup>; J.C. Defesche<sup>7</sup>; L. Tichy<sup>8</sup>; A.C. Alves<sup>2</sup>; H. Wand<sup>1</sup>; M. Williams<sup>9</sup>; E.J. Sijbrands<sup>10</sup>; R.A. Hegele<sup>11</sup>; J.W. Knowles<sup>1,12</sup>; M. Bourbon<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanford University, School of Medicine, Stanford, CA; 2) Instituto Nacional de Saude Doutor Ricardo Jorge, Lisbon, Portugal; 3) Centre for Cardiovascular Surgery and Transplantation, and Medical Faculty, Masaryk University, Brno, Czech Republic; 4) Sorbonne Universite, Hopital Pitie-Salpetriere, Paris, France; 5) University of North Carolina, Chapel Hill, NC; 6) Genomics England, London, United Kingdom; 7) Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands; 8) Centre of Molecular Biology and Gene Therapy, University Hospital Brno, Brno, Czech Republic; 9) Bristol Genetics Laboratory, National Health Service, Bristol, United Kingdom; 10) Erasmus Medical Center, Rotterdam, Netherlands; 11) Robarts Research Institute, Western University, London ON, Canada; 12) FH Foundation, Pasadena, CA

**Background:** Familial hypercholesterolemia (FH) is a heritable disorder characterized by elevated low-density lipoprotein (LDL) cholesterol levels and premature cardiovascular disease. Increasingly, patients with a clinical suspicion of FH are being offered genetic testing as part of diagnosis. However, determining pathogenicity of identified DNA variants remains a major challenge; this process may be rudimentary and often differs among laboratories. To improve accuracy, concordance and standardization, there has been increasing effort to implement the recent variant interpretation guidelines proposed by the American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG/AMP).

**Objectives:** To further improve standardized variant interpretation, the Clinical Genome Resource (ClinGen) FH Variant Curation Expert Panel (VCEP) has been tasked with optimizing the ACMG/AMP guidelines to become disease-specific for FH.

**Methods:** Following review of ACMG/AMP criteria, the FH VCEP met through frequent emails and conference calls to propose modifications. After multiple preliminary iterations, pilot testing to identify areas for additional refinement, debate and further commentary, consensus was reached.

**Results:** Here, we propose a consensus set of FH-ACMG/AMP guidelines, which focus first on the LDL receptor gene (*LDLR*), and include: 1) functional study criteria specifications and split-weighting for three 'levels' of relevant assays; 2) alteration of control population data frequency thresholds; 3) specific use and thresholds for *in silico* prediction tools; 4) additional specifications for copy number variants; and 5) co-segregation criteria specifications and split-weighting for large versus small families. Following ClinGen approval, these guidelines will be tested in a finalized pilot study of 50 variants to demonstrate improved variant classification compared to the original ACMG/AMP criteria.

**Conclusions:** Establishment of these specified guidelines will help to achieve a more evidence-based, standardized method for the classification of variants detected in FH patients worldwide. Ultimately, all ~3000 FH-associated variants currently deposited in the ClinVar database will be re-classified

using the FH-ACMG/AMP criteria, while promoting these guidelines as the new “gold standard” for variant classification among the FH community will ensure novel variants identified in real-time are subject to an appropriate standard of evaluation.

## **PgmNr 406: Pathogenic loss of function in genes associated to monogenic pediatric disorders in healthy elderly Brazilians improves molecular diagnosis interpretation.**

### **Authors:**

M.S. Naslavsky <sup>1</sup>; G.L. Yamamoto <sup>1,2</sup>; M.O. Scliar <sup>1</sup>; J. Wang <sup>1</sup>; Y.A.O. Duarte <sup>3</sup>; M.R. Passos-Bueno <sup>1</sup>; M. Zatz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Human Genome and Stem Cell Research Center, Department of Genetics and Evolutionary Biology, Biosciences Institute, University of Sao Paulo, São Paulo, Brazil; 2) 2-Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, Boston Children's Hospital and Harvard Medical School, Boston, MA 02115, USA; 3) Department of Epidemiology, Public Health School, University of Sao Paulo, São Paulo, Brazil

---

Large sequencing datasets, including whole exomes and genomes, are an important resource to assess allelic frequency of rare variation and improve pathogenicity interpretation in molecular diagnosis pipelines. Several studies have demonstrated that unaffected individuals may harbor potentially deleterious variants in clinically relevant genes. To further assess the incidence of putative loss of function previously assigned as pathogenic or likely pathogenic variants, a census-based cohort of 1172 unrelated Brazilian individuals aged 60 or over was analyzed in search for variants in genes related to childhood disorders with Mendelian inheritance. We have filtered 924 genes associated to monogenic pediatric disorders in a whole genome sequencing dataset. Initial filtering yielded over 4 thousand putative loss of function variants in 805 genes with at least a singleton passing depth and allele balance quality criteria. Removal of putative frameshifts caused by insertions/deletions and further filtering by an allelic frequency cutoff of 1% in both cohort and gnomAD, pathogenic assertions on ClinVar and monoallelic mode of inheritance provided a high confidence list of 32 single nucleotide substitutions in 21 genes with splicing, stop gain and stop loss predicted consequences. Forty individuals carry these variants (3.4% of the cohort) with individual frequency ranging from absent to 0.14% on gnomAD. No individual was found to be homozygous following the above-mentioned criteria of variant filtering. These results are compatible with other studies, but it is noteworthy that an elderly cohort may be biased towards health fitness and overall survival. Besides potential overestimation of pathogenicity classifications, true effects of these variants might be modulated by complex gene-gene interactions resulting in reduced penetrance. In fact, half of the carriers are highly admixed with less than 70% of any single ancestry. Further investigation on biallelic disorders and occurrence of compound heterozygosity is currently ongoing. Genomic datasets of individuals without prior health complaints are becoming common both within academic and private endeavors. Pathogenicity classification of incidental findings is challenging, therefore describing the incidence of potentially damaging variants in population-based datasets is essential. Furthermore, healthy elderly cohorts can play a role as a tool for filtering candidate variants as causes to Mendelian disorders.

## PgmNr 407: Genomic survival score percentiles informed by diseases make 15-year distinctions in lifespan.

### Authors:

P.R.H.J. Timmers<sup>1</sup>; N. Mounier<sup>2,3</sup>; K. Läll<sup>4,5</sup>; R. Mägi<sup>4,5</sup>; K. Fischer<sup>4,5</sup>; T. Esko<sup>4,7</sup>; J.F. Wilson<sup>1,6</sup>; Z. Kutalik<sup>2,3</sup>; P.K. Joshi<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) College of Medicine and Veterinary Medicine, Usher Institute for Population Health Sciences and Informatics, Edinburgh, United Kingdom; 2) University Center for Primary Care and Public Health, Lausanne, Switzerland; 3) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 4) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia; 5) Institute of Mathematics and Statistics, University of Tartu, Tartu, Estonia; 6) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom; 7) Broad Institute of Harvard and MIT, Cambridge, MA, USA

---

**Motivation:** Prediction of lifespan using genetic variants discovered by genome-wide association studies is challenging due to low power: a recent study of >1 million parental lifespans only revealed 12 genome-wide significant loci. However, many of the age-related diseases and risk factors underlying mortality each have genetic components that have been studied using large sample sizes, with many more loci identified. Carrying such variants has been shown to influence lifespan and we hypothesise that using this information can improve lifespan predictions.

**Method:** Here, we generate and test polygenic risk scores using published summary statistics for 1 million parental lifespans from UK Biobank and other European cohorts. We then assess to what extent such predictions can be improved by incorporating LD information, functional annotations, and genetic data on 16 independent diseases and disease risk factors using a Bayesian framework (bPRS).

**Results:** One standard deviation increase in bPRS (out-of-sample) associates with an increase of 1.82 (95% CI 1.63–2.01) years of life, improving upon conventional polygenic scores by more than 30%. When stratified into bPRS percentiles, parents in the top percentile live 15 years longer on average than those in the bottom percentile. The same individuals show a 30–80% decrease in the incidence of multiple age-related diseases and are 6 times as likely to survive past 90 years of age. We replicate these findings in the Estonian Biobank (N = 150,000), where mortality data are available for parents as well as subjects themselves.

**Conclusions:** Together, our findings suggest bPRS is able to make greater distinctions in long-term health and survival than conventional polygenic scores by borrowing strength from the genetic associations of multiple traits.

## PgmNr 408: A polygenic risk score to improve screening for fracture risk.

### Authors:

J. Keller-Baruch<sup>1,2</sup>; V. Forgetta<sup>1</sup>; M. Forest<sup>1</sup>; A. Durand<sup>3</sup>; S. Bhatnagar<sup>1</sup>; J. Kemp<sup>4,5</sup>; M. Nethander<sup>6,7</sup>; D. Evans<sup>8</sup>; J.A. Morris<sup>1</sup>; D.P. Kiel<sup>9</sup>; F. Rivadeneira<sup>10</sup>; H. Johansson<sup>11</sup>; N.C. Harvey<sup>12,13</sup>; D. Mellström<sup>7</sup>; M. Karlsson<sup>14</sup>; C. Cooper<sup>12,13,15</sup>; D.M. Evans<sup>4,5</sup>; R. Clarke<sup>16</sup>; J.A. Kanis<sup>11</sup>; E. Orwoll<sup>17,18</sup>; E.V. McCloskey<sup>19</sup>; C. Ohlsson<sup>7</sup>; J. Pineau<sup>3</sup>; W.D. Leslie<sup>20</sup>; C.M.T. Greenwood<sup>1,2,21,22</sup>; J.B. Richards<sup>1,2,23</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Clinical Epidemiology, Department of Medicine, Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Québec, Canada; 2) Department of Human Genetics, McGill University, Montréal, Québec, Canada; 3) School of Computer Science, McGill University, Montréal, Québec, Canada; 4) University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland, Australia; 5) MRC Integrative Epidemiology Unit, University of Bristol, UK; 6) Bioinformatics Core Facility, Sahlgrenska Academy, University of Gothenburg, Sweden; 7) Centre for Bone and Arthritis Research, Department of Internal Medicine and Clinical Nutrition, Institute for Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 8) California Pacific Medical Center Research Institute, San Francisco, California, USA; 9) Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Department of Medicine, Beth Israel Deaconess Medical Center and Harvard Medical School, Broad Institute of MIT & Harvard University, Boston, USA; 10) Department of Internal Medicine, Erasmus Medical Center, Rotterdam, Netherlands; 11) Centre for Metabolic Bone Diseases, University of Sheffield, Sheffield, UK, and Australian Catholic University, Melbourne, Australia; 12) Medical Research Council Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK; 13) National Institute for Health Research Southampton Biomedical Research Centre, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK; 14) Department of Orthopaedics and Clinical Sciences, Lund University, Skane University Hospital, Malmö, Sweden; 15) National Institute for Health Research Oxford Biomedical Research Centre, University of Oxford, Oxford, UK; 16) Clinical Trial Service Unit and Epidemiological Studies Unit, University of Oxford, UK; 17) Bone & Mineral Unit, Oregon Health & Science University, Portland, Oregon, USA; 18) Department of Medicine, Oregon Health & Science University, Portland, Oregon, USA; 19) Mellanby Centre for Bone Research, Centre for Integrated Research in Musculoskeletal Ageing, University of Sheffield, and Sheffield Teaching Hospitals Foundation Trust, Sheffield, UK; 20) Department of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada; 21) Gerald Bronfman Department of Oncology, McGill University, Montréal, Québec, Canada; 22) Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montréal, Québec, Canada; 23) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom

---

**Importance:** Screening programs often include large proportions of individuals at low risk of disease outcomes; polygenic risk scores of heritable risk factors could improve screening efficiency by targeting assessments to individuals at higher genetic risk. If successful, such evidence could help establish a clinical role for polygenic risk scores.

**Objective:** To test whether a polygenic risk score for heel ultrasound speed of sound (SOS) - a

heritable risk factor for fractures that is correlated with bone mineral density (BMD) - can effectively target a fracture screening program to higher risk individuals.

**Design, Setting, Participants:** 341,449 individuals from the UK Biobank were used to develop 600 polygenic risk score models for SOS with LASSO regression, with the optimal prediction model determined in a separate set of 5,335 individuals. The utility of this model in fracture risk screening was assessed on a subset of 5 validation cohorts (N=10,522) according to National Osteoporosis Guideline Group (NOGG) guidelines.

**Main Outcomes and Measures:** Performance metrics were the sensitivity and specificity to correctly identify individuals requiring treatment, with BMD-based FRAX probabilities as the reference-standard. Secondary outcomes included the proportion of the screened population requiring a clinical risk factor-based FRAX test (without BMD) and the proportion of individuals requiring a BMD-based FRAX test.

**Results:** Polygenic risk scores for SOS correlated with measured SOS ( $R^2=23.2\%$  in a testing set of 84,768 individuals). Without genetic pre-screening, guideline recommendations resulted in a high sensitivity and specificity for correct treatment assignment (99.6% and 97.1%, respectively) in the validation cohorts. However, 81% of the population required clinical risk factor-based FRAX assessments and 37% required BMD-based FRAX assessments. Limiting further assessment to individuals with a low polygenic risk score for SOS resulted in small changes to sensitivity and specificity (93.4% and 98.5%, respectively), while substantially reducing the proportions requiring clinical risk factor-based assessments and BMD-based FRAX assessments by 37% and 41%, respectively.

**Conclusions and Relevance:** The use of a polygenic risk score in fracture risk screening substantially decreases the number of individuals requiring detailed assessments, including BMD measurement, while maintaining a high sensitivity and specificity for correct treatment assignment.

## PgmNr 409: A clinical trial of intravenous 2-hydroxypropyl- $\beta$ -cyclodextrin for acute liver disease in infants with Niemann-Pick C.

### Authors:

M. Reynolds <sup>1</sup>; L. Linneman <sup>1</sup>; B. Warner <sup>1</sup>; Y. Turmelle <sup>1</sup>; X. Jiang <sup>1</sup>; M. Shinawi <sup>1</sup>; F. Porter <sup>2</sup>; D. Ory <sup>1</sup>; S. Cole <sup>1</sup>; P. Dickson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Washington University, Saint Louis, MO.; 2) Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD

---

Niemann-Pick C (NPC) is an autosomal recessive disease due to defective NPC1 or NPC2 proteins resulting in endo-lysosomal storage of unesterified cholesterol and lipids in both the central nervous system and the liver. Acute liver disease is often observed in the newborn period and may be self-limited or can be fatal. 2-hydroxypropyl- $\beta$ -cyclodextrin (2HPBCD) is a cholesterol-binding agent that reduces intracellular cholesterol accumulation in NPC, possibly by enhancing MCOLN-1 dependent secretion from the endo-lysosomal system. We are presently enrolling up to 15 infants < 6 months with direct hyperbilirubinemia due to NPC1 or NPC2 disease in clinical trial NCT03471143 , to determine whether intravenous 2HPBCD improves liver disease in NPC infants. Infants receive IV 2HPBCD twice a week for 6 weeks for a total of 12 doses, followed by monthly infusion during a 6-month extension study. Two patients have been enrolled in the study. The first patient is a 4 month old boy with NPC1 who presented with a direct bilirubin of 0.6 mg/dL, AST of 167 units/L, ALT of 94 units/L, and 5- $\alpha$ -cholan-3 $\beta$ , 5 $\alpha$ , 6 $\beta$ -triol N-(carboxymethyl)-amide ("bile acid B") of 57.2ng/mL. Direct bilirubin improved to 0.1 mg/dL two weeks after his first infusion and remained within normal limits after the 6 week infusion. AST and ALT remained elevated after 1 month at 114 units/L and 80 units/L, respectively. Bile acid B decreased to 10.5ng/mL. The second patient is a 2 month old boy who presented with a direct bilirubin of 5.7mg/dL, AST of 277 units/L, ALT of 124 units/L, and underwent his first infusion. The restoration of bile acid to within normal limits for Patient 1 suggests the drug is decreasing cholesterol storage. Additional follow-up as the patients undergo further infusions is planned.



## PgmNr 410: Systemic administration of AAV- *Slc25a46* relieves mitochondrial neuropathy of *Slc25a46* mutant mice.

### Authors:

L. Yang <sup>1,2</sup>; J. Slone <sup>1</sup>; Z. Li <sup>3</sup>; Y. Hu <sup>4</sup>; L. Queme <sup>5</sup>; M. Jankowski <sup>5</sup>; T. Huang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OHIO.; 2) Pediatrics of Xiangya Hospital, Central South University, Changsha, Hunan, 410008, China; 3) State Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan 410078, China; 4) Division of Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA; 5) Division of Anesthesia, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

---

Mitochondrial disorders could be the result of nuclear or mitochondrial DNA mutations and often affect multiple organ systems. Despite research advances in the molecular and cellular fields, the methods available for treating mitochondrial diseases are minimal. Advancements in gene therapy, however, may offer us a new opportunity to address this group of severe disorders. Mutations in the mitochondrial fusion-fission related gene *SLC25A46* cause peripheral neuropathy and optic atrophy in humans. Previously, we created a knockout mouse model that recapitulates the clinical phenotypes observed in humans. Here, we have treated *Slc25a46* mutant mice with AAV-PHP.B vector carrying the mouse *Slc25a46* gene via facial vein injection. Our results show that this AAV-PHP.B-*Slc25a46* vector is efficaciously transduced in the target organs as verified with a vector expressing eGFP. Consequently, we treated mutant mice with AAV-PHP.B-*Slc25a46* and found that treated animals have a significantly longer lifespan, improved weight gain, a reversal of central and peripheral neuropathy, and a restored mitochondrial function. Most notably, mitochondrial morphology was normalized. Overall, this study shows that gene therapy could be a practical approach to treating mitochondrial disorders caused by mutations in nuclear genes and that AAV-PHP.B is an excellent vector for delivering target genes to the central nervous system.

## PgmNr 411: Cell autonomous and cell-non-autonomous effects of NAGLU deficiency in Sanfilippo B syndrome.

### Authors:

I. Elsharkawi; S. Sukupolvi; S. Le; C. Mikulka; M. Sands; P. Dickson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pediatrics, Division of Genetics, Washington University in St. Louis, St. Louis, MO.

---

Cross-correction is the transfer of soluble lysosomal enzymes between neighboring cells, and its occurrence limits our ability to discriminate between cell-autonomous and non-cell autonomous effects of enzymatic activity. The lack of understanding of the cell types primarily involved in disease complicates our ability to study its pathogenesis and hinders the discovery of potential therapies. Sanfilippo B (MPS IIIB) is a lysosomal disease caused by deficient N-alpha acetylglucosaminidase (NAGLU) enzyme, characterized by progressive neurological deterioration in childhood. We aimed to study and differentiate cell-autonomous and non-cell-autonomous effects of NAGLU deficiency by generating a transmembrane form of NAGLU. In doing this, we aimed to create a NAGLU that would remain located within the lysosome, remaining enzymatically active while bound to the lysosomal membrane, in order to prevent cross correction. We generated an expression construct consisting of the cDNA encoding human NAGLU, a 6-glycine linker, and a C-terminal murine lysosome-associated membrane protein-1 (Lamp-1). We cloned this construct into a modified pmrROSA vector. We plan to clone this construct further into a lentiviral vector to transfer NAGLU-Lamp1 into MPS IIIB human fibroblasts. We plan to assess intracellular and extracellular NAGLU enzymatic activity to determine whether our transmembrane construct can express active protein that is retained in the lysosome. We subsequently plan to generate transgenic mice, employing the Cre-lox system to target specific cell lineages. The results and conclusion of this experiment will help us differentiate between cell-autonomous and non-cell-autonomous effects of NAGLU deficiency, with implications for future therapy targets.

## PgmNr 412: Enhancement of hepatic autophagy for therapy of argininosuccinic aciduria.

### Authors:

N. Brunetti-Pierri <sup>1,2</sup>; D.P. Perocheau <sup>3</sup>; A. De Angelis <sup>1</sup>; M. Orford <sup>3</sup>; S. Eaton <sup>3</sup>; P. Cuomo <sup>4</sup>; D. Paris <sup>4</sup>; A. Motta <sup>4</sup>; S. Waddington <sup>5</sup>; J. Baruteau <sup>3</sup>; L. Soria <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Telethon Inst Genetics & Medicine, Pozzuoli, Naples, Italy; 2) Department of Translational Medicine, Federico II University, Naples, Italy; 3) Great Ormond Street Institute of Child Health, University College London, London, UK; 4) Institute of Biomolecular Chemistry, National Research Council, Pozzuoli, Italy; 5) Institute of Women's Health, University College London, London, UK

---

Argininosuccinic aciduria (ASA) is a urea cycle disorder (UCD) caused by deficiency of argininosuccinate lyase (ASL). Clinical manifestations of ASA occur as a consequence of impaired ureagenesis and nitric oxide depletion. We recently found that enhancement of hepatic autophagy potentiates ureagenesis and increases *in vivo* ammonia detoxification. In the present study, we investigated the efficacy of autophagy enhancement for therapy of ASA.

ASL-deficient mice were intraperitoneally (i.p.) injected with Tat-Beclin-1 (TB-1), a cell penetrating autophagy inducing peptide. Following treatment, survival and growth were evaluated along with markers of autophagy activation in tissue homogenates by Western blot. Ureagenesis was measured by stable isotope in ASL-deficient mice receiving i.p. injection of <sup>15</sup>N-labeled NH<sub>4</sub>Cl. Whole-liver metabolome analysis was performed by <sup>1</sup>H-NMR.

Compared to vehicle-controls, survival was significantly increased in ASL-deficient mice treated with TB-1. Autophagic flux was enhanced in livers but not in brains of mice injected with TB-1 as shown by reduced LC3-II and p62 protein levels, and increased fragment p10 of BHMT. Consistent with our previous studies, autophagy activation was associated with increased incorporation of <sup>15</sup>N into urea, indicating enhanced ureagenesis. Liver metabolome profiling by <sup>1</sup>H-NMR showed that autophagy activation partially rescued metabolite imbalance of ASA. Notably, argininosuccinate and citrulline which are the biochemical hallmarks of ASA, were reduced in TB-1-treated mice and levels of key compounds such as, fumarate, succinate, and glucose were rescued by enhanced autophagy. In conclusion, our data show that activation of hepatic autophagy improves survival and ureagenesis and corrects the metabolic defects of ASL deficiency. These data confirm a key role of autophagy in nitrogen homeostasis and the efficacy of autophagy enhancer molecules for therapy of UCD.

## PgmNr 413: Surrogate biomarkers to assess efficacy of liver-directed genomic therapies in methylmalonic and propionic acidemias.

### Authors:

I. Manoli<sup>1</sup>; A.R. Pass<sup>1</sup>; O.A. Shchelochkov<sup>1</sup>; J.L. Sloan<sup>1</sup>; J. Gagne<sup>1</sup>; E.A. Harrington<sup>1</sup>; S. Koutsoukos<sup>1</sup>; B. Leitner<sup>2</sup>; L. Fletcher<sup>2</sup>; T. Cassimatis<sup>2</sup>; K.Y. Chen<sup>2</sup>; C.P. Venditti<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genomics and Metabolic Genetics Branch, NHGRI, NIH, Bethesda, MD; 2) Diabetes, Endocrinology and Obesity Branch, NIDDK, NIH, Bethesda, MD

Methylmalonic (MMA) and propionic acidemia (PA) are inborn errors of metabolism affecting adjacent steps in the propionyl-CoA oxidation pathway and sharing clinical and biochemical manifestations. Liver and/or kidney transplantation (LT/KT/LKT) are employed to stabilize severely affected patients, while efforts to develop genomic therapies for both disorders are underway. Disease-specific metabolites, such as serum methylmalonic and 2-methylcitric acid are highly variable and affected by dietary protein intake and renal function. With the aim of developing additional biomarkers that correlate with hepatic enzyme activity and disease burden, we have previously used a transgenic murine model of MMA (*Mut*<sup>-/-</sup>;Tg<sup>INS-MCK-Mut</sup>) to demonstrate that *in-vivo* 1-<sup>13</sup>C-propionate oxidation and plasma fibroblast growth factor-21 (FGF21) concentrations improve significantly in response to AAV8 hAAT *MUT* gene therapy. We evaluated 75 patients with MMA and 31 with PA at the NIH Clinical Center, including 20 transplanted MMA (LT/LKT, N=16 and KT, N=4) and 3 PA patients (LT, N=3). Isotopomer enrichment (<sup>13</sup>CO<sub>2</sub>/<sup>12</sup>CO<sub>2</sub> ratio) and cumulative percent of dose metabolized were measured using isotope ratio mass spectrometry in breath samples collected over two hours after an oral bolus of sodium 1-<sup>13</sup>C-propionate. Baseline VCO<sub>2</sub> production (ml/min) was measured by indirect calorimetry. The cumulative percent dose of 1-<sup>13</sup>C-propionate oxidized was 47.2 ± 7.0% (mean ± SD) in LT/LKT recipients compared to 26.9 ± 12.7 in non-transplanted MMA patients (*p*=0.008) and 67.6 ± 2.9% as opposed to 15.9 ± 13.3 (*p*<0.0001) in the PA cohort. Restored oxidative capacity post-LT/LKT was not different from healthy controls 46.81 ± 2.96% (N=16). Patients with KT (N=4) on the other hand had minimal improvement in 1-<sup>13</sup>C-propionate oxidation. Repeatability of the 1-<sup>13</sup>C-propionate breath test was tested in 6 controls and 13 MMA patients showing a coefficient of variation (CV) of 5.93 ± 3.84% and 12.87±12.53%, respectively. Serum MMA values in the same patients had a higher CV of 28.36 ± 24.90%. Circulating FGF21 concentrations were lower in LT/LKT patients 771.7 ± 822.1 vs. 6383.6 ± 8268.5 pg/ml in MMA (*p*<0.0001) and 201.1 ± 240.9 vs. 2316.8 ± 2912.9 (*p*=0.002) in PA LT recipients. 1-<sup>13</sup>C propionate oxidation and plasma FGF21 concentrations should be studied in concert with canonical metabolites to examine therapeutic effects of experimental liver-targeted genomic therapies for disorders of propionate oxidation.

## PgmNr 414: Designer triglycerides: A practical approach to deliver disease-specific alternative fatty acids, bypassing the metabolic block in patients with fatty acids oxidation disorders.

### Authors:

A.-W. Mohsen Mohammed <sup>1,2</sup>; A. Karunanidhi <sup>1</sup>; Y.L. Phua <sup>1</sup>; C. Van't Land <sup>1</sup>; J. Vockley <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, University of Pittsburgh, Pittsburgh, PA.; 2) Human Genetics, University of Pittsburgh, Pittsburgh, PA

**Introduction.** Mitochondrial fatty acid  $\beta$ -oxidation (FAO) disorders are life threatening, even with optimum treatment. FAO is the major source of energy for heart and is critical for skeletal muscle especially during physiologic stress. Symptoms of FAO disorders (FAODs) include cardiac conduction abnormalities, arrhythmias, cardiomyopathy, muscle weakness, hypoketotic hypoglycemia, and/or stress-related rhabdomyolysis. MCT oil bypasses long chain FAO but patients are still at risk for rhabdomyolysis and late onset cardiomyopathy. Triheptanoin (triheptanoylglycerol) in clinical trials improves cardiomyopathy and hypoglycemia of long chain FAODs but rhabdomyolysis persists, and more effective therapy to prevent the heterogeneous symptoms of these disorders is needed.

**Methods.** Fibroblasts from patients deficient in CPT II, very long chain acyl-CoA dehydrogenase (VLCAD), long chain 3-hydroxyacyl-CoA dehydrogenase (LCHAD), trifunctional protein (TFP), and medium chain acyl-CoA dehydrogenase (MCAD) were cultured with lipid stripped FBS. Cells were treated with heptanoic (C7), 2,6-dimethylheptanoic (dMC7), ENG05 (a designer lipid), or 4,8-dimethylnonanoic (dMC9) acids for 72 hr. Acetyl-, propionyl-, and succinylcarnitine were measured in media and TCA cycle intermediates were measured in cells.

**Results.** Media acetylcarnitine nearly doubled in VLCAD and TFP deficient cells when treated with dMC7, dMC9, or ENG05 compared to C7, while increasing in LCHAD, CPT II, and MCAD deficient cells treated with dMC9 by 10%, 25%, and 22%, respectively. With the exception of one VLCAD cell line, media propionylcarnitine was higher in C7 treatment compared to others. Media succinylcarnitine increased with dMC9 compared to C7 in all deficient cell except TFP. Intracellular succinate was higher with ENG05 treatment in two VLCAD (226% and 260%) and CPT II (161%) deficient cell lines compared to C7. Intracellular malate was higher with ENG05 in two VLCAD (22% and 90%), LCHAD (38%), TFP (210%), and CPT II (35%) deficient cells compared to C7.

**Discussion.** Persistence of rhabdomyolysis in patients treated with C7 suggests inadequate access to muscle tissue. We have shown that alternative fatty acids improve the cellular profile in fibroblasts of four key metabolites compared to C7 treatment. These findings suggest more efficient cellular utilization of the alternative fatty acids tested and potential advantage over C7. Muscle distribution in animal models remains to be demonstrated.

## PgmNr 415: The International Cantú Syndrome Registry: Patient findings and therapeutic options.

### Authors:

C.G. Nichols <sup>1</sup>; C. McClenaghan <sup>1</sup>; H.I. Roessler <sup>2</sup>; K. Duran <sup>2</sup>; K. Shields <sup>1</sup>; M.S. Remedi <sup>1</sup>; N.V.A.M. Knoers <sup>2</sup>; J.M. Lee <sup>1</sup>; E.P. Kirk <sup>3</sup>; I. Scurr <sup>4</sup>; S.F. Smithson <sup>4</sup>; G.K. Singh <sup>1</sup>; M.M. van Haelst <sup>2</sup>; G. van Haaften <sup>2</sup>; D.K. Grange <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Washington University in Saint Louis, Saint Louis, MO.; 2) Utrecht University, Utrecht, Netherlands; 3) University of New South Wales, Randwick NSW, Australia; 4) University Hospitals, Bristol, UK

---

Cantú syndrome (CS), first described in 1982, is caused by pathogenic variants in *ABCC9* and *KCNJ8*, which encode pore forming and regulatory subunits of ATP-sensitive potassium ( $K_{ATP}$ ) channels, respectively. Multiple case reports of individual patients have described the various clinical features of CS, but systematic studies have been lacking. To define the effects of genetic variants on CS phenotypes and clinical outcomes, we have developed a standardized REDCap-based registry for CS patients. We report phenotypic and associated genotypes on 75 CS patients, with confirmed *ABCC9* variants in 72 of the cases. Hypertrichosis and a characteristic facial appearance are present in all cases. Polyhydramnios during fetal life, lymphedema, patent ductus arteriosus (PDA), cardiomegaly, dilated aortic root, vascular tortuosity of thoracic and cerebral arteries and migraine headaches are common features, although even with this large group of patients, there is incomplete penetrance of CS-associated features, without clear correlation to genotype.

Many patients report taking cardiac medications currently, or having taken them in the past. These include ACE inhibitors, such as quinapril and enalapril, and beta-blockers, such as atenolol and bisoprolol, as well as previous treatment for pulmonary hypertension with sildenafil. Most patients with edema report current or previous medication treatment with diuretics, including furosemide, spironolactone and amiloride. There is currently no targeted therapy for CS, and it is unclear whether the above therapies are in all cases appropriate. However, we find that CS-associated cardiomegaly in mice carrying CS mutations knocked-in to the endogenous *ABCC9* locus is reversed by the sulfonylurea glibenclamide. Sulfonylureas directly inhibit overactive  $K_{ATP}$  channels, and thus may be a viable option for directed therapy. Hence, we intend to perform a clinical trial to test glibenclamide in CS patients.

## PgmNr 416: Clinical genetic results aided by a patient-provider app intended for diverse populations.

### Authors:

J. Chang<sup>1</sup>; M. Penon-Portmann<sup>1,2</sup>; M. Kvale<sup>1</sup>; J. Van Ziffle<sup>1,3</sup>; S. Rego<sup>1</sup>; T. Yip<sup>1</sup>; H. Hoban<sup>1</sup>; P. Martin<sup>1</sup>; P. Devine<sup>1,3</sup>; A. Tam<sup>1,2</sup>; U. Hodoglugil<sup>1</sup>; F. Chen<sup>1</sup>; N. Sahin-Hodoglugil<sup>1</sup>; A. Rajkovic<sup>1</sup>; A. Slavotinek<sup>1,2</sup>; M. Norton<sup>1</sup>; B. Koenig<sup>1</sup>; N. Risch<sup>1,4</sup>; P.Y. Kwok<sup>1,5</sup>; J.T. Shieh<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 2) Division of Medical Genetics, Pediatrics, Benioff Children's Hospital, University of California San Francisco, San Francisco, CA; 3) Department of Pathology, University of California San Francisco, San Francisco, CA; 4) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 5) Department of Dermatology, University of California San Francisco, San Francisco, CA

---

Genomic data can optimally advance health if results are accessible and understood by patients and providers. Too often, even in advanced health care systems, genetic test results exist in complex formats that make comprehension difficult for end-users. The complexity particularly impacts individuals tested by new technologies, as genes and conditions are often rare and unfamiliar. In the UCSF P3EGS study, part of the CSER2 Consortium, we have developed a user-friendly patient-provider application that integrates genomic and other medical data into a mobile platform that supports result communication and point-of-care access in the clinic. Here we demonstrate findings from this new app that we have implemented for the return of medical genetic results. First, we defined foundational elements for return of results that can be facilitated by an app. Second, we piloted the app in clinic with patients and providers, comparing paper to electronic modalities in individuals receiving results for undiagnosed conditions. Third, we performed patient surveys, with 39 returned results to date for analysis. We show the following: 1) The genetic testing results, specifically the diagnostic conclusions, were concordant in recipients and providers in 90% of responses. 2) Recipients preferred app-based information over paper, although there were exceptions, and 3) The findings were consistent among individuals of diverse socio-economic and educational backgrounds. These data suggest that patient-provider apps may play a novel role in facilitating results communication. As more people in the population are sequenced, further implementation and expansion of similar technologies may be needed. Further collaborative efforts will encourage cross-institutional technology development, data sharing and harmonization of outcome measures.

## PgmNr 417: Utilizing gene expression to prioritize genes in medical patients.

### Authors:

P.E. Evans; L.R. Rives; R.H. Hamid; N.J.C. Cox; Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt University Medical Center 1211 Medical Center Drive, Nashville, TN 37232

---

The Undiagnosed Diseases Network (UDN) is an NIH funded study, implemented to bring together clinical and research experts from across the country. The UDN consists of 12 clinical sites across the country. At Vanderbilt, we employ a variety of methods to find the underlying causes of a patient's phenotypes. This includes on-site visits by medical staff to examine the patient and verify reported phenotypes. When a patient's disease is thought to have a genetic cause, whole-genome or whole-exome sequencing is performed in the patient and any available family members. Variants are then examined for frequency, pathogenicity, and relevance to the patient's disease to create a list of candidate mutations that may contribute to the patient's disease.

In order to gain further insight into the disease etiology of several of our unsolved cases. We performed RNA-seq on blood samples from 39 individuals whose disease etiology could not be determined. This allowed us to compare expression levels of each gene in these patients to individuals within a healthy population. We compared our patient gene expression data to that of the Gene-Tissue Expression Consortium (GTEx). Genes which appeared as outliers in the distribution of expression within the GTEx whole blood samples, became candidate genes for further study in that patient, and any known diseases associated with these genes were examined.

We also compared predicted expression, using the PrediXcan method, of each gene in our patients to predicted expression within GTEx, BioVU (electronic health records at Vanderbilt linked to DNA) and to the observed expression within that patient to look for outliers in predicted expression. Since, the elastic net models that the PrediXcan models utilize usually favor common SNPs, deviation of the observed expression from the predicted expression may indicate that a previously uncharacterized rare SNP or variant is within this gene or its regulatory components. Furthermore, genes whose expression falls outside the distribution of predicted expression in GTEx or BioVU may indicate a large genetic burden of common variants attributable to a given gene. We outline several cases in which this approach has shown promising results, leading to better patient care, and providing answers for the families involved.



## PgmNr 418: New strategies for analyzing exomes to increase diagnostic yield from patients with rare disorders.

### Authors:

K. Schmitz-Abe <sup>1,2,3</sup>; Q. Li <sup>2,3</sup>; N. Nori <sup>2,3</sup>; J. Madden <sup>2,3</sup>; W. Wu <sup>2,3</sup>; J. Lin <sup>2,3</sup>; R. Rosen <sup>2,3</sup>; C. Genetti <sup>2,3</sup>; A. Beggs <sup>2,3</sup>; P. Agrawal <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Newborn Medicine and Neonatal Genomics Program; 2) Division of Genetics and Genomics; 3) The Manton Center for Orphan Disease Research, Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA, 02115

---

Whole exome sequencing (WES) is becoming increasingly important in enabling tailored interventions in 'personalized' or 'precision' medicine. Recent studies suggest that in patients with suspected genetic disorders, WES should be applied as a first-tier molecular test, in parallel with standard diagnostic care, to improve the cost-effectiveness of the diagnostic process. However, limitations of the current analytical methods, genetic and phenotypic heterogeneity, knowledge gaps in the gene-disease association, and interpretation/reporting differences often hinder the identification of disease-causing variants. Ultimately, only 25-30% of families receive a diagnosis after WES, leaving the genetic basis of the remaining families' diseases unknown. We have shown recently that periodic reanalysis of negative WES data using improved bioinformatic tools and up-to-date, gene-disease databases can identify additional candidate variants, potentially increasing the diagnostic yield to 45-55%. Despite the use of these approaches, many cases remain negative after reanalysis. Thus, there is a need to develop a comprehensive strategy for reanalysing negative WES that will further improve the diagnostic yield. In addition, a lack of standardization of this process leads to high variability in diagnostic yield, as well as lack of optimum care for patients who do not receive a diagnosis. In order to improve diagnostic outcomes for these patients, we have introduced several complementary methods to our pipeline including copy number variation (CNV) analysis, UK Biobank data and a small neural network-training algorithm for the optimization of the analysis process. The incorporation of 50,000 samples with phenotypic information and whole exome/genome data from the UK Biobank with proper statistical models will certainly improve the diagnostic yield. Here we report our cumulative experience from reanalysis of a WES cohort enrolled in The Manton Center using a custom-built, comprehensive variant detection and analysis pipeline in conjunction with updated phenotypic information, literature, statistical models and databases in collaboration with an interdisciplinary team. These advances allowed us to reach a confirmed or potential diagnosis for up to a third of previously negative CES cases, demonstrating a meaningful improvement in clinical diagnostic yield.

## **PgmNr 419: Large-scale genomic characterization of pediatric patients with rare diseases at Shriners Hospital for Children's Genomics Institute (SHCGI).**

### **Authors:**

K. Shazand; A. Gustafson; M. Gross-King; V.N. Bijanki; L. Wack; A. Carpenter; T. Morse; C. Hilton; K. Meader; S. Terkoski; D. Westberry; M. Maciel; G.S. Gottesman; M. Lalande

[View Session](#) [Add to Schedule](#)

**Affiliation:** Shriners Hospitals for Children Headquarters, 2900 N Rocky Point Dr., Tampa FL. 33607 - USA

---

Shriners Hospitals for Children is a one-of-a-kind international health care system of 22 facilities dedicated to improving the lives of children by providing specialty pediatric care, innovative research and training programs. Children with orthopedic conditions, burns, spinal cord injuries and cleft lip and palate are eligible for care regardless of their ability to pay.

To expand our capabilities and develop the best tools for our patients, SHCGI has recently initiated a genomic study involving 5,000 patients and their biologic parents (15,000 subjects in total). After informed consent is obtained, their genomes will be characterized by Massively Parallel Sequencing. Diseases such as cerebral palsy, osteogenesis imperfecta, idiopathic scoliosis, metabolic bone diseases, and spina bifida are among the disorders to be investigated. Identified variants will be associated with the family history and clinical information to help fully characterize each condition and its transmission pattern. Functional validation of the identified variants and involved pathways will follow this initial discovery phase.

So far, little is known about many of these disorders, although a strong genetic component is suspected. Our network of hospitals will participate in this unique project, and the resulting outcomes should allow us to achieve better support and targeted care for our patients as better diagnostic and therapeutic tools emerge.

## PgmNr 420: Lessons from the TOTEM trial, a phase 1/2 multicentre, open-label, single-arm study of low-dose PI3K inhibitor taselisib in adult patients with PI3KCA-related overgrowth (PROS).

### Authors:

M. Luu<sup>1</sup>; P. Vabres<sup>2</sup>; H. Devilliers<sup>1</sup>; R. Loffroy<sup>3</sup>; M. Carpentier<sup>4</sup>; C. Fleck<sup>4</sup>; A. Maurer<sup>1</sup>; M. Yousfi<sup>2</sup>; M. Bardou<sup>1</sup>; L. Faivre<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) INSERM, CIC1432, F-21000 Dijon, France ; Dijon Bourgogne University Hospital, Clinical Investigation Center, F-21000, France; 2) Dijon Bourgogne University Hospital, Centres de références Anomalies du Développement et Maladies Génétiques Cutanées Rares, GAD UMR1231 et FHU TRANSLAD, F-21000, France; 3) Dijon Bourgogne University Hospital, Department of interventional radiology, F-21000, France; 4) Dijon Bourgogne University Hospital, Clinical Research Department, F-21000, France

---

**Background:** Activating PIK3CA mutations in overgrowth syndromes provide a rationale for targeted PI3K inhibition as a drug therapy approach in PROS patients. Taselisib, a selective class I PI3K-inhibitor developed for breast cancer therapy, has been shown to suppress aberrant Pi3K hyperactivation in non-clinical pharmacological studies.

**Objective:** To establish the six-month tolerance of low-dose of taselisib in PI3KCA-related overgrowth (PROS) adult patients

**Methods:** The study was to enroll 30 adult patients with PROS in a phase IB/IIA multi-centre, open-label single arm trial. Using a 3+3 dose escalation design, a first batch of six patients received 1mg of taselisib daily before escalating at 2mg. The primary outcome was the occurrence of dose limiting toxicities defined as grade  $\geq 3$  adverse reaction (AR) during the first month of treatment. The study was to be interrupted on the occurrence of one life-threatening Suspected Unexpected Serious Adverse Reaction (SUSAR),  $\geq 2$  SUSAR, or  $\geq 5$  serious AR. All safety assessments were reviewed by an independent safety committee. Preliminary efficacy outcomes were assessed by measuring relative changes before and after treatment of a) tissue volume at affected and unaffected sites clinically and by dual energy X-ray absorptiometry, b) vascular outcomes and biologic parameters, c) quality of life. The study was conducted in accordance with the Declaration of Helsinki.

**Results:** A total of 19 patients were enrolled, 17 received taselisib and 11 completed the study. Whereas no serious AR occurred in the 1mg cohort (6 patients), the study was terminated after 11 patients were included in the 2mg cohort, due to two SUSAR (acute ileitis and a pachymeningitis). Five patients at 2mg had already completed the study, and taselisib was definitively withdrawn in the remaining 6. All the patients treated experienced at least 1 AR. Three patients had a grade 3 AR (the 2 SUSAR and 1 parvovirus infection). Gastrointestinal disorders, nervous system disorders and infections accounted for respectively 43% (43/113), 17% (19/113) and 12% (13/113) of AR. Preliminary efficacy data suggest improvement such as pain reduction, chronic bleeding cessation or

improved quality of life.

**Conclusion:** The first therapeutic trial of a PI3KCA-inhibitor in PROS patients suggests that, despite promising efficacy results, low-dose Taselisib can induce serious drug-related AR that prevent consideration of long term use.

## PgmNr 421: Adenine base editing of the *DUX4* somatic polyadenylation signal in facioscapulohumeral muscular dystrophy.

### Authors:

D. Sikrova; J. Balog; S.M. van der Maarel

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, Leiden University Medical Center, Leiden, Zuid Holland, Netherlands

Facioscapulohumeral muscular dystrophy (FSHD) is an inherited myopathy caused by sporadic expression of the transcription factor double-homeobox 4 (*DUX4*) in skeletal muscle. Expression of *DUX4* in skeletal muscle results from incomplete epigenetic repression of its locus residing within the D4Z4 macrosatellite repeat in the 4q subtelomere. Since *DUX4* is a transcriptional activator, once expressed it induces improper transcriptional changes in muscle cells, eventually leading to cell death. Interestingly, epigenetic de-repression of the D4Z4 repeat itself is not sufficient to cause FSHD as it needs to occur on a specific genetic background of chromosome 4q. There are two main subtelomeric 4q allelic variants (termed 4qA and 4qB) with only de-repression of the D4Z4 repeat on 4qA resulting in FSHD. This is due to polymorphisms between 4qA and 4qB immediately distal to the last D4Z4 repeat unit, creating only on 4qA alleles an extra exon for *DUX4* that contains a somatic polyadenylation signal (sPAS) for *DUX4* in muscle cells. It has been shown that even a single nucleotide substitution in the conserved *DUX4* PAS hexameric motif has a negative effect on the cleavage and polyadenylation of the mRNA precursor leading to gene downregulation.

Here, we wanted to explore if mutating endogenous *DUX4* sPAS would lead to reduced transcript production and thus attenuating its downstream harmful effects on myogenic cells. To achieve efficient and precise mutagenesis of the *DUX4* sPAS, we made use of CRISPR/Cas9-guided adenine base editing. The adenine base editor (ABE) allows direct conversion of targeted adenine(s) into guanine(s) and as such bypasses the need of template-dependent homologous recombination pathway which usually represents the limiting step when introducing specific nucleotide changes at the region of interest. Conveniently, the *DUX4* sPAS is a good substrate for ABE-mediated A to G mutagenesis as a nearby protospacer adjacent motif (PAM) site positions the last three adenines of the *DUX4* sPAS in the active editing window of ABE. The results of this study provide insight into the role of the *DUX4* sPAS in FSHD pathogenesis and the regulation of *DUX4* in muscle cells.

## PgmNr 422: A CRISPR-Cas13a based strategy that tracks and degrades toxic RNA in myotonic dystrophy type 1.

### Authors:

N. Zhang <sup>1</sup>; B. Bewick <sup>1</sup>; G. Xia <sup>2</sup>; D. Furling <sup>3</sup>; T. Ashizawa <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology, Houston Methodist Research Institute, Houston, TX77030, USA; 2) Department of Neurology and Neuroscience, University of New Mexico, NM87131, USA; 3) Sorbonne Universite, Inserm, Association Institut de Myologie, Centre de Recherche en Myologie, CRM, F-75013 Paris, France

---

Cas13a, an effector of type VI CRISPR-Cas systems, is an RNA guided RNase with multiplexing and therapeutic potential. This study employs the *Leptotrichia shahii* (*Lsh*) Cas13a and a repeat-based CRISPR RNA (crRNA) to track and eliminate toxic RNA aggregates in myotonic dystrophy type 1 (DM1) – a neuromuscular disease caused by CTG expansion in the *DMPK* gene. We demonstrate that *Lsh*Cas13a cleaves CUG repeat RNA in biochemical assays and reduces toxic RNA load in patient-derived myoblasts. As a result, *Lsh*Cas13a reverses the characteristic adult-to-embryonic missplicing events in several key genes that contribute to DM1 phenotype. The deactivated *Lsh*Cas13a can further be repurposed to track RNA-rich organelles within cells. Our data highlights the reprogrammability of *Lsh*Cas13a and the possibility of using Cas13a as a means to treat other microsatellite expansion diseases.

## PgmNr 423: An NIA/AA inspired polygenic risk score framework for cognitive decline in old age.

### Authors:

A.M. Moore<sup>1,2,3</sup>; T.J. Filshtein<sup>4</sup>; L. Dumitrescu<sup>1,2</sup>; A. Harrati<sup>5</sup>; F. Elahi<sup>6</sup>; E.C. Mormino<sup>7</sup>; D.M. Mungas<sup>8</sup>; T. Hedden<sup>9</sup>; L.G. Apostolova<sup>10</sup>; A.J. Saykin<sup>11</sup>; D. Chasioti<sup>11</sup>; R. Buckley<sup>12</sup>; T.J. Hohman<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Memory and Alzheimer's Center, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA; 5) School of Medicine, Stanford University School of Medicine, Palo Alto, CA; 6) Department of Neurology, Memory and Aging Center, University of California, San Francisco, CA; 7) Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Palo Alto, CA; 8) UC Davis Alzheimer's Disease Research Center, University of California, Davis, CA; 9) Icahn School of Medicine at Mount Sinai, New York City, NY; 10) Department of Neurology, Indiana University School of Medicine, Indianapolis, IN; 11) Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN; 12) Department of Neurology, Massachusetts General Hospital/Harvard Medical School, Boston, MA

---

**Objective:** Understanding genetic risk for Alzheimer's disease (AD) is crucial to optimizing early detection and intervention. The National Institute on Aging and Alzheimer's Association have proposed the A/T/N AD research framework for interactions between three main biomarkers of disease, amyloid plaques (A), phosphorylated tau tangles (T), and neurodegeneration (N), and how they contribute to cognitive decline associated with AD. We developed a novel polygenic risk score (PRS) based on the A/T/N framework to integrate genetic risk for AD biomarkers into the framework (the A/T/N PRS). We then compared the predictive capabilities of this PRS score for AD risk based on a case-control design. **Methods:** We used summary statistics from genome wide association studies (GWAS) of cerebrospinal fluid (CSF) phosphorylated-tau (ptau<sub>181</sub>), and amyloid- $\beta$  (A $\beta$ <sub>42</sub>), as well as late-onset AD risk genes and left hippocampal volume (LHIPV) to calculate PRS scores for 1182 subjects in the Alzheimer's Disease Neuroimaging Initiative (ADNI). The  $\alpha$ -cutoff was 0.01 for index single nucleotide polymorphisms (SNPs) and the physical distance for linkage disequilibrium (LD) clumping was 200kb. Linear regression validated performance of PRS scores for CSF biomarkers and hippocampal volume on respective outcomes with covariates for age, sex, and years of education. Additionally, CSF models were covaried for batch and LHIPV for magnet strength and intracranial volume. A generalized linear model assessed AD risk PRS performance on AD diagnosis. PRS scores were Blom transformed and averaged to calculate an A/T/N PRS. We evaluated the performance of the A/T/N PRS and AD risk PRS scores on baseline and longitudinal cognitive composites of executive function and memory by comparing p-values and R<sup>2</sup> values. **Results:** The A/T/N PRS showed superior predictive performance on executive function decline, accounting for 1.33% of the variance (whole model, 12.50%), compared to 0.40% variance explained by the AD risk PRS (whole model, 11.57%). The A/T/N PRS and the AD risk PRS performed comparably for baseline executive function, baseline and change in memory performance. **Discussion:** The A/T/N PRS was developed to model the amyloidosis, tauopathy, neurodegeneration framework that drives AD progression, and outperforms a PRS for AD risk to predict change in executive function. Integration of genetic risk across biomarkers

relevant to AD may represent a more refined strategy to model disease progression.



## PgmNr 424: Characterizing the mechanisms of *SLC35A2* epilepsy in human iPSC-derived neurons.

### Authors:

D. Lai <sup>1</sup>; P. Sosicka <sup>2</sup>; A.K. Ressler <sup>1</sup>; M.J. Boland <sup>1</sup>; H.H. Freeze <sup>2</sup>; E.L. Heinzen <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Genomic Medicine, Columbia University, New York, NY; 2) Human Genetics Program, Sanford-Burnham-Prebys Medical Discovery Institute, La Jolla, CA; 3) Department of Pathology and Cell Biology, Columbia University, New York, NY

---

Germline genetic variants of *SLC35A2*, which encodes a Golgi-localized UDP-galactose transporter (UDPGaT) essential for cellular galactosylation, have been implicated in one type of congenital disorder of glycosylation associated with intractable seizures and in rare X-linked developmental and epileptic encephalopathy. Recently, we have identified post-zygotically-acquired, *de novo*, loss-of-function *SLC35A2* variants in 17% of non-lesional focal epilepsy cases, however, mechanisms of *SLC35A2* in the epileptogenic process remain elusive. This study seeks to study how pathogenic *SLC35A2* variants affect neural network activity. We have established human induced pluripotent stem cell lines (hiPSC; 713-5, isogenic control) CRISPR-edited to harbor either a pathogenic missense variant (c.910T>C p.Ser304Pro; *SLC35A2*<sup>S304P/Y</sup>) or a frameshift indel (*SLC35A2*<sup>-Y</sup>). Immunofluorescent (IF) staining for NANOG and SOX2 confirms the isogenic control, *SLC35A2*<sup>S304P/Y</sup> and *SLC35A2*<sup>-Y</sup> exhibit pluripotent characteristics. However, expression of pluripotent marker TRA1-81, which recognizes a galactose-containing keratan sulfate epitope on podocalyxin, was lost in *SLC35A2*<sup>S304P/Y</sup> and *SLC35A2*<sup>-Y</sup> compared to the isogenic control suggesting *SLC35A2* variants encode loss-of-function proteins. This was further confirmed using a MALI lectin binding assay, which detects terminal sialic acid on glycoproteins, that showed reduced binding in *SLC35A2*<sup>S304P/Y</sup> and *SLC35A2*<sup>-Y</sup> protein lysates compared to the isogenic control. Loss of transporter activity is likely due to loss of UDPGaT expression in *SLC35A2*<sup>S304P/Y</sup> and *SLC35A2*<sup>-Y</sup> hiPSCs as shown by reduced IF staining and no co-localization with GM-130 (Golgi marker) as observed in the isogenic control. To investigate how *SLC35A2* variants impact neural network activity and development, hiPSC-derived neurons were plated on multi-electrode array. Compared to the isogenic control, *SLC35A2*<sup>S304P/Y</sup> exhibit synchronous activity at an earlier onset compared to the isogenic control as quantified by the spike train tiling coefficient, mutual information, and percentage of spikes in network spikes. This suggests that *SLC35A2* variants influence the development of neural network connectivity, which may be contributing to the epileptogenic process. Further characterization of the underlying mechanisms of *SLC35A2*-associated epileptogenesis across an allelic series and the exploration of the potential of galactose treatment in this model system are underway.

## PgmNr 425: Applying genomic medicine to patients with mesial temporal lobe epilepsy: Predicting response to treatment.

### Authors:

I. Cendes<sup>4</sup>; M.C. Molck<sup>1,4</sup>; R. Secolin<sup>1,4</sup>; M.K.M. Alvim<sup>2,4</sup>; F.R. Torres<sup>1,4</sup>; C.L. Yasuda<sup>2,4</sup>; B.S. Carvalho<sup>3,4</sup>; F. Cendes<sup>2,4</sup>; I. Lopes-Cendes<sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics and Genomic Medicine. School of Medical Sciences. University of Campinas (UNICAMP), Campinas, SP, Brazil.; 2) Department of Neurology, School of Medical Sciences. University of Campinas (UNICAMP), Campinas, SP, Brazil.; 3) Department of Statistics, Institute of Mathematics, Statistics and Computer Sciences; University of Campinas (UNICAMP), Campinas, SP, Brazil.; 4) Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, SP, Brazil.

---

Mesial temporal lobe epilepsy (MTLE) is the most common form of focal epilepsy in adult patients. A large proportion of patients with MTLE is refractory to treatment with antiepileptic drugs (AEDs) and may be eligible for surgical treatment, which is effective in most patients. To date, the presence of hippocampal sclerosis (HS) is the only factor used to predict poor response to AED in patients with MTLE. The long delay to indicate surgery for AED-refractory MTLE can negatively impact the quality of care offered to these patients. Recently, we proposed an algorithm to predict whether a patient with MTLE will be refractory to AED therapy combining information from clinical variables as well as SNPs in candidate genes. We were able to achieve an accuracy of 0.8177 using 56 SNPs and the presence of HS (Silva-Alves et al., doi:10.1371/journal.pone.0169214). In the present study, we aim to replicate and validate this algorithm in an independent cohort and to test whether the use of a limited number of SNPs could be enough to identify patients who are refractory to AED-therapy. For this, we selected the ten topmost significant SNPs obtained in the previous report and genotyped a total of 253 patients with MTLE, who were classified into two groups, 38 AED-responsive and 215 AED-refractory. We obtained an accuracy of 0.799, with a sensitivity of 0.92, in the prediction of AED-refractory patients. We successfully validated our previous results in an independent cohort and demonstrated that by using only 10 SNPs, it is possible to predict with high accuracy and sensitivity patients who will be refractory to AED therapy. Thus, suggesting that by incorporating genetic testing into the medical evaluation of patients with MTLE, one can indicate epilepsy surgery sooner to those predicted to be AED-refractory, which is likely to significantly improve the quality of care offered to these patients. Study supported by FAPESP, SP, Brazil.

## PgmNr 426: Zebrafish modeling for the clinic: Rapid *in vivo* functional testing of patient variants for clinical applications.

### Authors:

B. Jussila; T. Brock; C. Hopkins; R. Fekete

[View Session](#) [Add to Schedule](#)

**Affiliation:** NemaMetrix, Inc., Eugene, OR.

---

Genomic sequencing has become more prevalent in clinical settings. The result has been a rapid increase in identification of suspect variants for causing disease. Unfortunately, many of the variants identified are either not annotated, or they are assigned a Variant-of-Uncertain-Significance (VUS) status. To improve diagnostic rates, strong (PS3) assessments of pathogenic or benign status can be obtained from functional studies. In this work, we use zebrafish (*Danio rerio*) as an intact animal model to measure the functional effects of patient-derived genetic variation. Two advantages of modeling in zebrafish are a high level of conservation with human disease associated genes (~95%) and a large clutch capacity enabling high throughput applications, such as drug testing and behavioral assessment. Using CRISPR/Cas9-mediated gene editing, we targeted the zebrafish ortholog of human syntaxin binding protein 1 (*STXBP1*), denoted *stxbp1a*, as a test case for modeling genetic contribution to pediatric epilepsy and encephalopathy. The variants chosen for the functional studies include VUS and established pathogenic and benign alleles. In a complementary approach, morpholino (MO) knock-down coupled with transient mRNA rescue with variant cDNA is being explored as a rapid assay method for detecting variant pathogenicity. The result is a toolkit in zebrafish for uncovering the functional consequence of genomic variation. The zebrafish studies integrate into our Precisome Platform - rapid, actionable, and quantifiable results for variant pathogenicity measured in disease-proxy animals ('Clinical Avatars'). The platform combines the complementary variant modeling in *C. elegans* (worms) and *D. rerio* (fish) to create a correlative body of evidence for pathogenicity that is strengthened by the conservation of biology between 3 species: worm, fish and humans. For example, the functional data presented will explore the connection of patient-specific phenotype and the conserved morphological, molecular, and behavioral activity of the Clinical Avatar animals. By including zebrafish into the Precisome Platform, strong correlative assessment is achieved for assigning pathogenicity to gene variations observed in the patient.

## PgmNr 427: Anti-sense oligonucleotide therapy delays seizure onset and extends survival in a mouse model of *Scn8a* encephalopathy.

### Authors:

M.H. Meisler <sup>1,3</sup>; G.M. Lenk <sup>1</sup>; P. Jafar-Nejad <sup>2</sup>; L.D. Huffman <sup>3</sup>; C.E. Smolen <sup>1</sup>; J.L. Wagnon <sup>1</sup>; H. Petit <sup>1</sup>; R.J. Giger <sup>3,4</sup>; F. Rigo <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Human Gen, Univ Michigan, Ann Arbor, Michigan; 2) Ionis Pharm, Carlsbad, California; 3) Neuroscience Program, Univ Michigan, Ann Arbor, Michigan; 4) Dept Cell and Devel Biol, Univ Michigan, Ann Arbor, Michigan

---

The gene *SCN8A* encodes the voltage-gated sodium channel Na<sub>v</sub>1.6, which is localized at the axon initial segment and at nodes of Ranvier in the CNS and PNS. Exome sequencing has identified several hundred *de novo* mutations of *SCN8A* in patients with epileptic encephalopathy (MIM#614558), a severe disorder characterized by early onset seizures, developmental delay and cognitive impairment. Many patients with *SCN8A* encephalopathy are nonambulatory and nonverbal, and the condition is resistant to standard anti-epileptic drugs. Most patient mutations of *SCN8A* result in gain-of-function changes leading to elevated channel activity and neuronal hyperactivity. To test therapeutic interventions, we generated a conditional mouse model of the patient mutation p.Arg1872Trp (*Bunton-Stasyshyn, Wagnon et al, Brain 2019*). We used this model to evaluate the effectiveness of anti-sense oligonucleotides (ASOs) to compensate for neuronal hyperactivity by reducing the abundance of the Na<sub>v</sub>1.6 mRNA. *Scn8a*<sup>cond/+</sup>, *E2A-CRE* mice were treated by intracerebroventricular injection of an ASO that decreases the abundance of the *Scn8a* transcript by up to 50%. We observed a dose-dependent increase in length of seizure-free survival of the mutant mice from 2 weeks for untreated mice to 9 weeks in mice receiving two treatments of ASO. The ASO-treated mice did not exhibit any of the side effects that can result from greater reduction of Na<sub>v</sub>1.6 function, such as muscle wasting and ataxia.

These experiments provide pre-clinical evidence for the effectiveness of ASO therapy to treat an intractable childhood epilepsy. Reduction of neuronal excitability with an *SCN8A*-ASO is potentially applicable to other epilepsies. *Supported by NIH R01 NS34509.*

## PgmNr 428: HMI-202 and gene therapy for metachromatic leukodystrophy (MLD).

### Authors:

J. Gingras; T. St-Martin; K. Gall; T. Seabrook; N. Avila; I. Rivas; S. Seidel; M. Chittoda; M. Mercaldi; L. Hennessey; T. Wright; S. Krupa; O. Francone; A. Seymour

[View Session](#) [Add to Schedule](#)

**Affiliation:** Homology Medicines, Inc., Bedford, Massachusetts

---

Metachromatic leukodystrophy, commonly known as MLD, is an inherited autosomal recessive lysosomal storage disorder with a great unmet medical need. This fatal neurodegenerative disease occurs in three forms: late infantile (prevalence of 1 in 40,000), juvenile, and adult. The late infantile and juvenile forms represent the majority of the MLD patients and mortality at 5 years is estimated at 75% and 30%, respectively. Most commonly, MLD is caused by mutations in the ARSA gene and patients suffering from the disease are deficient in arylsulfatase-A (ARSA) enzyme. The disease is characterized by accumulation of supraphysiologic levels of lipids (sulfatides) in the brain, spinal cord and peripheral organs, which become toxic. This excess sulfatide leads to the destruction of myelin, a key protective layer of the nerve fibers, resulting in nerve damage. Herein, we are reporting preclinical gene therapy data in a murine model of MLD where a single intravenous dose of HMI-202 (AAVHSC15-hARSA) crossed the blood-brain barrier and led to a dose-response relationship in vector genome copies, ARSA protein and ARSA enzymatic activity in the central nervous system (CNS). HMI-202 expression patterns and corresponding enzymatic activity were rapidly detected in key biologically relevant regions of the brain (brainstem, cortex, cerebellum and white matter tracks), spinal cord (gray matter and ascending white matter tracks of the posterior column) and peripheral nervous system (dorsal root ganglion and sciatic nerve) in lysosomes of both neuronal and glial cellular profiles. In summary, HMI-202 is a promising gene therapy in development for the treatment of MLD based on mouse data demonstrating its rapid onset of action and ability to achieve hARSA activity levels at or above the therapeutic threshold (10-15% of normal hARSA activity). Based on these preclinical data, IND-enabling studies of CNS gene therapy development candidate HMI-202 have been initiated.

## PgmNr 429: *PLP1* gene suppression therapy for Pelizaeus-Merzbacher disease using artificial miRNA.

### Authors:

K. Inoue <sup>1</sup>; H. Li <sup>1</sup>; H. Okada <sup>2</sup>; S. Suzuki <sup>1</sup>; K. Sakai <sup>3</sup>; H. Izumi <sup>1</sup>; Y. Matsushima <sup>1</sup>; N. Ichinohe <sup>3</sup>; Y. Goto <sup>1</sup>; T. Okada <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of MR & BD Res, Natl Inst Neurosci, NCNP, Kodaira, Tokyo, Japan; 2) 2.Department of Molecular and Medical Genetics, Graduate School of Medicine, Nippon Medical School, Tokyo, Japan; 3) Dept of Ultrastructural Research, Natl Inst Neurosci, NCNP, Kodaira, Tokyo, Japan

---

Genomic duplication encompassing the entire *PLP1* gene is the most common cause of Pelizaeus-Merzbacher disease (PMD), an X-linked hypomyelinating leukodystrophy showing severe motor and cognitive developmental delay with various neurological phenotypes. Although the exact molecular and cellular mechanisms underlying *PLP1* duplication, which causes severe hypomyelination in the central nervous system, remain largely elusive, *PLP1* overexpression is likely the fundamental cause of this devastating disease. Here, we investigated if adeno-associated virus (AAV)-mediated gene-specific suppression may serve as a potential cure for PMD by correcting quantitative aberrations in gene products. We developed an oligodendrocyte-specific *Plp1* gene suppression therapy using artificial miRNA under the control of human *CNP* promoter in a self-complementary AAV (scAAV) platform. A single direct brain injection achieved widespread oligodendrocyte-specific *Plp1* suppression in the white matter of WT mice. AAV treatment in *Plp1* transgenic mice, a *PLP1* duplication model, ameliorated cytoplasmic accumulation of Plp1, preserved mature oligodendrocytes from degradation, restored myelin structure and expression of myelin genes, and improved survival and neurological phenotypes. Together, we provide evidence that AAV-mediated gene suppression therapy using artificial miRNA can serve as a potential cure for PMD resulting from *PLP1* duplication and possibly be applied to the development of therapies for other genomic disorders.

## PgmNr 430: Epigenome-editing of *SNCA* gene expression: A novel strategy for gene therapy of Parkinson's disease.

### Authors:

B. Kantor<sup>1,2</sup>; L. Tagliafierro<sup>3,4</sup>; E. Ilich<sup>1,2</sup>; L. Brown<sup>1,2</sup>; J. Gu<sup>3,4</sup>; O. Chiba-Falek<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NEUROBIOLOGY, DUKE, Durham, North Carolina.; 2) VIRAL VECTOR CORE, DUKE, Durham, North Carolina; 3) NEUROLOGY, DUKE, Durham, North Carolina; 4) Center for Genomic and Computational Biology, DUKE, Durham, North Carolina

**Background:** The overexpression of *SNCA* play an important role in the pathogenesis of Parkinson's Disease (PD) while, normal physiological levels of gene are crucial to maintain neuronal function. A so-far unmet need is the development of new therapeutic strategies targeting the regulatory mechanisms of *SNCA* expression to fine-tune *SNCA* levels, versus previous approaches that targeted directly the mRNA or the protein product resulting in robust reduction of *SNCA* levels associated with neurotoxicity.

**Objective:** To translate our knowledge of the mechanisms involved in the regulation of *SNCA* gene to next generation drug discovery based on tightly controlled repression of *SNCA* overexpression.

**Methods:** We developed a novel strategy to intervene with the transcription regulation of *SNCA*. The developed system based on the targeted DNA-methylation editing at *SNCA*-intron 1 and comprises of an all-in-one lentiviral vector (LV), carrying CRISPR/dCas9 fused with the catalytic domain of DNA-methyltransferase 3A (DNMT3A). Aiming to further optimize the delivery system, we created a safer integrase-deficient version of the LV.

**Results:** Applying the gRNA-dCas9-DNMT3A into human induced pluripotent stem cells (hiPSC)-derived 'aged' dopaminergic neurons from a PD-patient with the *SNCA* triplication resulted in fine-tuned downregulation of *SNCA*-mRNA and protein levels. Furthermore, the reduction in *SNCA* levels by the gRNA-dCas9-DNMT3A system rescued disease-related cellular-phenotypes characteristics of the *SNCA*-triplication/hiPSC-derived dopaminergic neurons, e.g. mitochondrial ROS-production, cell viability, and exacerbated nuclear aging signatures measured by DNA damage and abnormal nuclear envelop morphology. Next, we validated the vectors *in vivo* utilizing PD rodent models. Following the administration of LV-gRNA/dCas9-DNMT3A particles into the rat brain, we demonstrated an efficient decrease in *Snc*a-mRNA and protein levels.

**Conclusions:** Collectively, these *proof-of-concept* experiments provide the foundation and validation for advancing the novel epigenetic editing-based system further towards PD therapeutic strategy for application in a clinical setting. The system outlined here would be highly attractive for developing 'next generation drugs' as prevention and/or disease modifying interventions for PD, Alzheimer's disease and other neurodegenerative diseases and pathologies associated with dysregulation of gene expression.



## PgmNr 431: A comprehensive haplotype targeting strategy for allele-specific *HTT* suppression in Huntington disease.

### Authors:

C. Kay<sup>1</sup>; J.A. Collins<sup>1</sup>; N.S. Caron<sup>1</sup>; L. de Andrade Agostinho<sup>3,4,5</sup>; H. Findlay-Black<sup>1</sup>; L. Casal<sup>1</sup>; D. Sumathipala<sup>6</sup>; V.H.W. Dissanayake<sup>6</sup>; M. Cornejo-Olivas<sup>7,8</sup>; F. Baine<sup>9,10</sup>; A. Krause<sup>9</sup>; J.L. Greenberg<sup>10</sup>; C.L.A. Paiva<sup>3</sup>; F. Squitieri<sup>2</sup>; M.R. Hayden<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Molecular Medicine and Therapeutics, Univ British Columbia, Vancouver, British Columbia, Canada; 2) Fondazione IRCCS Casa Sollievo della Sofferenza Research Hospital, San Giovanni Rotondo, Italy; 3) PPGNEURO, Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro, Brazil; 4) Centro Universitário UNIFAMINAS, Muriaé/MG, Brazil; 5) Hospital do Câncer de Muriaé, Muriaé/MG, Brazil; 6) Human Genetics Unit, University of Colombo, Sri Lanka; 7) Neurogenetics Research Center, Instituto Nacional de Ciencias Neurológicas, Lima, Peru; 8) Center for Global Health, Universidad Peruana Cayetano Heredia, Lima, Peru; 9) Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; 10) Division of Human Genetics, Department of Pathology, University of Cape Town, South Africa

---

Huntington disease (HD) is a fatal neurodegenerative disorder caused by a gain of function mutation in *HTT*. Suppression of mutant *HTT* expression has emerged as a leading therapeutic strategy for HD, with allele-selective approaches targeting *HTT* SNPs now in clinical trials. Haplotypes associated with the HD mutation (A1, A2, A3a) represent panels of allele-specific gene silencing targets for efficient treatment of HD patients of Northern European and indigenous South American ancestry. Here we extend comprehensive haplotype analysis of the HD mutation to key patient populations of Southern European, South Asian, Middle Eastern, and admixed African ancestry. In each of these populations, the HD mutation occurs predominantly on the A2 *HTT* haplotype. Analysis of HD haplotypes across all affected population groups enables rational selection of candidate target SNPs for development of allele-selective gene silencing therapeutics worldwide. Targeting SNPs on the A1 and A2 haplotypes in parallel is essential to achieve treatment of the most HD patients in populations where HD is most prevalent. Current allele-specific approaches will leave a majority of patients untreated in populations where the HD mutation occurs most frequently on the A2 haplotype, including Southern Europeans, South Asians, and patients of Middle Eastern origin. To develop strategies for allele-specific suppression of the A2 *HTT* haplotype in these patients, we demonstrate preclinical development of potent and selective ASOs targeting defining A2 *HTT* haplotype SNPs. On the basis of comprehensive *HTT* haplotype analysis, we further estimate the maximum proportion of patients that may be treated with three or four allele targets in different populations worldwide, informing current allele-specific *HTT* silencing strategies.



## PgmNr 432: Long-term effect of the gene therapy using AAV vectors with the human intrinsic GLUT1 promoter for *Glut1*-deficient mice.

### Authors:

S. Nakamura <sup>1</sup>; H. Osaka <sup>1</sup>; S. Muramatsu <sup>2,3</sup>; N. Takino <sup>2</sup>; M. Ito <sup>2</sup>; E.F. Jimbo <sup>1</sup>; K. Shimazaki <sup>4</sup>; T. Onaka <sup>5</sup>; S. Ohtsuki <sup>6</sup>; T. Terasaki <sup>7</sup>; T. Yamagata <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of pediatrics, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan; 2) Division of Neurology, Jichi Medical University, Tochigi, Japan; 3) Center for Gene and Cell Therapy, The Institute of Medical Science, The University of Tokyo, Japan; 4) Department of Neurosurgery, Jichi Medical University, Tochigi, Japan; 5) Division of Brain and Neurophysiology, Department of Physiology, Jichi Medical University, Tochigi, Japan; 6) Department of Pharmaceutical Microbiology, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan; 7) Division of Membrane Transport and Drug Targeting, Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan

**Background:** Glucose transporter 1 deficiency syndrome (GLUT1DS, OMIM #606777) is an autosomal dominant disorder caused by haplo-insufficiency of *SLC2A1*, and results in impaired hexose transport into the brain. Previously, we generated an AAV vector in which *SLC2A1* was expressed under synapsin I promoter (AAV-h*SLC2A1*), and suggested that AAV-h*SLC2A1* administration improved motor function of heterozygous knock-out *Glut1* (GLUT1<sup>+/-</sup>) mice. In this study, we developed another AAV vector that approximates human physiological GLUT1 expression.

**Methods:** We generated the AAV vector that expresses GLUT1 under the human GLUT1 promoter (AAV-GLUT1). AAV-GLUT1 was administered to GLUT1<sup>+/-</sup> mice via intra-cerebroventricular injection (CNS-local;  $6.5 \times 10^{10}$  vg/mouse,  $3.25 \times 10^{12}$  vg/kg). We analyzed GLUT1 mRNA/protein expression in the brain and other major organs and cerebral microvasculature by lectin-staining. We also examined motor function using rota-rod test and glucose levels in the blood and cerebrospinal fluid (CSF). Additionally, we confirmed that exogenous GLUT1 protein distribution in the brain and other organs after intra-cardiac injection of AAV-GLUT1 ( $7.8 \times 10^{11}$  vg/mouse,  $3.9 \times 10^{13}$  vg/kg).

**Results and conclusions:** After intra-cerebroventricular injection of AAV-GLUT1, exogenous GLUT1 was mainly expressed in endothelial cells in the brain, and partially in neural cells and oligodendrocytes. Exogenous GLUT1 also expressed in the cerebral cortex, hippocampus, and thalamus. Cerebral microvasculature was increased compared to un-injected control GLUT1<sup>+/-</sup> mice. AAV-GLUT1 improved the motor function and CSF-glucose levels of GLUT1<sup>+/-</sup> mice. Exogenous GLUT1 expression was not detected in other organs after intra-cerebroventricular injection of AAV-GLUT1. Exogenous GLUT1 expression was maintained in endothelial cells and neural cells for 24 months. After intra-cardiac injection, modest levels of exogenous GLUT1 were detected throughout the brain. Exogenous GLUT1 was also detected in the liver and in the muscle of the lower limb. AAV-based vector is a promising approach for the treatment of GLUT1DS patients.

## PgmNr 433: Loss-of-function variants in >4 million humans suggest that partial LRRK2 inhibition is a safe therapeutic strategy for Parkinson's disease.

### Authors:

N. Whiffin<sup>1,2</sup>; I.M. Armean<sup>2,3</sup>; A. Kleinman<sup>4</sup>; J.L. Marshall<sup>2</sup>; E.V. Minikel<sup>2</sup>; J. Goodrich<sup>2,3</sup>; K.J. Karczewski<sup>2,3</sup>; B.B. Cummings<sup>2,3</sup>; L. Francioli<sup>2,3</sup>; K. Laricchia<sup>2,3</sup>; Q. Wang<sup>2,3</sup>; A. Guan<sup>4</sup>; B. Alipanahi<sup>4</sup>; P. Morrison<sup>4</sup>; M.A.S. Baptista<sup>5</sup>; K.M. Merchant<sup>5</sup>; J.S. Ware<sup>1,2</sup>; M. Daly<sup>2,3,6</sup>; T. Esko<sup>2,7</sup>; C. Hultman<sup>8,9</sup>; R.J.F. Loos<sup>9</sup>; L. Milani<sup>7</sup>; A. Palotie<sup>2,3,6</sup>; C. Pato<sup>10</sup>; M. Pato<sup>10</sup>; D. Saleheen<sup>11,12</sup>; P.F. Sullivan<sup>8,13</sup>; J. Alföldi<sup>2,3</sup>; P. Cannon<sup>4</sup>; D.G. MacArthur<sup>2,3</sup>; the 23andMe Research Team and the gnomAD Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Imperial College London, London, UK; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Massachusetts General Hospital, Boston, MA; 4) 23andMe, Inc.; 5) Michael J. Fox Foundation; 6) Institute for Molecular Medicine Finland, Helsinki, Finland; 7) University of Tartu, Tartu, Estonia; 8) Karolinska Institutet, Stockholm, Sweden; 9) Icahn School of Medicine at Mount Sinai, New York, NY; 10) SUNY Downstate Medical Centre, New York, NY; 11) University of Pennsylvania, Philadelphia, PA; 12) Center for Non-Communicable Diseases, Karachi, Pakistan; 13) University of North Carolina, Chapel Hill, NC

Human genetics is a powerful tool to study the phenotypic impact of the modulation and inhibition of candidate drug targets. In particular, variants causing loss-of-function (LoF) of protein-coding genes provide natural in vivo models of gene inactivation, and are thus valuable indicators of the potential for on-target toxicity of therapeutic inhibitors.

Missense variants in *LRRK2* that cause gain of kinase function are a known cause of Parkinson's disease, suggesting inhibition of LRRK2 kinase activity as a promising therapeutic strategy. The biological consequences of such inhibition, however, have not been well characterised in humans. Furthermore, toxicity concerns have arisen from preclinical homozygous knock-out studies in model organisms, which have shown abnormal lung, liver, and kidney phenotypes.

Here, we systematically analyse predicted LoF variants across 141,456 sequenced participants in the Genome Aggregation Database (gnomAD), 49,960 exome sequenced individuals from the UK Biobank, and >4 million genotyped participants from 23andMe. After thorough manual curation, we identify 1,458 carriers of 136 unique high-confidence heterozygous LoF variants in *LRRK2*. Through Western blotting of LoF carrier lymphoblasts and CRISPR-engineered cardiomyocytes we demonstrate that six LoF variants, representing 85.2% of the carriers, result in reduced LRRK2 protein levels.

We manually curate available phenotypic data and conduct a phenome-wide association study. These analyses show that heterozygous LoF of the *LRRK2* gene is not associated with reduced life expectancy or with any specific phenotype or disease state; in particular, there is no evidence of adverse lung, liver, or kidney phenotypes.

These results suggest that partial life-long reduction of LRRK2 protein levels is consistent with a

normal, healthy lifespan, and therefore, therapeutics that downregulate LRRK2 are likely to be well-tolerated by Parkinson's disease patients. Furthermore, our results demonstrate the value of large-scale genomic databases combined with deep phenotypic characterisation of human LoF carriers to target validation in drug discovery.

## PgmNr 434: Loss of peroxisomal ACOX1 induces autoimmunity whereas a *de novo* gain-of-function variant induces elevated ROS and Schwann cell loss.

### Authors:

H.L. Chung<sup>1,2,18</sup>; M.F. Wangler<sup>1,2,3</sup>; P.C. Marcogliese<sup>1,2</sup>; J. Jo<sup>2,4</sup>; T.A. Ravenscroft<sup>1,2</sup>; S. Sadeghzadeh<sup>5</sup>; D. Li-Kroeger<sup>1,2</sup>; R.E. Schmidt<sup>6</sup>; A. Pestronk<sup>6,10</sup>; J.A. Rosenfeld<sup>1</sup>; L. Burrage<sup>1</sup>; M.J. Herndon<sup>6</sup>; S. Chen<sup>1</sup>; A. Shillington<sup>7,8</sup>; M. Vawter-Lee<sup>8,9</sup>; R. Hopkin<sup>7,8</sup>; J. Rodriguez-Smith<sup>8,10</sup>; M. Hendrickson<sup>8,10</sup>; B. Lee<sup>1</sup>; A.B. Moser<sup>11</sup>; R.O. Jones<sup>11</sup>; P. Watkins<sup>11</sup>; J. Chae<sup>12,14</sup>; S. Mar<sup>13</sup>; C.E. Prada<sup>7,8</sup>; R. Bucelli<sup>15</sup>; S. Yamamoto<sup>1,2,3,16</sup>; H.K. Lee<sup>2,3,4,16</sup>; T.P. Vogel<sup>17</sup>; H.J. Bellen<sup>1,2,3,16,18</sup>; Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX 77030, USA; 3) Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, USA; 4) Department of Pediatrics, Section of Neurology, Baylor College of Medicine, Houston, TX 77030 USA; 5) Department of Psychology, Harvard University, Cambridge, MA 02138, USA; 6) Department of Pathology and Immunology, Division of Neuropathology, Washington University School of Medicine, St. Louis, MO 63110, USA; 7) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 8) Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH, USA; 9) Division of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 10) Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 11) Division of Neurogenetics, Kennedy Krieger Institute, Johns Hopkins School of Medicine, Baltimore, Maryland 21205, USA; 12) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea; 13) Department of Neurology, St. Louis Children's Hospital, Washington University School of Medicine, St. Louis, MO 63110, USA; 14) Department of Pediatrics, Seoul National University College of Medicine, Seoul, Republic of Korea; 15) Department of Neurology, Washington University School of Medicine, St. Louis MO 63110, USA; 16) Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA; 17) Department of Pediatrics, Section of Immunology Allergy and Rheumatology, Baylor College of Medicine, Center for Human Immunobiology, Texas Children's Hospital, Houston, TX 77030 USA; 18) Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA

---

ACOX1 (Acyl-CoA oxidase 1) encodes the rate-limiting enzyme in very-long-chain fatty acid (VLCFA)  $\beta$ -oxidation, and produces  $H_2O_2$  in peroxisomes. We report a *de novo* mutation (p.N237S) in individuals who exhibit different phenotypes than ACOX1 deficiency patients. Loss of *dACOX1* in flies is semi-lethal, and eclosing flies exhibit progressive motor deficits and abnormal immune responses which can be rescued by Bezafibrate, an inhibitor for ELOVL1. In contrast, expression of p.N237S causes a gain-of-function that stabilizes ACOX1 as a dimer, produces elevated levels of reactive oxygen species (ROS) in insulating glia, and leads to demise of glia and neurons. Similarly, probands carrying p.N237S exhibit loss of Schwann cells, motor and sensory neurons. Treatment of flies or primary Schwann cells with an antioxidant strongly suppresses these phenotypes. Our data show that peroxisomes play a critical role in insulating glia across phyla, that loss of ACOX1 causes an autoimmune disease, and that gain of function results in ROS-driven glial death.



## PgmNr 435: A comprehensive approach to developing biomarkers tracking the progression of Parkinson disease.

### Authors:

J. Zhu <sup>1</sup>; T. Carlile <sup>1</sup>; D. Sangurdekar <sup>1</sup>; D. Lin <sup>1</sup>; H. McLaughlin <sup>1</sup>; R. Challa <sup>1</sup>; P. Cullen <sup>1</sup>; C. Sun <sup>1</sup>; M. Weisz <sup>3</sup>; O. Goldstein <sup>3</sup>; A. Mirelman <sup>3</sup>; J. Cedarbaum <sup>2</sup>; N. Giladi <sup>3</sup>; A. Orr-Urtreger <sup>3</sup>; B. Zhang <sup>1</sup>; M. Penny <sup>1</sup>; S. John <sup>1</sup>; M. Lalioti <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Translational biology, Biogen, Boston, MA, USA; 2) Early Clinical Development, Biogen, Boston, MA, USA; 3) Tel Aviv Sourasky Medical center, Israel

---

Parkinson's disease (PD) is a complex neurodegenerative disorder characterized by motor and non-motor symptoms that worsen over time. Currently, there is no effective clinical and/or biological markers that can track PD disease progression.

To develop novel biological markers that can model disease characteristics, severity and progression, we are recruiting a cohort of 800 Ashkenazi Jewish individuals either manifesting or genetically at-risk for PD, and healthy controls for both cross-sectional and longitudinal studies. So far, we have recruited 620 subjects for the cross-sectional study and 135 of them will be followed longitudinally. Individuals include carriers and non-carriers of LRRK2, GBA mutations.

For all the subjects, we collected demographic data, baseline PD history and severity, clinical assessments, including MDS-UPDRS score, fluid biomarkers, digital data from wearable devices and imaging data from MRI and DaT-SPECT.

For genetic/genomic biomarker development, we collected both whole blood and peripheral blood monocytes from each subject and measured gene expression using RNA-Seq, chromatin accessibility using ATAC-Seq, and DNA methylation using MethylationEPIC BeadChip assays. Differential gene expression, chromatin accessibility and methylation will be assessed between different groups. Currently, gene expression data of whole blood and monocytes from first 400 subjects have passed quality control. Moderate correlation between gene expression and methylation with traditional measurement of PD progression (MDS-UPDRS score) was observed. The genomic biomarker candidates will be used in a multi-modal analysis with clinical, digital, imaging and fluid biomarker data in order to classify patients into subgroups that share the same characteristics and develop models that can track the severity and progression of the disease.

## PgmNr 436: Genetic spectrum of pediatric movement disorders and management implications: A single centre experience from India.

### Authors:

N. Rathod; A. Hegde; V. Patel

[View Session](#) [Add to Schedule](#)

**Affiliation:** PEDIATRIC NEUROLOGY, JASLOK HOSPITAL & RESEARCH CENTRE, MUMBAI, MAHARASHTRA, India

---

**Introduction:** The advent of next generation sequencing(NGS) has led to accurate genetic diagnosis and better understanding of the spectrum of pediatric movement disorders(MD). We aim to describe the genetic profile of our patients with Movement Disorder (MD).

**Methodology:** We analyzed 138 patients with primary MD. 89of them with suspected genetic etiologies underwent NGS. We classified them according to the International Parkinson and Movement Disorder Society Task Force recommendations. The results were correlated with their clinical profile, neuroimaging and family genetics, when available, to determine their significance.

**Results:**36/89, 32/89, 9/89, 10/89, 2/89 belonged to the dystonia, ataxia, hereditary spastic paraparesis(HSP), paroxysmal movement disorders(PMD) and parkinsonism group respectively. Causative genes were found in 55/89(61%).

Most common phenotype was dystonia with 19/36(52.7%) genetic positivity. DYT related genes(DYT5,DYT6,DYT11) were the most common while few others were PANK2, PLA2G6, GLB1 and HEXA. Large number of patients with clinically suspected neurotransmitter disorders with pharmacological responsiveness were genetically negative.24/32(75%) patients presenting predominantly with ataxia had positive results with the common genes being ATM,ITPR1,SLC2A1, APTX and POLR3A. 8/9(88%) with HSP were positive for causative genes like SPG11, SPG47, SPG3A, SPG33 and SPG7. Amongst patients with PMD, 3/10(30%) were positive with PNKD , CACNA1A , ATP1A3 gene mutations.1/2(50%) patients with parkinsonism was positivewith LRRK2 mutation.

**Conclusion:** In movement disorders, genetics helps largely with confirmation of diagnosis, prognostication, avenues for research, and antenatal protection. Genetic diagnosis provided disease specific treatment , highlighting the importance of early and specific genetic diagnosis .

## **PgmNr 437: Pharmacogenetic analysis of rs2205986 associated with glatiramer acetate-induced liver injury in multiple sclerosis patients.**

### **Authors:**

K.M. Reynolds <sup>1</sup>; J.T. Beales <sup>1</sup>; M.R. Rimmasch <sup>1</sup>; J.C. Denny <sup>2</sup>; M.F. Davis <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah, USA; 2) Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

---

Multiple sclerosis (MS) is an autoimmune disease that develops when immune cells attack the myelin sheaths coating neurons in the central nervous system. Disease modifying therapies (DMT) can help reduce the frequency of demyelination events and MS severity, but are associated with a variety of adverse drug reactions (ADR), including drug-induced liver injury (DILI). Physicians often monitor liver condition for patients on DMTs with liver enzyme tests that measure patients' levels of alkaline phosphatase (AlkP), aspartate aminotransferase (AST) alanine aminotransferase (ALT), and bilirubin (TBil). A study performed by Kowalec et. al in 2018 investigated the genetic elements of DILI linked to the MS medication interferon- $\beta$  and discovered variant rs2205986 is associated with an increased DILI risk due to elevated AlkP and AST levels ( $\beta = 0.35$ ,  $\beta = 0.73$ ). Another common MS DMT known to be connected with DILI is glatiramer acetate (GA). In order to test if rs2205986 is also associated with an increased DILI risk in GA patients, 375 patients with references to GA in their electronic medical records located in Vanderbilt University Medical Center's Synthetic Derivative were manually reviewed to extract GA treatment timelines. 58 Caucasian patients had both liver lab values taken during treatment periods and genotyping information available. For each patient, the highest value for each test was identified and standardized to the upper limit of normal. With these liver enzyme values, a linear regression analysis was performed in PLINK v1.90 with patient age, sex, and first four principal components as covariates. Associations between rs2205986 and elevated AST, ALT, and TBil levels were insignificant ( $P = 0.33$ ,  $P = 0.32$ ,  $P = 0.51$ ). rs2205986, however, was significantly associated with elevated AlkP levels ( $P = 0.01$ ). With each additional variant allele, AlkP levels increase on average by 0.27 units/L (95% confidence interval: 0.09 - 0.46). This discovered association could aid physicians in screening potential GA candidates to prevent DILI, as well as contribute to the growing interest in developing personalized treatment plans accounting for patients' genetic information.



## PgmNr 438: Antisense oligonucleotide therapy in a humanized mouse model of *MECP2* duplication syndrome.

### Authors:

Y. Shao <sup>1,2</sup>; Y. Sztainberg <sup>1,2</sup>; P. Jafar-Nejad <sup>3</sup>; F. Rigo <sup>3</sup>; H. Zoghbi <sup>1,2,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030; 2) Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Baylor College of Medicine, Houston, Texas 77030; 3) Ionis Pharmaceuticals, 2855 Gazelle Court, Carlsbad, California 92010; 4) Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas 77030; 5) Departments of Pediatrics, Baylor College of Medicine, Houston, Texas 77030,

---

**Background:** Many intellectual disability disorders are due to copy number variations, and to date there have been no treatment options tested for this class of diseases. *MECP2* duplication syndrome (MDS) is one of the most common genomic rearrangements in males and results from duplications spanning the methyl-CpG binding protein 2 (*MECP2*) gene locus. Previously, we have shown that antisense oligonucleotide (ASO) therapy can successfully reduce MeCP2 levels in an MDS mouse model and reverse the disease-like phenotypes. However, our previous MDS mouse model carried one transgenic human allele and one mouse allele, with the latter being protected from human specific *MECP2*-ASO targeting. In humans, the two *MECP2* alleles are identical and because MeCP2 is a dosage-sensitive protein, one must ensure that the ASO is titrated to target the human allele such that MeCP2 levels are reduced from 2X to 1X. **Methods:** We generated a new "humanized" mouse model of MDS, that carries two human *MECP2* alleles, and no mouse endogenous allele. Moreover, we tested the effects of a human-specific *MECP2*-ASO through intracerebroventricular injection, which approximates that intrathecal delivery approach in humans. **Results:** The humanized mouse model of MDS showed elevated MeCP2 levels and using various behavioral tests we found they recapitulated the majority of the human patient phenotypes. We also found that the *MECP2*-ASO efficiently downregulates MeCP2 expression throughout the brain by quantifying the RNA and protein levels in 7 different brain regions. Furthermore, *MECP2*-ASO in a dose-dependent manner decreases MeCP2 levels in the brain. *MECP2*-ASO also ameliorates behavioral deficits including exploratory, learning and memory as well as motor coordination behavior, without any dose-limiting toxic effect or safety concern. We characterized the pharmacodynamic effect of the *MECP2*-ASO on MeCP2 and selected MeCP2-regulated genes during the duration of the treatment and found that majority normalized. **Conclusion:** Our results demonstrate that central nervous system administration of *MECP2*-ASO is well tolerated, has beneficial effects, and support a feasible translatable approach for the treatment of MDS.

## PgmNr 439: Epilepsy and autism spectrum disorder: Sequential plasmatic level and global quality of life assessment in antiepileptic drug treatment in Mali.

### Authors:

F.F. Doumbia <sup>1,3</sup>; M. Sangare <sup>1,3</sup>; M.S. Haidara <sup>1</sup>; M. Kouyate <sup>1</sup>; O. Sidibe <sup>1</sup>; H. Dolo <sup>1</sup>; S.S. Diakite <sup>1,3</sup>; A. Karembe <sup>1</sup>; H.G. Kim <sup>2</sup>; M. Diakite <sup>1</sup>; G. Awandare <sup>3</sup>; D. Geschwind <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculty of Medicine and Odontostomatology (FMOS), University of Sciences, Techniques and Technologies of Bamako (USTTB), Bamako, Mali; 2) Neurological Disorder Center, Qatar Biomedical Research Institute (QBRI), Hamad Bin Khalifa University, Doha, Qatar; 3) West African Centre for Cell Biology of Infectious Pathogens (WACCBIP), University of Ghana, Accra, Ghana; 4) Center for Autism Research and Treatment, Semel Institute, UCLA, LA, USA

---

**Introduction:** Epilepsy and Autism Spectrum Disorder (ASD) are the two faces of a coin. We hypothesized that appropriate antiepileptic treatment will result in a better quality of life for epileptic patients with or without ASD.

**Aim:** To evaluate the quality of life of epileptics with/out ASD on antiepileptic drugs (AED) in Mali.

**Methods:** We followed 90 epileptics with/out ASD from October 2017 to May 2019. We recorded the duration, intensity, and the frequency of epileptic seizures before and during treatment (M1 at M3) and used the McGill to assess the global quality of life. Plasmatic AED levels were determined at first month (M1) and third month (M3) timepoints after treatment initiation.

**Results:** Comorbidity ASD/epilepsy was 8.9%. The age group 16-30 years represented 50%. Our study participants received phenobarbital (73.3%), valproic acid (7.8%) and carbamazepine (18.9%). Side effects occurred in 13.3% of participants. With 79% treatment adherence rate, the AED was under-dosed (42%), normo dosed (50%) or overdosed (8%). At M1, epileptic seizures occurred <1/month in 26.7%, had minimal intensity in 94.4%, and lasted <1 minute in 16.7 %. The global quality of life was better 94.5%. The plasmatic levels of AED were different from M1 to M3  $p < 0,000$ .

**Discussion:** The non-observance rate of the AED treatment was non-negligible. Nevertheless, our data show that AED prescription and follow up based only on body weight does not guarantee adequate plasma levels to control seizures. Epileptics with/out ASD will benefit from sequential AED, Electroencephalogram (EEG) and quality of life assessment. The occurrence of side effects should prompt for screening for known polymorphisms (SNPs) in Cytochrome CYP450 genes. Altogether, our ultimate goal is to revolutionize the EAD treatment for personalized medicine in Mali.

## PgmNr 440: A pilote animal model study for genetic vaccine of addiction, sensitivity reduction of methamphetamine by using crispr cas9 technology.

### Authors:

A. Haghighatfard; R. Ghamari

[View Session](#) [Add to Schedule](#)

**Affiliation:** department of genetic, Tehran medical sciences branch, Islamic Azad university, Tehran, Iran

Methamphetamine is a strong addictive neurotoxin that stimulates the central nervous system and affects several pathways and metabolisms in different organs. Recently, the clustered regularly interspaced short palindromic repeats-CRISPR-associated protein 9 (CRISPR-Cas9) systems was developed, which improves sequence-specific gene editing in cell lines, organs, and animals. CRISPR system would selectively disrupt the targeted location without affecting other region of DNA or other RNAs. Several long non-coding RNAs (lncRNAs) were found associated with tendency and high sensitivity to methamphetamine abuse. lncRNA named AK036791 is related to depression and sensitivity to stimulants such as methamphetamine.

We set out to design and use a CRISPR system to disrupt expression of AK036791 in a mouse model of methamphetamine exposure. Two group of male mouse were used (30 wild-type C57BL/6 mice for each group, 7-8 weeks old, male, 20-25 g). The groups of mouse were then given once-daily injections of METH (2 mg/kg) or saline for five consecutive days (day 3-7) followed by two injection-free days (day 8-9). On day 10, mouse were given a challenge injection of either 2 mg/kg METH or saline. Test group were CRISPR/Cas9 knock-in mouse model to induce doxycycline-regulated Cas9 induction that enables widespread gene disruption in multiple tissues and that limiting the duration of Cas9 expression. Forced swimming test performed in both group and expression level of AK036791 was assessed in nucleus accumbens by using quantitative Real time PCR.

Findings revealed showed three time reduction in expression level of AK036791 in nucleus accumbens of test group ( $p < 0.003$ ). Also Forced swimming test results showed depression level of knocked in mouse were significantly ( $p < 0.002$ ) lower.

It seems that designed CRISPR/Cas9 could induce reduction of expression of genes that are involved in methamphetamine sensitivity and depression. It may lead to new kind of vaccination for addiction as well as reduction of dependency and depressive symptoms. CRISPR/Cas9 also can use for developing knock-in and knock-out mice that may help to shed light to molecular mechanisms of methamphetamine in central nervous system.

## **PgmNr 441: Translating behavioral epigenetics into the primary care setting.**

### **Authors:**

J.R. Murphy <sup>1,2</sup>; E. Corwin <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Laney Graduate School, Emory University, Atlanta, GA.; 2) School of Nursing, Emory University, Atlanta, GA

---

The concept of screening the epigenome for patterns of methylation and unique biomarkers that underlie common disorders is advancing in the literature. Technologies such as Illumina Infinium HumanMethylation450 BeadChip array may provide a mechanism by which scientists and clinicians can collaboratively screen, diagnose, and initiate precise treatment modalities for neurodevelopmental and psychiatric disorders of childhood. In addition, the field of behavioral epigenetics provides evidence by which clinicians understand the gene by environment influences that associate with clinical symptoms of mood and behavioral disturbance. Primary care advanced practice nurses (APRN) are ideally positioned to conduct risk assessments, employ novel epigenetic screening technologies, and make diagnoses within a biopsychosocial framework that more precisely captures the dynamic course of disease and illness. In addition, APRN's are trained in therapeutic communication and may initiate counseling on the psychosocial and ethical implications of undergoing a genetic evaluation. A conceptual model for translating behavioral epigenetics into the primary care setting is proposed.

## PgmNr 442: Tryptophan hydroxylase 1 and 2 genetic polymorphisms implicated in childhood onset high-aggression.

### Authors:

A. Singh <sup>1,2</sup>; C.C. Zai <sup>1,2</sup>; E. Koyama <sup>1</sup>; N. Freeman <sup>1</sup>; S.A. Shaikh <sup>1</sup>; M. Tampakeras <sup>1</sup>; K. Vieux <sup>1</sup>; S.Y. Cheema <sup>1</sup>; J. Beitchman <sup>1,2,3</sup>; J.L. Kennedy <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Brain Science, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

---

The serotonin neurotransmission system has been implicated in childhood onset-aggression, but its role in the development of aggressive behaviours is unclear. The first and rate-limiting step of serotonin biosynthesis involves tryptophan hydroxylase (*TPH*). The purpose of this study was to investigate whether two single nucleotide polymorphisms (SNPs), rs1800532 (A218C) in the tryptophan hydroxylase 1 (*TPH1*) gene and rs4570625 (G703T) in the tryptophan hydroxylase 2 (*TPH2*) gene, are significantly associated with aggressive behaviour in a sample of children with a history of extreme, persistent and pervasive aggression. These SNPs were also assessed for possible interactions between each other, anxiety, and a history of maltreatment in association with internalizing and externalizing behaviours. From the currently compiled dataset at the Centre for Addiction and Mental Health in Toronto, Canada, the 2 SNPs were genotyped in 141 high aggression cases and 169 healthy controls. SNP genotype was analyzed for association to aggressive behaviour, anxiety, maltreatment history and internalizing/externalizing behaviours using chi-squared tests, linear, and logistic regression models. The T/T and G/T genotypes of rs4570625 were significantly associated with aggressive behaviour ( $p=0.024$ ), suggesting that T allele carriers at this SNP are at higher risk for the development of such behaviours in childhood. Externalizing behaviour in the high-aggression cases was also associated with both rs1800532 ( $p=0.006$ ) and rs4570625 ( $p=0.015$ ). In addition, for T-allele carriers of rs4570625, a history of maltreatment was associated with externalizing behaviour ( $p=0.013$ ). Although anxiety was not associated with either SNP, more females were affected by anxiety in the HA cases, suggesting that males and females have different comorbidities and different mechanisms in the development of aggressive behaviours. These results suggest that *TPH2* genetic variations and environmental factors, such as maltreatment, play important roles in the development of childhood-onset aggressive and related behaviours. More studies are required to further understand the development of such behaviours, which may lead to implications for potential means of intervention.

## PgmNr 443: A genetic screen in *Drosophila* to identify new human disease candidate genes associated with altered dopamine homeostasis.

### Authors:

F. Shi <sup>1</sup>; S.L. Deal <sup>2</sup>; D. Bei <sup>3</sup>; E.S. Seto <sup>4</sup>; S. Yamamoto <sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Rice University, Houston, TX; 2) Program in Developmental Biology, Baylor College of Medicine, Houston, TX; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX

---

The regulation of the neuromodulator dopamine (DA) is a vital component to many biological functions. In humans, dysregulation of DA leads to social and behavioral deficits as well as many debilitating neurological disorders. The genes required for DA synthesis and secretion are well-conserved, allowing model organism researchers to study the cellular and molecular processes behind DA regulation in vivo. In addition to DA's use in the nervous system, insects use DA as a precursor of melanin to color and harden their cuticle. Through a genome-wide RNAi screen in *Drosophila*, we identified 136 genes that affect cuticle pigmentation. Further experiments using the *Drosophila* activity monitor system revealed that many of these genes affect behavior when knocked down in DArgic cells, indicating they may be involved in DA signaling in the nervous system in addition to their role in cuticle pigmentation. Interestingly, ~85% of the genes identified from this screen are conserved in mammals, suggesting that defects in these genes may lead to alterations in DA signaling in humans.

In this study, we visualized the expression of a subset of our hits by using the T2A-GAL4 system to determine the expression of these genes in DArgic neurons. We expressed a CD8::GFP reporter construct in the expression domain of the genes of interest, dissected the fly brains, performed co-immunostaining with a TH antibody, and visualized the signal with a confocal microscope. For five *Drosophila* genes tested, which corresponds to eight human genes [pfk (PFKL, PFKM, PFKL), lola (ZBTB20), ctrip (TRIP12), CG4328 (LMX1A, LMX1B), and CG16758 (PNP)], expression was found in a subset of DArgic neurons. The selective expression of these genes suggests they may not have global effects on DA regulation, but may instead have cell-specific effects. As different DArgic neurons play specific roles in the brain, identifying the genes responsible for cell-specific DA regulation will allow us to examine the molecular basis behind behavioral deficits caused by defects in specific DA neuronal populations.

## PgmNr 444: Investigating the burden of treatable genetic diseases in psychiatric populations: A novel avenue for precision medicine.

### Authors:

V. Sreiretnakumar <sup>1</sup>; C. Zai <sup>1</sup>; J. Kennedy <sup>1</sup>; J. So <sup>1, 2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurogenetics Section, Centre for Addiction and Mental Health (CAMH), Toronto, Ontario, Canada;  
2) The Fred A. Litwin Family Centre in Genetic Medicine, University Health Network & Mount Sinai Hospital, 60 Murray Street, Toronto, Canada, M5T 3L9

---

Genetic diseases are individually rare but collectively common. Many genetic conditions can mimic mental health disorders, with psychiatric symptoms that are difficult to treat with regular medications. Discovery of rare genetic diseases in psychiatric patients would reveal specific management options, including treatments that may ameliorate psychiatric symptoms or help regular medications to work better, and provide information about the chances of other family members being affected. In a previous pilot study of 2046 psychiatric patients we found an enrichment of variants associated with four treatable genetic diseases, which included Niemann-Pick disease type C, Wilson disease, homocystinuria, and acute intermittent porphyria (AIP), including a 240X enrichment of AIP in comparison to the general population. In this study, we expand our investigation of treatable genetic disorders in psychiatric populations (n=2304) by screening for pathogenic variants associated with 108 treatable genetic disorders using next-generation sequencing (Illumina HiSeq 2500/NovaSeq 6000). Specifically, SCZ (n=445), bipolar (n=568), major depression (n=304), obsessive compulsive disorder (n=368) and generalized anxiety disorder (n=619) cohorts will be screened. Re-contactable patients will be prioritized to facilitate clinical molecular and biochemical diagnostic follow-up for validation of genetic screening results. We expect to see an increased prevalence of treatable genetic disorders within the psychiatric population relative to the general population. Additionally, we expect an increased carrier frequency amongst psychiatric patients that could potentially be playing a role the pathophysiology of psychiatric disorders. Importantly, patients found to be carriers or affected with treatable genetic diseases will be followed in the clinical setting and given appropriate genetic counseling, where appropriate. Discovering genetic diseases in psychiatric patients will shift how health care is delivered to these vulnerable patients by addressing underlying conditions rather than masking symptoms with medications and will especially help patients who don't respond to regular medications. This will lead to significant cost savings to the health care system. Ultimately, this study will pave the way for the identification of specific characteristics associated with psychiatric symptomatology in treatable genetic diseases that will allow for earlier targeted screening and treatments.



## **PgmNr 445: Mood stabilizing medications affect gene expression, development, and calcium signaling in iPSC derived neuronal derivatives.**

### **Authors:**

Jiang <sup>1</sup>; K. Johnson <sup>2</sup>; L. Kassem <sup>1</sup>; W. Zheng <sup>3</sup>; W. Corona <sup>1</sup>; S.D. Detera-Wadleigh <sup>1</sup>; S. Christopher <sup>1</sup>; F.J. McMahon <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) NIMH/NIH, Bethesda, MD.; 2) NINDS/NIH, Bethesda, MD; 3) NCATS/NIH, Rockville, MD

---

Bipolar Disorder (BD) is a genetically heterogeneous psychiatric disorder of unknown etiology that presents major challenges for the study of disease biology and drug development. Induced pluripotent stem cells (iPSCs), which can be differentiated into neurons and glia, provide a cellular model system suitable for studies of cases and controls as well as studies that assess the impact of therapeutic agents on cellular development and differentiation. As part of the Amish-Mennonite Bipolar Genetics study (AMBiGen), we are collecting clinical data, DNA, and fibroblasts from probands diagnosed with BD and their relatives, all ascertained within genetically-isolated Amish and Mennonite communities in the Americas. Fibroblasts have been reprogrammed to iPSCs with Sendai virus and differentiated into neural progenitor cells (NPCs), neurons, and astrocytes using standard protocols. Pilot studies are underway in cells from 4 probands and 4 sex-matched, unaffected siblings (2 clones each).

Assessments include morphology, action potentials, gene expression profiles, and cellular response to therapeutic dosages of established treatments such as lithium and valproic acid (VPA). Preliminary results suggest that long-term treatment with VPA, but not lithium, greatly reduced proliferation of NPCs by the MTT assay, and promoted neuronal differentiation in both cases and controls. Long-term treatment with lithium greatly reduced neuronal calcium response to glutamate stimulation in both cases and controls, based on single-cell calcium imaging. Gene expression microarray data showed that cells in different differentiation stages demonstrated the most significant differences, some genes were differentially-expressed in cases compared to controls. VPA treatment of neural progenitor cells increased neurogenesis and growth of neurites. In addition, long-term VPA treatment during the process of differentiating neural progenitor cells into neurons can greatly reduce the number of differentially-expressed genes in cases versus controls, which may reflect some therapeutic effects of VPA. These preliminary results, if replicated, could point to cell-autonomous phenotypes in neural cells derived from people diagnosed with BD



## PgmNr 446: Body mass index influences a nicotine metabolism biomarker: A Mendelian randomization study in European ancestry smokers.

### Authors:

H.K. Giratallah<sup>1,2</sup>; J. Pouget<sup>1,2,4</sup>; M.J. Chenoweth<sup>1,2</sup>; A. El-Boraie<sup>1,2</sup>; C. Lerman<sup>3</sup>; J. Knight<sup>4,5</sup>; R.F. Tyndale<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pharmacology and Toxicology, University of Toronto, Toronto, Canada; 2) Campbell Family Mental Health Research Institute, CAMH, Toronto, Canada; 3) Department of Psychiatry, Annenberg School for Communication, and Abramson Cancer Center, University of Pennsylvania, Philadelphia, United States; 4) Department of Psychiatry, University of Toronto, Toronto, Canada; 5) Data Science Institute, Lancaster University, Lancaster, UK

---

**Background:** CYP2A6, a genetically variable enzyme that metabolizes nicotine and tobacco-specific nitrosamines, alters smoking behaviours and the risk for tobacco-related diseases such as COPD and lung cancer. CYP2A6 activity is phenotyped by the Nicotine Metabolite Ratio (NMR), a ratio of nicotine's metabolites (3'-hydroxycotinine/ cotinine). NMR is negatively correlated with body mass index (BMI), but the biology underlying this relationship is unclear. Smoking is also negatively associated with BMI; higher NMR increases smoking quantity, and could thereby reduce BMI. Alternatively, higher BMI could alter nicotine and metabolite kinetics and resulting NMR. Taylor *et al.*, 2018 provided some evidence for an influence of BMI on NMR in some but not all cohorts. Here we used Mendelian Randomization (MR) with updated instrumental variables to evaluate both potential causal influences, in order to determine whether NMR causes altered BMI or vice versa. **Methods:** Our clinical sample comprised 935 European-ancestry smokers from a cessation trial (PNAT2 NCT01314001). As an instrument for BMI, we performed LD clumping via PLINK to identify LD-independent SNPs from 941 conditional SNPs identified in a large genome-wide association study, which captured ~6% of the variation in BMI (Yengo *et al.*, 2018). As an instrument for NMR, we used our 7-variant CYP2A6 wGRS which captured ~34% of the variation in NMR (El-Boraie *et al.*, 2019). We used one-sample, bidirectional MR on individual-level data. Analyses were adjusted for relevant covariates. Sensitivity analyses were performed using MR-Egger regression to assess the validity of the resulting MR models. **Results:** LD clumping produced 334 SNPs which we subsequently used as instrumental variables for BMI in the MR models. We detected a significant causal influence for BMI on NMR ( $p \leq 0.001$ ) but no causal effect for NMR on BMI. Sensitivity analyses revealed consistent directions of effect, and the MR Egger intercept was non-significant ( $p > 0.1$ ) indicating no clear evidence of horizontal pleiotropy. **Conclusions:** Using MR we provide evidence for a causal relationship between BMI and NMR (i.e. higher BMI causes lower NMR). In contrast, we did not detect evidence that the NMR causally alters BMI. Our findings raise the possibility that BMI alters the kinetics of one or both metabolites that comprise the NMR.

## **PgmNr 447: Genetic polymorphisms in *CYP2D6*: Antipsychotic-induced weight gain in the UK Biobank sample.**

### **Authors:**

I. Austin-Zimmerman; J. Thygesen; H. Irizar; B. Wang; Y. Wannasuphoprasit; A. McQuillin; E. Bramon

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Psychiatry, University College London, London, United Kingdom

---

Management of the metabolic side effects induced by antipsychotics are an important concern in the successful treatment of patients with psychosis. *CYP2D6* is a major metabolic pathway for many antipsychotics. There is growing evidence demonstrating a relationship between *CYP2D6* genotype and clinical outcomes. This study aims to consider whether *CYP2D6* metabolic status influences antipsychotic-induced weight gain.

The UK Biobank is a large, population cohort study including approximately 500,000 volunteers (age 40-70). Participants underwent extensive baseline assessments, including genome-wide genotyping and clinical interviews. We extracted genetic and clinical data for all individuals taking antipsychotics. Genetic data were phased using PHASE software to allow for star-allele allocation according to the Pharmacogene Variation Consortium guidelines. The sample was categorised into three phenotypic groups based on their Gaedigk activity score: poor, intermediate and extensive metabolisers. Ultra-rapid metabolisers were combined with extensive in this analysis. Linear regression models were conducted with *CYP2D6* metabolic phenotype as the predictor, body mass index (BMI) and waist circumference as the outcome measures, and age, sex and whether or not the drug taken was a primary *CYP2D6* substrate included as covariates.

We identified 2,863 individuals within the UK Biobank taking antipsychotics. Of these, 143 were poor metabolisers, 218 intermediate and 2,502 extensive. Overall, no significant association between *CYP2D6* phenotype and increased BMI or waist circumference was identified. We did find that taking a drug metabolised primarily through *CYP2D6* was predictive of higher BMI and waist circumference. When we conducted the analysis on 24 individual drugs rather than all antipsychotics together, we found a significant association between increased BMI and waist circumference in poor compared to extensive metabolisers for individuals taking fluphenazine.

The sample included in this paper constitutes one of the largest studies of *CYP2D6* poor and intermediate metabolisers conducted thus far. We did not identify any significant association between *CYP2D6* phenotype and increased BMI or waist circumference. Our results suggest that antipsychotic-induced weight gain is driven more by the type of antipsychotic drug than by *CYP2D6* metabolic phenotype. Further investigation of the impact of metabolic status on response to specific psychotropic drugs is required.

## PgmNr 448: X-chromosome variation influences on a nicotine metabolism biomarker predictive of smoking behaviors differ by sex in African American cigarette smokers.

### Authors:

M.J. Chenoweth <sup>1</sup>; L.S. Cox <sup>2</sup>; C. Lerman <sup>3</sup>; J. Knight <sup>4,5</sup>; R.F. Tyndale <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health and the Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada; 2) Department of Population Health, University of Kansas School of Medicine, Kansas City, KS, USA; 3) Department of Psychiatry, Annenberg School for Communication, and Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA, USA; 4) Data Science Institute and Lancaster University Medical School, Lancaster, UK; 5) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada

---

Cigarette smoking behaviors, which differ between men and women, are also influenced by pharmacogenetic variation. The genetically polymorphic CYP2A6 enzyme inactivates nicotine, the major psychoactive compound in cigarettes, to cotinine and then to 3'hydroxycotinine. The Nicotine Metabolite Ratio (NMR; ratio of 3'hydroxycotinine/cotinine) is a heritable (~60-80%) biomarker of the rate of nicotine metabolism. The NMR varies with sex and is associated with smoking behaviors and disease risk. Previous genome-wide association studies of the NMR that controlled for sex and examined only the autosomes (chromosomes 1 - 22) captured up to 38% of NMR variation. We investigated whether X-Chromosome genetic variation is associated with the NMR in African American smokers from two smoking cessation clinical trials (NCT01314001 and NCT00666978). Standard GWAS and X-Chromosome specific quality control steps were undertaken prior to genotype imputation. Males and females were analyzed separately. Within each trial, additive genetic models (phenotype = standardized square root NMR) adjusted for population substructure and NMR covariates (age, BMI, and use of menthol cigarettes). Results from the two individual trials were then meta-analyzed; variants with info scores >0.7 and minor allele frequencies (MAF) >1% were considered. In males (n = 351), the top variant was rs5963533 (beta = - 2.50 for A vs. G allele; P = 4.7e-9; MAF = 1.4%), explaining ~9.0% of NMR variation. The rs5963533 variant is located ~49kb downstream of *TSPAN7*, a gene previously associated with X-linked intellectual disability. In contrast to males, rs5963533 was not associated with the NMR in females (beta = - 0.22; P = 0.29). In females (n = 502), the top variant was rs148897912 (beta = 0.88 for G vs. A allele; P = 2.6e-6, MAF = 3.4%), explaining ~3.5% of NMR variation. The rs148897912 variant is located ~48kb upstream of *ENOX2*, which belongs to a family of growth-related NADH oxidases. In contrast to females, rs148897912 was not associated with the NMR in males (beta = 0.45; P = 0.21). Our findings suggest that X-Chromosome variation may represent an additional source of genomic variability in the NMR, with differing impacts in males and females. A greater understanding of the variation in this nicotine metabolism biomarker associated with smoking behaviors and tobacco-related disease risk may improve treatment strategies for both men and women.

## PgmNr 449: Evaluation of a weighted genetic risk score for the prediction of CYP2A6 activity in African-Ancestry smokers.

### Authors:

A. El-Boraie <sup>1</sup>; M.J. Chenoweth <sup>1</sup>; K. Fukunaga <sup>2</sup>; T. Mushiroda <sup>2</sup>; M. Kubo <sup>2</sup>; J. Knight <sup>3,4</sup>; J.G. Pouget <sup>4,5</sup>; C. Lerman <sup>6</sup>; N.L. Nollen <sup>7</sup>; L.S. Cox <sup>7</sup>; R.F. Tyndale <sup>1,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada; 2) Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 3) Data Science Institute and Medical School, Lancaster University, Lancaster, UK; 4) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 5) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health and Division of Brain and Therapeutics, Toronto, Ontario, Canada; 6) Department of Psychiatry and USC Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, USA; 7) Department of Population Health, University of Kansas School of Medicine, Kansas City, Kansas, USA

**Background:** The Nicotine Metabolite Ratio (NMR, 3-hydroxycotinine/cotinine), an index of nicotine metabolic inactivation by CYP2A6, is associated with smoking behaviors and cessation outcomes. However, the NMR cannot be measured in non-, former- or intermittent-smokers, for example in studies of tobacco-related disease risk or for other CYP2A6 drug substrates. The NMR is highly heritable (60-80%) and we previously curated a CYP2A6 weighted genetic risk score (wGRS) for European-ancestry populations (EUR) capturing 33.8% of NMR variance. Several loss-of-function CYP2A6 alleles are unique to African-ancestry populations (AFR), and ~60% of hits identified in an AFR-stratified NMR genome-wide association study (GWAS) were not shared with EUR, warranting the need for an AFR-based wGRS.

**Methods:** Training set: N=954 AFR smokers from two trials (PNAT-II; NCT01314001) and (KIS-III; NCT00666978); replication set: N=216 AFR smokers from a third trial (Q2L; NCT01836276). Variations of the wGRS model included independent GWAS signals along with known functional CYP2A6 \* alleles, calculated as the sum of risk alleles weighted by their effect sizes.

**Results:** The final model consisted of 11 genetic variants, 8 of which are unique to AFR. In the training and replication sets, the AFR wGRS explained 32.2% ( $P<0.001$ ) and 34.3% ( $P<0.001$ ) of the variance in the NMR, respectively. The AFR wGRS demonstrated a fair ability (receiver-operating characteristic curve=0.73;  $P<0.001$ ) to discriminate NMR slow from normal metabolizers. There was a negligible influence of known NMR covariates (e.g. sex, age, BMI) and ancestral population substructure (genomic principal components) on wGRS robustness (unique variance explained=32.0%;  $P<0.001$ ), despite the geographical diversity in recruitment sites in the training set. We combined wGRSs, previously described for EUR and described here for AFR, and using wGRS-stratified analyses were able to replicate previous unique NMR-stratified cessation outcomes between metabolizer groups: the interaction between treatment and NMR metabolizer group was 2.25 ( $P=0.02$ ), compared to 2.12 ( $P=0.03$ ) for treatment and predicted wGRS metabolizer group.

**Conclusions:** We present a useful approach for translating CYP2A6 genotypes into a gene activity score in AFR populations; the wGRS is robust to non-genetic covariates and fine population

substructure. Furthermore, the AFR wGRS is comparable to its EUR wGRS counterpart in its ability to capture NMR variation.

## **PgmNr 450: Pharmacological annotation of polygenic risk in individuals to direct precision intervention in complex disorders.**

### **Authors:**

M. Cairns <sup>1</sup>; W. Reay <sup>1</sup>; J. Atkins <sup>1</sup>; M. Green <sup>2</sup>; V. Carr <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) School of Biomedical Sciences and Pharmacy, The University of Newcastle, Callaghan, New South Wales, Australia; 2) The University of New South Wales, Sydney, Australia

---

Individuals with complex disorders typically have a heritable burden of common variation that can be indexed by the summation of their genic risk with respect to genome-wide association summary statistics, expressed as a polygenic risk score (PRS). While each component of polygenic risk has a relatively small effect size and are small targets for therapeutic intervention, we hypothesized that some individuals may harbour significant variation that is specifically enriched in clinically actionable pathways. To test this, we developed a framework, designated the pharmagenic enrichment score (PES), to quantify common variant enrichment in pathways with known drug targets. This approach was exemplified using a cohort of schizophrenia cases, using common variants associated with drug-oriented gene-sets derived from genome wide association study summary statistics. A large proportion of individuals had elevated PES in one or more of eight candidate clinically actionable gene sets identified, including a significant proportion of patients with relatively low genome-wide schizophrenia PRS, and a subgroup displaying treatment resistance. These observations suggest that the PES approach could provide a mechanism to integrate an individual's common variant risk to inform personalized interventions, including drug repositioning for complex disorders, such as schizophrenia.

## PgmNr 451: Epigenetic modification of hypothalamic-pituitary-adrenal (HPA) axis genes *CRHR1* and *CRHR2* may be associated with treatment response in clinical depression.

### Authors:

A.J. Lisoway<sup>1,2</sup>; C.C. Zai<sup>1,2,3,4</sup>; A.K. Tiwari<sup>1,3</sup>; J.L. Kennedy<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Brain Science, Campbell Family Mental Health Research Institute, Centre for Addiction & Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Science, University of Toronto; 3) Department of Psychiatry, University of Toronto; 4) Department of Laboratory Medicine and Pathobiology, University of Toronto

**Background:** Stress and dysregulation of the HPA axis plays an important role in the development and severity of depressive symptoms. Epigenetic modifications are known to significantly impact the relationship between stress system genes and the environment via the HPA axis but are currently under investigated in regard to their effect on clinical depression and antidepressant treatment. Therefore, we investigated DNA methylation sites across six well-known HPA axis genes and examined their associations with antidepressant treatment response in clinically depressed patients with and without comorbid anxiety.

**Methods:** Clinically depressed patients with complete follow-up data (n=157) were selected from our naturalistically treated sample (IMPACT). Patients were referred to our study for failure to adequately respond, or intolerance to, their psychiatric medication. Saliva samples were collected at enrollment and DNA methylation levels were interrogated using the Infinium HumanMethylation450 Beadchip. Thirty CpG probes located across the promotor regions of *CRHR1*, *CRHR2*, *FKBP5*, *NR3C1*, *NR3C2*, and *SKA2* were selected to quantify methylation levels. Change in depression symptom severity, as assessed by the Beck Depression Inventory (BDI-II), was used to measure antidepressant response and remission status over eight weeks. The relationship between methylation levels and antidepressant response/remission status was investigated using non-parametric tests, linear, or logistic regression.

**Results:** DNA methylation levels at two specific CpG sites within *CRHR1* and *CRHR2* nominally differed between responders/non-responders (p=0.06) and between remitters/non-remitters (p=0.05). Furthermore, when examining these CpG sites, remitters and responders had slightly lower methylation levels as compared to non-responders/non-remitters.

**Discussion:** Lower levels of DNA methylation within HPA axis related genes may be associated with response to antidepressant treatment and improvement in clinical depression. Generally, hypomethylation in promoter regions is associated with increased gene expression. Thus, lower methylation levels across promoter regions of *CRHR1* and *CRHR2* could increase expression of these genes and influence signaling within the feedback loop of the HPA axis. Further pharmacogenetic studies, incorporating genotypes of functional SNPs within these genes, are required to fully

understand the effect of DNA methylation changes on antidepressant treatment outcome.



## PgmNr 452: High rates of genetic diagnosis in psychiatric and neurodevelopmental patient populations: A practical tool to predict and guide genetic diagnosis.

### Authors:

J. So <sup>1,2,3,4,5</sup>; J. Suddaby <sup>1</sup>; V. Sriretnakumar <sup>2,3</sup>; B. Barsanti-Innes <sup>2</sup>; H. Faghfoury <sup>1,4</sup>; T. Gofine <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Fred A. Litwin Family Centre in Genetic Medicine, University Health Network and Mount Sinai Hospital, Toronto, Canada; 2) Centre for Addiction and Mental Health, Toronto, Canada; 3) Department of Laboratory Medicine and Pathobiology, University of Toronto, Canada; 4) Department of Medicine, University of Toronto, Canada; 5) Ontario Shores Centre for Mental Health Sciences, Whitby, Canada; Department of Psychiatry, University of Toronto, Canada

---

**OBJECTIVE:** Although studies of genetic diagnostic rates have been undertaken in patient populations with neurodevelopmental disorders (NDs) and neurological disorders, there is a paucity of literature on genetic diagnosis in psychiatric populations, particularly the vulnerable population of patients with concomitant ND and psychiatric impairment. In this prospective cohort study, we investigated the genetic diagnostic rate in 151 adult psychiatric patients from two centres in Ontario, Canada, including a large subset (73.5%) with concurrent ND, and performed phenotypic analysis to determine the strongest predictors for the presence of a genetic diagnosis.

**METHODS:** Patients 16 years of age or older and affected with a psychiatric disorder plus at least one of ND, neurological disorder, congenital anomaly, dysmorphic features or family history of ND were recruited through the genetics clinics between 2012 and 2016. Patients underwent genetic assessment and testing according to clinical standards. Chi-squared test was used for phenotypic comparisons. Binary logistic regression analysis was performed to determine which phenotypic features were predictive of genetic diagnosis types.

**RESULTS:** There was an overall genetic diagnostic rate of 45.7% in the total cohort, comprising primarily single gene (18.5%) and chromosomal (15.9%) disorders. Schizophrenia was significantly more common in patients with chromosomal disorders, while mood and anxiety disorders were significantly more common in those with single gene disorders. Regression analysis using statistically significant phenotypic associations showed that presence of a single gene disorder diagnosis was best predicted by the combination of having a neurological phenotype, specifically seizures and migraines, mood disorder and family history of ND, while presence of a chromosomal diagnosis was best predicted by the combination of neurological phenotype, specifically migraines, ND and schizophrenia.

**CONCLUSIONS:** Based on the results of this study, we propose a clinical tool to aid physicians working with the psychiatric patient population to determine which individuals would benefit from genetic assessment.

## PgmNr 453: The effect of startle-related genes on anxiety symptom severity at baseline and following pharmacological treatment.

### Authors:

J. Tomasi<sup>1,2</sup>; A.K. Tiwari<sup>1,3</sup>; C.C. Zai<sup>1,2,3,4</sup>; D. Herbert<sup>1</sup>; A. Shahmirian<sup>1</sup>; N. King<sup>1</sup>; N. Freeman<sup>1</sup>; J.L. Kennedy<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Brain Science, Centre for Addiction and Mental Health, Toronto, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, Canada; 4) Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

**Background:** Anxiety symptoms are observed across psychiatric disorders, with these symptoms leading to worse prognosis. Startle response magnitude increases during states of anxiety, and the presence of exaggerated startle reactivity is a strong indicator of anxiety disorder. Furthermore, initial startle reactivity predicts antidepressant treatment outcome in psychiatric populations. Several genes have shown associations with variations in startle reactivity, and the effect of these genes on anxiety symptom severity at baseline and following treatment across psychiatric patients has never been investigated. We hypothesized that startle-related genes will be associated with baseline anxiety symptom severity and/or the change in anxiety symptoms following 8 weeks of pharmacological treatment across psychiatric disorders, and in an anxiety disorder subset.

**Methods:** The sample consists of patients from the pharmacogenetic IMPACT study that provided salivary DNA samples and had anxiety symptom severity determined by the Generalized Anxiety Disorder 7-item (GAD-7) scale. We selected 19 genes and 22 SNPs associated with startle reactivity for a hypothesis-driven gene set analysis. Analyses were conducted in general psychiatric patients (N=508), and in the subsample of anxiety disorder patients (N=298). Linear regressions were performed to examine the effect of the SNPs on baseline GAD-7 score and the change in GAD-7 score after 8 weeks.

**Results:** Across psychiatric patients, two SNPs, one from *MAOA* and one from *NPSR1*, were nominally associated with baseline GAD-7 score ( $p(\text{unadjusted})=0.017, 0.023$ ). A SNP from *NLGN1* was nominally associated with the change in GAD-7 score after 8 weeks across the entire psychiatric sample ( $p(\text{unadjusted})=0.018$ ). Among anxiety disorder patients, nominally significant associations with change in GAD-7 score after 8 weeks were found for a SNP from *ANK3* and a SNP from *CNR1* ( $p(\text{unadjusted})=0.019, 0.043$ ).

**Conclusions:** A subset of genetic variants related to startle reactivity is associated with baseline anxiety severity and the change in anxiety severity following pharmacological treatment across psychiatric patients. This study provides support for exploring startle-related genes to identify biomarkers of pathological anxiety and to predict pharmacological treatment response for patients suffering with anxiety.

## PgmNr 454: Characterization of the retinal phenotype in a humanized FD mouse model with defective *ELP1* splicing.

### Authors:

A. Chekuri <sup>1,2</sup>; E. Morini <sup>1,2</sup>; X. Wang <sup>3</sup>; E. Logan <sup>1</sup>; A. Krauson <sup>1</sup>; M. Salani <sup>1</sup>; P. Dietrich <sup>4</sup>; I. Dragatsis <sup>4</sup>; L. Vandenberghe <sup>3</sup>; S. Slaugenhaupt <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Massachusetts General Hospital/Harvard Medical School, Massachusetts, USA.; 2) Department of Neurology, Massachusetts General Hospital Research Institute and Harvard Medical School, Boston, Massachusetts, USA; 3) Grousbeck Gene Therapy Center, Schepens Eye Research Institute and Massachusetts Eye and Ear Infirmary, Boston, Massachusetts, USA; 4) Department of Physiology, the University of Tennessee, Health Science Center, Memphis, TN, USA.

---

Familial dysautonomia (FD) is an autosomal recessive neurodegenerative disorder caused by a splice mutation in the gene encoding Elongator complex protein 1 (*ELP1*, also known as *IKBKAP*). This mutation results in skipping of exon 20 and tissue specific reduction of *ELP1* protein levels, predominantly in the central and peripheral nervous system. Although FD patients exhibit a complex neurological phenotype due to the degeneration of sensory and autonomic neurons, progressive retinal degeneration severely impacts quality of life. Two different mouse models have been previously generated to study the retinal phenotype that results from complete loss of *ELP1*. However, neither of these models accurately recapitulates the tissue specific defective splicing observed in FD patients. Our recently developed FD mouse model, *TgFD9; Ikbkap*<sup>Δ20/flox</sup>, was generated by introducing the human *ELP1* transgene with the FD mutation into a mouse heterozygous for a hypomorphic *ELP1* allele and a knock-out allele. This mouse exhibits most of clinical features of the disease and, importantly, accurately models the tissue specific splicing defects observed in FD patients. Detailed characterization of the retinal phenotype in this FD mouse model was performed to investigate the pathology associated with the splice mutation in *ELP1*. Optical Coherence Tomography (OCT) analysis to investigate retinal thickness revealed significant reduction in the thickness of the retinal nerve fiber layer (RNFL) when compared to control littermates. Analysis of retinal morphology revealed progressive loss of retinal ganglion cells (RGC's) and significant reduction in outer nuclear layer (ONL) thickness, indicating loss of photoreceptors. RGC cell counting using whole mount retina was also performed to substantiate the loss of RGC's. Our findings suggest that our novel FD mouse model will provide a platform in which to test the *in vivo* efficacy of *ELP1* splicing modulation to increase functional *ELP1* in the retina.

## PgmNr 455: Identifying the genetic cause of retinitis pigmentosa in a multigenerational pedigree using Extremely Low-Coverage Whole Genome Sequencing (XLC-WGS).

### Authors:

J. Lazaro Guevara <sup>1,2</sup>; B. Flores-Robles <sup>3</sup>; S. Frodsham <sup>1</sup>; E. Javier-Hernandez <sup>2</sup>; E. Labrador-Sánchez <sup>3</sup>; V. Pinillos-Aransay <sup>3</sup>; A. Elena-Ibáñez <sup>3</sup>; J. López-Martín <sup>3</sup>; M. Pezzolesi <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Nephrology and Hypertension, Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, Utah; 2) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah; 3) Department of Rheumatology, San Pedro Hospital, Logroño, Spain; 4) Diabetes and Metabolism Research Center, University of Utah School of Medicine, Salt Lake City, Utah

---

Retinitis Pigmentosa (RP) is a rare, genetic eye disorder that cause progressive vision loss. RP has a slow progression and late onset and, to date, more than 60 causal genes are known. RP diagnoses are made through ophthalmological examinations. The decreasing cost of next-generation sequencing is providing opportunities to leverage precision genomics to establish molecular diagnoses for a variety of human diseases. Given RP's clinical and genetic heterogeneity, genetic screening may be able to help pinpoint a genetic cause and facilitate its diagnosis. Because whole-genome sequencing (WGS) is still cost prohibitive in a clinical setting, we investigated the utility of extreme low coverage WGS (~0.5X; XLC-WGS) as a cost-effective way to identify causal variants in a multigenerational pedigree with RP. XLC-WGS was performed in 17 members of this pedigree, including 3 individuals with clinical diagnosis of RP. The XLC-WGS data was processed using Illumina's DRAGEN pipeline and filtered using various genotype quality scores (GQX; no filter, GQX ≥ 10, or GQX ≥ 30. The resulting variants from all 3 analyses were passed to 2 independent variant interpretation tools: Expert Variant Interpreter from enGenome and Variant Interpreter from Illumina; both analyses applied ACMG-AMP guidelines. Both programs identified a nonsense mutation (c.1625C>G; p.Ser542Ter) in exon 4 of the retinitis pigmentosa 1 (*RP1*) gene on chromosome 8 as the most likely causal variant. p.Ser542Ter has previously been shown to be causative of a recessive form of RP. XLC-WGS identified 2 homozygous p.Ser542Ter carriers among the 3 sequenced RP cases. This region had insufficient coverage to identify this variant in the remaining case. In addition to these homozygous carriers, 3 non-affected pedigree members were found to be heterozygous for this variant. Improved sensitivity in detecting this variant was observed with the less stringent genotype quality score filters. Although these less stringent filters led to less confident variant calls, coupled with bioinformatic tools to guide variant interpretation, XLC-WGS was able to identify the disease-causing variant in this family. In conclusion, in the context of multigenerational pedigrees, where multiple affected and non-affected individuals are available, we have shown that XLC-WGS is a cost-effective approach that may have clinical utility in detecting pathogenic disease variants that co-segregate in families enriched for genetic disease.

## PgmNr 456: Concurrent hearing, genetic and HCMV screening in a general newborn population.

### Authors:

Z. Peng<sup>1</sup>; Q. Wang<sup>2</sup>; J. Xiang<sup>1</sup>; J. Sun<sup>3,4</sup>; Y. Yang<sup>1</sup>; J. Guan<sup>2</sup>; D. Wang<sup>2</sup>; C. Song<sup>5,6</sup>; L. Guo<sup>7</sup>; H. Wang<sup>2</sup>; Y. Chen<sup>8</sup>; J. Leng<sup>8</sup>; X. Wang<sup>9</sup>; J. Zhang<sup>3</sup>; B. Han<sup>2</sup>; J. Zou<sup>10</sup>; C. Yan<sup>1</sup>; L. Zhao<sup>2</sup>; H. Luo<sup>1</sup>; Y. Han<sup>11</sup>; W. Yuan<sup>11</sup>; H. Zhang<sup>9</sup>; W. Wang<sup>12</sup>; X. Xun<sup>13,14</sup>; Y. Yin<sup>1</sup>; C. Morton<sup>5,15,16</sup>; L. Zhao<sup>9</sup>; S. Zhu<sup>13,14,17</sup>; J. Shen<sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) BGI Genomics, BGI-Shenzhen, Shenzhen, Guangdong, China; 2) Department of Otolaryngology-Head and Neck Surgery, Chinese PLA Institute of Otolaryngology, Chinese PLA General Hospital, Beijing, China; 3) Tianjin Medical Laboratory, BGI-Tianjin, BGI-Shenzhen, Tianjin, China; 4) Binhai Genomics Institute, BGI-Tianjin, BGI-Shenzhen, Tianjin, China; 5) Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 6) Children's Hospital of Chongqing Medical University, Chongqing, China; 7) Jining Maternal and Child Health Care Service Center, Jining, China; 8) Tianjin Women and Children's Health Centre, Tianjing, China; 9) BGI Clinical Laboratory, BGI-Shenzhen, Shenzhen, China; 10) MGI, BGI-Shenzhen, Shenzhen, China; 11) Wuhan BGI Clinical Laboratory, BGI-Shenzhen, Wuhan, China; 12) BGI-Beijing, BGI-Shenzhen, Beijing, China; 13) BGI-Shenzhen, Shenzhen, China; 14) China National GeneBank, BGI-Shenzhen, Shenzhen, China; 15) Manchester Center for Audiology and Deafness, School of Health Sciences, University of Manchester, Manchester, UK; 16) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 17) Shenzhen Engineering Laboratory for Innovative Molecular Diagnostics, Shenzhen, China

---

Current universal newborn hearing screening programs do not be designed to detect mild, late-onset, and drug-induced hearing loss, nor reveal genetic etiology. The NGS-based test can identify human single nucleotide variations and viral sequence from blood, facilitating the identification of hereditary and acquired hearing loss patients.

In one of our recent study, we screened newborns for 20 hearing loss-related genetic variants from 2012 to 2017. Then, we followed 12,778 genetically non-negative newborns and revealed a higher rate of hearing loss by three months of age among referrals from the initial hearing screening (60% vs. 5.0%,  $P < 0.001$ ) and a lower rate of lost-to-follow-up/documentation (5% vs. 22%,  $P < 0.001$ ) in the genotype-positive group than in the genotype inconclusive group. Importantly, genetic screening detected 13% more hearing-impaired infants than hearing screening alone and identified 2,638 (0.23% of total) newborns predisposed to preventable ototoxicity undetectable by hearing screening. Our results proved that incorporating genetic screening improves the effectiveness of newborn hearing screening programs by elucidating etiologies, discerning high-risk subgroups for vigilant management, identifying additional children who may benefit from early intervention, and informing at-risk newborns and their maternal relatives of increased susceptibility to ototoxicity.

Besides, human cytomegalovirus (HCMV) is one of the main contributors for acquired hearing loss. Previous studies demonstrated when infants with  $>1 \times 10^4$  copies/ml of viral DNA in peripheral blood were at a higher risk for infection-related hearing loss. We then developed an NGS-based test that is capable of detecting deafness-related mutations and HCMV sequence simultaneously from a simulated dried blood spot. The availability of the test would facilitate the detection of HCMV in

newborns at birth without an extra cost. Our pilot results showed that the test is sensitive when the viral load is as low as  $1 \times 10^4$  copies/ml. Thus, we propose that the test, which concurrently detects deafness-related mutations and HCMV, is promising to compensate conventional hearing screening program. A more comprehensive clinical validity and utility analysis is warranted.

## PgmNr 457: *LTA4H* role to montelukast response in asthmatic children and young adults.

### Authors:

C. Maroteau <sup>1</sup>; G. Lassi <sup>1</sup>; S. Young <sup>2</sup>; A. Platt <sup>3</sup>; C.N.A Palmer <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Genomics Research, Discovery Science, BioPharmaceuticals R&D, AstraZeneca, Cambridge, U.K.; 2) Precision Medicine, Oncology R&D, AstraZeneca, Cambridge, U.K.; 3) Translational Science and Experimental Medicine, Research and Early Development, Respiratory, Inflammation and Autoimmune, BioPharmaceuticals R&D, AstraZeneca, Cambridge, U.K.; 4) Centre for Pharmacogenetics and Pharmacogenomics, Ninewells Hospital and School of Medicine, University of Dundee, Dundee, U.K.

---

Leukotrienes play a central pathophysiological role in asthma. Recent change in NICE guidelines recommend using leukotriene receptor antagonists (LTRA) as an alternative to low dose inhaled corticosteroid (ICS) for both paediatric and adult asthmatic patients. However, 35% to 78% of asthmatics do not respond to LTRA treatment. Several studies have shown that variants in leukotriene pathway candidate genes are associated with interindividual variability in response to montelukast, a selective cysteinyl leukotriene 1 (*cysLT1*) receptor antagonist.

To test the role of one of these variants, regulatory variant rs2660845 in *LTA4H*, in the response to montelukast, we selected and genotyped 88 asthmatics starting montelukast treatment before the age of 18 from the GoSHARE study cohort of Caucasian individuals in the Tayside and Fife areas of Scotland. Non-responders were defined as being on montelukast for at least a year with more than one exacerbation or discontinuing the drug. Responders were defined as taking montelukast for at least a year with no or one exacerbation. An association analysis of the candidate SNP rs2660845 with montelukast response was performed using logistic regression and adjusted for age, gender and date of first salbutamol medication. In addition, a fixed effect meta-analysis was performed on GoSHARE, BREATHE and Tayside RCT paediatric Scottish cohorts' results.

A trend was found between *LTA4H* genotype (regulatory variant rs2660845) and exacerbations in the GoSHARE cohort ( $p=0.062$ ). Response to montelukast in the A/A group was higher (OR = 9.4, 95% CI: 0.6752 – 3.216) than those in the G allele carriers (AG+GG). Reported associations in BREATHE ( $n=210$ ) and Tayside RCT ( $n=62$ ) were respectively OR=4.4, 95% CI: 1.77 – 10.96 and OR=9.6, 95% CI: 1.00 – 92.19. Meta-analysis of those results showed a highly significant association between rs2660845 common variant ( $f_g=0.30$ ) and exacerbation responses to montelukast (OR=5.30, 95% CI: 2.38 – 11.77,  $p=4.29 \times 10^{-5}$ ).

Our study suggested that genetic variation in *LTA4H* may contribute to variability in montelukast response. Individuals under 18 years-old carrying the rs2660845 G allele are significantly less responsive to montelukast treatment, as they have increased risk of exacerbation due to the down regulation of *LTA4H* activity. These findings have the potential to help personalise therapy for individuals with asthma by identifying subgroups of responders to this commonly used medication.



## PgmNr 458: Expression of *SMARCD1* interacts with age in association with asthma control on inhaled corticosteroid therapy.

### Authors:

M.J. McGeachie<sup>1</sup>; J.E. Sordillo<sup>2</sup>; A. Dahlin<sup>1</sup>; A.L. Wang<sup>1</sup>; S. Lutz<sup>2</sup>; K.G. Tantisira<sup>1</sup>; R. Panganiban<sup>3</sup>; Q. Lu<sup>3</sup>; S. Sajuthi<sup>4</sup>; C. Rios<sup>4</sup>; R. Kelly<sup>1</sup>; B. Saef<sup>4</sup>; C. Eng<sup>5</sup>; S.S. Oh<sup>5</sup>; A.T. Kho<sup>6</sup>; D.C. Croteau-Chonka<sup>1</sup>; S.T. Weiss<sup>1</sup>; B.A. Raby<sup>1,7</sup>; A.C.Y. Mak<sup>4</sup>; J.R. Rodriguez-Santana<sup>8</sup>; E.G. Burchard<sup>4</sup>; M.A. Seibold<sup>5</sup>; A.C. Wu<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Channing Division of Network Medicine, Brigham & Women's Hosp, Boston, Massachusetts.; 2) PREcisiOn Medicine Translational Research (PROMoTeR) Center, Department of Population Medicine, Harvard Pilgrim Health Care Institute and Harvard Medical School, Boston, MA; 3) Program in Molecular and Integrative Physiological Sciences, Departments of Environmental Health and Genetics & Complex Diseases, Harvard T.H. Chan School of Public Health, Boston, MA; 4) Center for Genes, Environment and Health, Department of Pediatrics, National Jewish Health, Denver, CO; 5) Department of Medicine, University of California San Francisco, San Francisco, CA; 6) Computational Health Informatics Program, Boston Children's Hospital, Boston, MA; 7) Division of Pulmonary and Respiratory Diseases, Boston Children's Hospital, Boston, MA; 8) Centro de Neumología Pediátrica, San Juan, PR

**Background:** Global gene expression levels are known to be highly dependent upon gross demographic features including age, yet identification of age-related genomic indicators has yet to be comprehensively undertaken in a disease and treatment-specific context.

**Objective:** To discover genomic indicators specific to response to inhaled steroids in individuals with asthma by accounting for age-dependent genomic interactions.

**Methods:** We used gene expression data from CD4+ lymphocytes in the Asthma BioRepository for Integrative Genomic Exploration (Asthma BRIDGE), an open-access collection of subjects participating in genetic studies of asthma with available phenotype and gene expression data. Replication population participants were Puerto Rico islanders recruited as part of the ongoing Genes environments & Admixture in Latino Americans (GALA II), who provided nasal brushings for transcript sequencing. The main outcome measure was chronic asthma control as derived by questionnaires. Genomic associations were performed using regression of chronic asthma control score on gene expression with age in years as a covariate, including a multiplicative interaction term for gene expression times age.

**Results:** The *SMARCD1* gene (SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 1) interacted with age to influence chronic asthma control on inhaled corticosteroids, with a doubling of expression leading to an increase of 1.3 units of chronic asthma control per year (95% CI [0.86, 1.74],  $p = 6 \times 10^{-9}$ ), suggesting worsening asthma control with increasing age. This result replicated in GALA II ( $p = 3.8 \times 10^{-8}$ ). Cellular assays confirmed the role of *SMARCD1* in glucocorticoid response in airway epithelial cells.

**Conclusion:** Focusing on age-dependent factors may help identify novel indicators of asthma medication response. Age appears to modulate the effect of *SMARCD1* on asthma control with inhaled



corticosteroids.

## **PgmNr 459: Rapid and ultra-rapid whole genome sequencing in acutely ill neonates: University of California Davis experience.**

### **Authors:**

S. Shankar <sup>1</sup>; S. Gandhi <sup>1</sup>; M. Martin <sup>1</sup>; S. Caylor <sup>2</sup>; L. Farnes <sup>2</sup>; S. Chowdhury <sup>2</sup>; W. Benson <sup>2</sup>; D. Dimmock <sup>2</sup>; K. Rauen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Genomic Medicine, MIND Institute UC Davis, Sacramento, California. 95817; 2) Rady Children's Institute for Genomic Medicine 3020 Children's Way, San Diego, CA 92123

---

We report on UC Davis Genomic Medicine experience in collaboration with the Rady Genomic Institute as one of the sites for Project Baby Bear. The Division of Genomic Medicine worked closely with all the key players in NICU, PICU and UC Davis laboratory including neonatologists, intensivists, nurses, NICU and PICU nursing managers, residents, fellows and pediatric surgeons. The initial presentation was by the Rady team in August 2018 during the site visit. This was followed by individual small group presentation to NICU, PICU and Pediatric surgery teams by the UC Davis Genomic Medicine team and flyers were distributed to remind them of the opportunity to perform Rapid and Ultra rapid Whole Genome Sequencing (rWGS & urWGS) through Project Baby Bear. Our Precision Genomic group established a protocol and created flow charts and smart phrases to facilitate the process of identification, recruitment, evaluation and sending of genetic tests. As a screening process for Baby Bear, the PICU and NICU patient list was screened daily to identify potential patients. The UC Davis Children's Hospital pediatric team have in general been enthusiastic participants. Using the inclusion/exclusion criteria provided by Project Baby Bear, a total of 29 patient families have been approached thus far and 25 have consented; three families declined due to "no interest in genetic testing." The clinical presentation of the babies enrolled thus far include arrhythmias, epilepsy, skeletal dysplasia, hydrops fetalis, congenital diaphragmatic hernia, Hirschsprung disease, congenital heart defects, distinct facial features, hypocalcemia, seizures, gastroschisis, failed extubation, tracheomalacia and subdural hematoma/fluid collections. Test results in all 25 individuals have been received: positive: 8/25 (32%), variant of uncertain significance (VUS): 3/25 (12%) and negative: 14/25 (56%). Finding a positive test result made a difference in the management of all babies in whom positive test result was identified. Although negative results did not directly influence management, it prevented further workup for genetic evaluations in the immediate neonatal period for all individuals. The option of rWGS and urWGS is changing the practice of genomic medicine, moving care from symptom-based treatment of illness to targeted treatments and predictive management of potential future health care concerns.

## PgmNr 460: PubCaseFinder and PhenoTouch for differential diagnosis in rare disease.

### Authors:

T. Fujiwara <sup>1</sup>; Y. Yamamoto <sup>1</sup>; J.D. Kim <sup>1</sup>; T. Takagi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Database Center for Life Science, Kashiwa-shi, Chiba, Japan; 2) Toyama University of International Studies, Toyama-shi, Toyama, Japan

Next-generation sequencing-based diagnostic tests have been attempted for the diagnosis of affected individuals with suspected rare diseases. The percentage of affected individuals diagnosed by clinical exome sequencing is up to 40% in several studies, unfortunately, other ~60% of affected individuals with suspected rare diseases don't receive a diagnosis. They will likely lose opportunities such as optimization of clinical management and early intervention. In this situation, phenotype-driven differential diagnosis systems such as Phenomizer and GDDP have been implemented in recent years. These systems provide a ranked list of rare diseases on the basis of the phenotypic similarity between an affected individual and rare diseases, and the top-listed diseases represent the most likely differential diagnosis. The performance of these systems is greatly influenced by the quantity and quality of both underlying databases of disease-phenotype associations (DPAs) and end-user input.

To improve the DPA databases, we developed a text mining approach to extend the coverage of DPAs and a phenotype-driven differential diagnosis system PubCaseFinder (<https://pubcasefinder.dbcls.jp>) using published case reports. A series of experiments showed that the performance of PubCaseFinder could substantially be improved thanks to the extension of the DPA database, and our proposed approach is deemed practically useful because manual curation will take enormous time and cost.

Human Phenotype Ontology (HPO) terms are the input to PubCaseFinder as affected individual's phenotypes. However, the existence of a huge number of HPO terms including synonyms and abbreviations preclude the finding of appropriate HPO terms by text search. To address this problem, we developed PhenoTouch, a new phenotype input assistant tool which was implemented in PubCaseFinder. We mapped 3,642 HPO terms to 991 FMA terms using external links, and then we mapped these FMA terms to 2,536 OBJ files of BodyParts3D/Anatomography. PhenoTouch facilitates searching for HPO terms using a human 3D model constructed by these OBJ files. In the case of selecting the region of the mandible in the human 3D model, users can easily get HPO terms related to the mandible such as retrognathia, micrognathia, narrow jaw, and broad jaw which do not include the word "mandible".

We believe our proposed approach and PhenoTouch will be valuable for improving upon phenotype-driven differential diagnosis systems for rare diseases.

## PgmNr 461: Enabling precision medicine in large exome-sequenced cohorts.

### Authors:

S. Khalid; S. Balasubramanian; C. Gonzaga-Jauregui; C.V. Hout; A. Hawes; B. Ye; N. Banerjee; L. Habegger; A.L. Blumenfeld; J. Staples; A.R. Shuldiner; M. Cantor; J.G. Reid; A. Baras

[View Session](#) [Add to Schedule](#)

**Affiliation:** Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY

---

Precision medicine aims to leverage genetic data of patients and family members to preemptively identify disease risks, prevent or manage disease conditions, and tailor therapeutic choices. At the Regeneron Genetics Center (RGC), we are building a comprehensive database of human genetic variation which currently comprises of exome sequences of about 500,000 individuals coupled with deidentified Electronic Health Record (EHR) data. Such data allow us to identify novel and known genetic variation in the population and systematically evaluate their effect on patient health. To this end, we have implemented a scalable workflow using Spark Resilient Distributed Dataset architecture to integrate various databases of clinically relevant variants, such as ClinVar and Human Gene Mutation Database (HGMD), with RGC's ever growing sequencing and EHR data. Our workflow allows us to query genetic variant databases based on various criteria.

As an example, we discuss our survey of medically actionable variants from the exome sequencing of 50,000 individuals from the UK Biobank. We identify pathogenic (P) and likely pathogenic (LP) variants in the ACMG59 genes list; a set of genes known to cause or predispose individuals to disease and where medical intervention is expected to improve outcome(s) in terms of mortality or the avoidance of significant morbidity. Using a stringently defined "strict" set of pathogenic variants from ClinVar with an assertion criteria of  $\geq 2$  stars, we show that approximately 2% of individuals have an actionable variant. Using more relaxed criteria for defining pathogenic variants provides a larger estimate ranging from 2 -7%. We further analyze the 239 LP predicted loss-of-function (pLOF) variants from the strict set (315 carriers, 0.63%) in genes where truncation is the expected mechanism of disease. We evaluate LP pLOF variants in terms of: 1. Stop causing variants that escape NMD, affect  $< 10\%$  of the protein, 2. non-canonical splice variants, and 3. start-lost variants with alternate initiation sites downstream.

We highlight specific considerations for merging different databases of pathogenic variants, interpreting pLOF variation and the importance of systematic and scalable evaluation of genetic variants. Combining this approach with EHR data will allow us to evaluate novel variants in known disease genes, identify incorrectly reported pathogenic variants and refine pathogenic annotations in terms of penetrance and prevalence.

## **PgmNr 462: First-line preventative genetic screening: Disease penetrance in Tier 1 inherited diseases in an all-comers population is similar to family history selected populations.**

### **Authors:**

J. Lu <sup>1</sup>; J. Grzymski <sup>2</sup>; G. Elhanan <sup>2</sup>; E. Smith <sup>2</sup>; C. Rowan <sup>3</sup>; N. Slotnick <sup>2</sup>; S. Dabe <sup>2</sup>; K. Schlauch <sup>2</sup>; R. Read <sup>2</sup>; W.J. Mefcalf <sup>2</sup>; B. Lipp <sup>2</sup>; H. Reed <sup>2</sup>; E. Levin <sup>1</sup>; J. Kao <sup>1</sup>; M. Rashkin <sup>1</sup>; J. Bowes <sup>1</sup>; A. Bolze <sup>1</sup>; K. Dunaway <sup>1</sup>; N. Washington <sup>1</sup>; A. Slonim <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Helix, San Carlos, California.; 2) Desert Research Institute, Reno, Nevada; 3) Renown Health, Reno, Nevada

---

Clinical testing guidelines are focused on family history or demographic/ ethnic background as penetrance estimates for inherited genetic disease are often derived from populations with known disease. It is unknown if disease penetrance in unselected populations will be lower than in selected populations. Here we report on disease penetrance in three inherited autosomal dominant diseases: Familial Hypercholesterolemia (LDLR, PCSK9, APOB), Hereditary Breast and Ovarian Cancer (BRCA1, BRCA2) and Lynch Syndrome (MLH1, MSH2, MSH6, PSM2) in unselected adult volunteers (n = 23,709), who underwent clinical exome sequencing as a part of the Healthy Nevada Project (HNP) in Northern Nevada (Renown Health, Reno, Nevada) from March 15, 2018, to Sept 30, 2018. We identified more than 290 carriers of pathogenic/ likely pathogenic alleles, >80% of which have medical records. We report that over 90% of these carriers are undetected in the medical system with over 80% without any documented family history. In addition, 26% of these carriers have already manifested with disease, with a median age of first presentation of relevant disease at age 60, 45 and 62 for HBOC, LS and FH. When examined more closely, we find that disease penetrance in unselected carriers for LS and FH is not different from carriers identified from family history. 40% of unselected LS carriers present with colorectal or other cancers by age 40, similar to disease registries of LS patients. For FH, we estimate that 50% of unselected FH carriers will present with atherosclerotic related cardiovascular disease by age 60. We also find BRCA carriers have greatly elevated risk for HBOC. With ~90% of disease presentation occurs after 40, and 70% before the age of 65, our experience suggests that a population health based approach to genetic screening in younger patients can provide substantial clinical, economic, and patient benefit.

## PgmNr 463: Feasibility and potential impact of exon skipping as a therapeutic for neurofibromatosis type I (NF1).

### Authors:

D. Wallis <sup>1</sup>; A. Leier <sup>1</sup>; H. Liu <sup>1</sup>; M. Daniel <sup>1</sup>; B. Korf <sup>1</sup>; R.A. Kesterson <sup>1</sup>; G. Dickson <sup>2</sup>; L. Popplewell <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, Univ Alabama at Birmingham, Birmingham, ALABAMA.; 2) Molecular Cell Biology, Centre of Gene & Cell Therapy, School of Biological Sciences, Royal Holloway - University of London, Egham, Surrey, TW20 0EX, UK

---

NF1 is a rare genetic disorder affecting ~1:3000 individuals. Almost all patients develop benign cutaneous and plexiform neurofibromas, but there is also a ~10% lifetime risk for development of malignant peripheral nerve sheath tumors. In addition 40-60% of afflicted children exhibit learning disabilities. NF1 encodes neurofibromin and mutations lead to hyperactivation of the Ras pathway. Current therapeutics primarily target the Ras pathway, but these are not very effective and are limited by toxicity. NF1 is a large protein and it is unclear how critical regions and domains outside the well described Gap-related domain (GRD) are for functionality. Potentially, some are dispensable and can be deleted or skipped through exon skipping which may lead to a truncated but still functional protein, despite genetic mutation. In silico analysis has identified that 45 of 57 total exons could be skipped either as a single exon or two consecutive exons while still maintaining the translational reading frame. We performed additional in silico analysis to predict which exons could be skipped with minimal loss of NF1 function and maximum therapeutic potential. In efforts to identify exon skips that are likely to be asymptomatic we prioritized exon candidates based on: known skips in unaffected individuals, lack of skips reported in NF1 individuals, and exons that are not within critical NF1 functional domains. Next, we performed a cDNA functional screen utilizing a novel Nf1 cDNA system to determine the effects of exon skipping on in vitro NF1 expression and GRD function. Both in silico and in vitro data will be presented. We find that cDNAs representing loss of some exons are able to maintain GRD function. Furthermore, we note that deletion of some exons leads to lower levels of NF1 protein expression presumably due to loss of protein stability. We show that after both GRD function and NF1 expression levels are considered jointly, skipping of specific exons may be therapeutic. This represents a personalized medicine approach for the treatment of NF1.

## PgmNr 464: The genetic diversity gap: Precision medicine is leaving most of us behind.

### Authors:

B. Murarka; P. Fung

[View Session](#) [Add to Schedule](#)

**Affiliation:** Spectrum Bio Bank, Portola Valley, CA.

---

### The Genetic Diversity Gap

The future is here but it won't work for everyone. Specifically, Polygenetic Risk Scores are bringing precision medicine into mainstream medical practice. Unfortunately, in precision medicine, your data source is your destiny. Billions of people will get little benefits from DNA based medical breakthroughs.

Why? 85% of DNA used in research and drug development is from people of European descent. This is creating "The Genetic Diversity Gap" for the vast majority of people. The Genetic Diversity Gap has profound implications on access to precision medicine. Future generations of underrepresented communities face a rapidly compounding problem that will become insurmountable unless we start to address it today.

Spectrum BioBank is the world's only diversity-first genome database. We are a non-profit organization, funded by private donors, fixated on bridging this gap through public awareness of the problem, directly sequencing contributions from underrepresented communities, and enabling inclusive research to close this gap permanently.

Uniquely, we build direct relationships with DNA contributors from underrepresented communities so that our research partners can recontact and engage with these diverse communities. In January of 2020 we'll launch free sequencing for the major underrepresented populations and their extremely heterogeneous subpopulations (Hispanic-Latinos consists of Mexican, Puerto Rican, South American, etc. yet another reason we need this research).

We are a multidisciplinary team of experts from Consumer Genetics, Academia, Software Development, and Customer Acquisition who are passionate about eliminating The Genetic Diversity Gap.

In this talk we will share our plan to solve the gap, our initial research focuses, and a call to action for the ASHG membership to help close the gap.

## PgmNr 465: Automating analyses of high-throughput genotyping global screening array data.

**Authors:**

D.E. Condon; P.F. Cherukuri; L. Carmichael; K. Lubieniecki; J. Lubieniecka; S. Bartaria; K. Meis

[View Session](#) [Add to Schedule](#)

**Affiliation:** Sanford Health, Imagenetics, Sioux Falls, SD.

---

**Background:** Sanford Health is using the Illumina Global Screening Array (GSA) towards implementation of precision medicine and precision public health. GSA allows screening for medically actionable variants and known pharmacogenomic variants among its 661,125 assays. This system needs routine checking for proficiency of laboratory technicians, competency of the system in general, and other potential experiments involving comparisons. In order to facilitate this checking, we developed a pipeline, using Databricks as a platform, that allows concordance checks between experimental test results and known standards, confirming that the GSA is performing as expected.

**Hypothesis:** The GSA requires periodic checking to ensure that the genotyping chip is reporting correct results. We report a system which can easily run concordance using the cluster computing framework Apache Spark on Databricks.

**Methods & Results:** Genotype data is received from the Molecular Genetics Lab in a binary Illumina GTC format and is converted to a per sample Variant Call Format (VCF) file. All sample data within a directory is processed into a single dataframe, which allows for rapid concordance calculation. This allows highly parallel, fast computational analysis of large datasets. The concordance check is done with variable sets for “truth” and “test” datasets. We use the 1000 Genomes project data as a standard of truth set. However, our workflow can be used for other purposes, for example, to verify whether the GSA gives the same results regardless of source tissue, or that the GSA performs the same at different sample DNA concentrations, or whether the location of where GSA data was generated matters. GSA data is received and prepared for analyses on the Databricks platform, which allows highly parallel, fast data analysis of large data sets. The output organizes concordance according to seven types of genotype discordance and five types of genotype concordance. This can compare a single GSA run compared with the 1000 Genomes project in as little as 39 minutes. We have developed a system that automatically runs concordance checks on GSA results, and can compare with other GSA runs and against a 1000 Genomes truth set. This is beneficial for ensuring that results are being reported accurately to the patients. GSA results maintain a high concordance rate with the 1000 genomes project.

**Conclusion:** We have created a system for automated and easy concordance checking of the GSA



## PgmNr 466: Using CRISPR-Cas9 to generate zebrafish models of human disease for drug discovery.

**Authors:**

R. Noche<sup>1,4</sup>; X. Cui<sup>1,4</sup>; E. Pannia<sup>1,4</sup>; J. Dowling<sup>1,2,3,4,5,6,7</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Zebrafish Genetics and Disease Models Core Facility; 2) Genetics and Genome Biology Program; 3) Division of Neurology; 4) Hospital for Sick Children, Toronto, Canada; 5) Department of Pediatrics; 6) Department of Molecular Genetics; 7) University of Toronto, Toronto, Canada

---

Advances in next generation sequencing have greatly accelerated the identification of novel disease-associated gene mutations. However, a prevailing challenge continues to be functional validation of these variants, particularly for rare diseases. We have established the Zebrafish Genetics and Disease Models Facility that combines shared expertise and infrastructure at the Hospital for Sick Children; creating a pipeline from gene identification to functional validation and drug discovery. With a fee for service, cost recovery model, our facility aims to provide the services required to efficiently generate and analyze zebrafish models that accurately recapitulate human disease. We use a high throughput CRISPR-Cas9 mutagenesis system along with high resolution melt (HRM) analysis to generate mutations in zebrafish that are targeted to putative human disease loci. Additionally, we offer zebrafish phenotypic analysis and drug discovery services. To date, we have worked with over 29 individual labs to provide our services including generation of over 35 targeted mutations in over 25 zebrafish genes with a success rate of >85%. Here we present the results of our high-throughput mutation generation effort, as well as the preliminary characterization of our first successful zebrafish mutant strains.

## PgmNr 467: Genetic basis of pharmacokinetic variability in African Americans.

### Authors:

T. De <sup>1</sup>; C. Alarcon <sup>1</sup>; A. Gaisin <sup>1</sup>; J. Avitia <sup>1</sup>; B. Lec <sup>1</sup>; E. Smithberger <sup>2</sup>; Y. Xu <sup>2</sup>; A. Ott <sup>1</sup>; M.A. Perera <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pharmacology, Northwestern University, Chicago, Illinois.; 2) University of Chicago, Chicago, Illinois

**Background:** Genetic variation contributes substantially to the differences in drug response and in particular enzyme activity within and across populations. African Americans (AAs) suffer disproportionately from adverse drug reactions. Using the unique genomic architecture of AAs and probe drugs for established drug metabolizing enzymes, we uncovered population-specific SNPs that may explain health disparity in drug response.

**Methods:** Primary hepatocyte suspension cultures derived from 60 AA were treated with four probe drugs for specific cytochrome P450 enzymes (CYPs) [midazolam (CYP3A4/5), bupropion (CYP2B6), mephenytoin (CYP2C19), phenacetin (CYP1A2)] to identify the SNPs associated with metabolite formation rate (MFR) using genome-wide association and imputed gene based methods.

**Results:** We performed four GWAS specific to each probe drug and identified rs35672164 (beta: -0.9,  $p=6.9 \times 10^{-8}$ ) and rs17882547 (beta: -1.1,  $p=2.9 \times 10^{-7}$ ) associated with hydroxy-mephenytoin MFR. Within GTEx, rs17882547 is a multiple-tissue expressed quantitative trait loci (eQTL) for *CYP2C19* and is also in linkage disequilibrium (LD) with *CYP2C19\*2*, associated with poor clopidogrel response, reduced phenytoin metabolism and acenocoumarol maintenance dose. rs35672164 is a non-coding variant within *CYP2C18* and is only found in African ancestry populations. rs73993690 (beta: -2.13,  $p=8.6 \times 10^{-7}$ ) was associated with hydroxy-midazolam MFR and is a blood eQTL for *PRKCA*, a kinase that plays an important role in hepatic CYP3A4 turnover. rs10411010 (beta: -1.15,  $p=1.97 \times 10^{-7}$ ) was associated with hydroxy-bupropion MFR. Lastly, rs264558 (beta: 1.14,  $p=2.35 \times 10^{-8}$ ) was associated with acetaminophen MFR. Gene based analysis using magma identified *CYP2C19* ( $p=1.13 \times 10^{-6}$ ), *CYP2C9* ( $p=9.34 \times 10^{-6}$ ) and *CYP2C18* ( $p=3.34 \times 10^{-6}$ ) to be significantly associated with hydroxy-mephenytoin MFR and *ZIM3* ( $p=2.44 \times 10^{-5}$ ), a Zinc finger-domain gene significantly associated with hydroxy-bupropion MFR. Zinc finger proteins are major regulators of CYP gene expression.

**Conclusion:** We identified common population-specific genetic loci that were significantly associated with probe drug MFR, a clinically relevant endophenotype. The identified genetic loci may represent regulatory variants that predispose individuals to adverse events related to several different classes of medications. Our findings may help towards predicting adverse drug events and improving treatment outcomes in AAs.

## PgmNr 468: Host genome-wide association study of infant susceptibility to shigella-associated diarrhea.

### Authors:

D. Duchon <sup>1</sup>; R. Haque <sup>2</sup>; G. Wojcik <sup>3</sup>; L. Chen <sup>1</sup>; P. Korpe <sup>1</sup>; B. Kirkpatrick <sup>4</sup>; W. Petri <sup>5</sup>; P. Duggal <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States; 2) International Center for Diarrhoeal Disease Research, Bangladesh, Dhaka, Bangladesh; 3) Stanford University, Stanford, CA, United States; 4) University of Vermont, Burlington, VT, United States; 5) University of Virginia, Charlottesville, VA, United States

---

Shigella is a leading cause of moderate-to-severe diarrhea in African and South Asian children and the causative agent of shigellosis and dysentery. Associated with 80 - 165 million cases of diarrhea and up to 600,000 deaths annually, exposure to shigella is ubiquitous in many regions while colonization or infection is heterogenous. To characterize the host-genetic susceptibility to shigella-associated diarrhea, we performed two independent genome-wide association studies (GWAS) including 589 Bangladeshi infants, 429 from the PROVIDE birth cohort and 160 infants from a Cryptosporidium-focused study birth cohort in Dhaka, Bangladesh. We classified children as ever having shigella associated diarrhea or not in the first 13 months of life. A qPCR Ct distribution of the ipaH gene, carried by all four shigella species and enteroinvasive E. coli, identified a total of 143 infants with a shigella-associated diarrheal event and 446 infants with no evidence of shigella-associated diarrhea within their first 13 months of life. Host GWAS's were performed using the Illumina Infinium 5 Multiethnic Global Array and analyzed under an additive genetic model. A joint analysis (imputed variants N=6,547,362) identified loci of interest on chromosomes 11 (rs582240, within the *KRT18P59* pseudogene, average MAF=29.4%,  $P = 8.37 \times 10^{-8}$ ) and 8 (rs12550437, within the lincRNA *RP11-115J16.1*, average MAF=38.1%,  $P = 1.69 \times 10^{-7}$ ). This study suggests host genetic factors may influence the response to shigella colonization and pathogen-associated diarrhea. Additional replication and further research on the function of these genes and their association with shigellosis and other pathogen-associated diarrheal diseases is warranted.

## PgmNr 469: Prediction of personalized radiotherapy cytotoxicity score based on transcriptomics data.

### Authors:

C. Li <sup>1</sup>; J. Peacock <sup>2</sup>; S. Eschrich <sup>2</sup>; X. Liu <sup>1</sup>; J. Torres-Roca <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of South Florida, COPH, Tampa, FLORIDA.; 2) Moffitt Cancer Center, Tampa, FLORIDA.

**Background:** Radiotherapy (RT) is the one of the most commonly adopted therapeutics to treat cancer. Through ionizing radiation, RT primarily causes irreversible DNA damage to tumor cells which will lead to cell death. Although efforts have been made to improve the quality of RT, a key question remains to be answered is whether exploiting individual tumor biology could benefit patients receiving RT. To address this issue, we aimed to develop a new model to predict personalized RT cytotoxicity score (RT-CTS) using tumor transcriptomics data. This proposed model would then be able to guide the delivery of individualized treatment strategies.

**Methods:** We retrieved publicly available NCI-60 cancer cell line data, i.e. gene expression data, and survival curves. We utilized the linear quadratic (LQ) model to calculate cytotoxicity ( $\alpha$  and  $\beta$ ). First,  $\alpha$  and  $\beta$  in the 60 cell lines were calculated based on the LQ model. Then gene expression data were used to predict  $\alpha$  and  $\beta$  using linear support vector machine with 10-fold cross validation (CV). Finally, the performance of the models was validated using two independent clinical data sets of breast and lung cancer patients. Statistical analysis was based on multi-variable COX regression.

**Results:** Thirteen and seventeen highly informative genes were selected using max-min Markov blanket method to predict  $\alpha$  and  $\beta$ , respectively. The 10-fold CV R-squared was 0.82 (root mean square error (RMSE)= 0.20) for  $\alpha$  and 0.65 (RMSE= 0.03) for  $\beta$ . Using the LQ model, our predicted  $\alpha$  and  $\beta$ , and each patient's RT dosing schedule, we calculated a new score, RT-CTS, to estimate the number of clones of tumor cells killed for each individual patient in the two validation cohorts. In the lung cohort, patients with a predicted RT-CTS  $> 10^{-12}$  (RT-CTS12) had an improved 1-yr local control (LC) (84.3% vs 57.1%,  $p<0.001$ ) and overall survival (OS)(4.8 vs 1.9 yrs,  $p=0.04$ ). In the RT breast cohort, patients who achieved RT-CTS12 had superior 8 yr LC (91% vs 72%  $p=0.03$ ) that is significant on multi-variable analysis (Hazard ratio= 0.29,  $p<0.014$ ).

**Conclusion:** Based on gene expression data, we developed and validated a machine learning model that predicts RT-CTS, an individualized RT cytotoxicity score. The model can be used to provide a personalized RT dosing schedule given tumor cell expression. This study highlights the importance of transcriptomics and its impact on personalized medicine.

## PgmNr 470: Combining clinical and candidate gene data into a risk score for azathioprine-associated leukopenia in routine clinical practice.

### Authors:

P. Anandi; A.L. Dickson; Q. Feng; W. Wei; W.D. Dupont; D. Plummer; G. Liu; K.A. Barker; V.K. Kawai; K. Birdwell; N. Cox; C.M. Stein; A. Hung; C.P. Chung

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt University Medical Center, Nashville, TN

### Background/Purpose:

Azathioprine (AZA) is widely used to treat rheumatic diseases and for organ transplantation. However, serious adverse events, such as leukopenia, can limit treatment. Our ability to predict AZA-associated leukopenia is limited to the evaluation of the enzyme thiopurine methyl transferase (TPMT). We hypothesize that a risk score comprised of clinical factors and selected candidate gene variants—in addition to TPMT—could improve the prediction of AZA-associated leukopenia.

### Methods:

This is a case-control study using de-identified records from a clinical practice-based biobank at a tertiary medical center. We identified 425 Caucasian patients who received prescriptions for AZA. Subjects with a leukocyte count of  $<4,000$  WBC/ $\mu$ L while receiving AZA prescriptions were classified as cases. Subjects who did not develop leukopenia were classified as controls. We reviewed the clinical records, collected clinical variables and genotyped 71 genetic candidates of which 60 passed quality control. We built two scores using the coefficients from multivariate regressions. The first model was a logistic regression including TPMT status, age, and sex. The second model also included weight, clinical indication for AZA use, concurrent prescription of other medications, AZA daily dose, and genetic candidates. To avoid overfitting, we used ridge regression for the second model.

### Results:

216 patients had AZA-associated leukopenia and 209 were control subjects. Younger age, lower weight, a lupus diagnosis or organ transplant, and being a carrier of the *TPMT* haplotype were significantly associated with leukopenia. The area under the ROC curve (AUC) of a score based on a regression that included TPMT status, age, and sex was 0.59 (95% C.I.: 0.54-0.64). After shrinkage, the ridge model included 11 clinical and 47 genetic candidates from 8 genes (*TPMT*, *NUDT15*, *XDH*, *MOCOS*, *ABCC4*, *ITPA*, *AOX1*, and *GST*); the AUC of the score from that model was 0.77 (95% C.I.:0.73-0.82).

### Conclusion:

This proof-of-concept study investigates the utility of adding clinical and genetic variables to TPMT to improve discrimination of the risk for AZA-associated leukopenia. After shrinkage, we were able to show that a combined model, including clinical and genetic variants, improved our ability to predict leukopenia. Further studies in larger populations with independent replications are needed to further develop prediction scores that are useful for routine clinical practice.



## PgmNr 471: Implementation of pharmacogenomics into primary care across a large multi-state healthcare system.

### Authors:

J.F. Baye <sup>1,2,3</sup>; N. Petry <sup>1,4</sup>; A. Massmann <sup>1</sup>; A. Aifaoui <sup>1</sup>; C. Hajek <sup>1,3</sup>; R.A. Wilke <sup>1,3</sup>; A. Schultz <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sanford Imagenetics, Sioux Falls, SD; 2) South Dakota State University College of Pharmacy and Allied Health Professions, Brookings, SD; 3) University of South Dakota Sanford School of Medicine, Vermillion, SD; 4) North Dakota State University College of Pharmacy, Fargo, ND

---

Individualized medicine – the use of molecular, genetic and other patient specific features to customize patient care – is a paradigm shift in allopathic medicine. There are a number of different approaches to individualizing care, and multiple institutions throughout the United States have established programs for implementing various technologies. Pharmacogenomics (PGx) is among the most advanced and effective forms of individualized medicine to date. PGx utilizes known variation in drug metabolizing enzymes to predict and avoid certain medications based on an individual's genetic makeup. Furthermore, technologies such as programmed clinical decision support (CDS) provide prescribers with drug-gene interactions and therapy recommendations at the point of care, therefore making PGx timely, specific, and actionable.

The individualized medicine initiative at Sanford Health – named Imagenetics (merging Internal Medicine and Genetics) – began offering PGx testing in 2015 starting with *CYP2C19* for clopidogrel activity. An 8-gene prospective PGx panel consisting of *CYP2C19*, *CYP2C9*, *CYP2D6*, *CYP3A5*, *DPYD*, *SLCO1B1*, *TPMT*, and *VKORC1* was subsequently introduced in 2016. This PGx panel has allowed us to capture the majority of drug-gene interactions listed in the Clinical Pharmacogenomics Implementation Consortium (CPIC) guidelines, and additional expansion of PGx tests is currently underway.

Presently, we offer PGx testing to all patients within the Sanford Health system, which consists of more than 2 million patients across 9 different states in the upper Midwest. With a centralized genomic laboratory performing PGx testing across such a large territory, integrated technology is crucial. To maximize utility of the genomic laboratory data, PGx CDS was integrated into the electronic medical record (EMR) across the entire enterprise. Furthermore, clinical pharmacist tools and workflows were implemented to improve efficiency of review. Video conferencing, a centralized information resources (intra web), and a robust online education program are a few of the other tools used to implement and maintain this large PGx program. We present major challenges encountered during our first few years of PGx implementation, and we propose several novel solutions.

## PgmNr 472: Quantifying known and novel actionable pharmacogenetic variations in whole-exome sequencing data of 1,091 Chinese subjects.

### Authors:

H. Chung <sup>1</sup>; H.C. Yu <sup>1</sup>; C.Y. Chan <sup>1</sup>; K.S. Yeung <sup>1</sup>; J.F.T. Chau <sup>1</sup>; M.Y. Lee <sup>1</sup>; W.L. Yang <sup>1</sup>; J. Yang <sup>1</sup>; C.S.M. Tang <sup>2</sup>; M.M. Garcia-Barcelo <sup>2</sup>; Y.L. Lau <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Paediatrics and Adolescent Medicine, The University of Hong Kong, Hong Kong, China; 2) Division of Paediatric Surgery, Department of Surgery, The University of Hong Kong, Hong Kong, China

Pharmacogenetics (PGx) is an important area of precision medicine. Currently, lack of large-scale exome/genome-wide data precludes clinical application of pre-emptive PGx testing in Chinese. Since whole exome sequencing (WES) data can be used to extract PGx information for individuals undergoing a diagnostic test, we analyze known and novel variants in clinically actionable pharmacogenes in a cohort recruited for the evaluation of pediatric genetic diseases in Hong Kong.

**Methods:** Sequencing data of 1,091 unrelated Chinese subjects in the HKU Pediatric Exome Cohort were analyzed using a two-tiered approach. The 1st-tier analysis targeted known PGx variants based on a curated actionable variants list (PharmGKB Level 1/ ethnic-specific high-confidence variants by literature search). The 2nd-tier analysis focused on discovery of rare, predicted deleterious variants in 121 genes with PharmGKB Level 1 and 2 evidence.

**Results:** We identified 34 PGx variants (28.3%) in the curated variant list, comprising of 41 drug-gene combinations. The most prevalent actionable variant is rs9934438 in VKORC1 (93.8%), followed by rs4244285 in CYP2C19 (40.9%) and rs2108622 in CYP4F2 (39.0%). In addition to known variants, we discovered 106 novel variants in 121 PharmGKB Level 1/2 genes that are i) not found in PharmGKB; ii) <1% in gnomAD ; and have either iii) CADD>20 or REVEL>0.7 (missense variants) or (iii) “high-confidence” annotation in LOFTEE (LOF variants). They are implicated in 125 drug-gene combinations and demonstrate diverse functional effects such as non-synonymous nucleotide substitution (n=93), premature gain of stop codon (n=9) and disruption of reading frame (n=4). Of the 1,091 Chinese subjects, 99.3% harbored at least 1 known actionable PGx variant (median=4) and 56.7% had at least 1 novel predicted deleterious PGx variant. Focusing on warfarin PGx, we observed high occurrence of VKORC1 actionable variants (rs9934438) in Chinese, while actionable CYP2C9 variants were rare compared to Caucasians.

**Conclusion:** To date, this is the largest Chinese PGx study using WES data. We have demonstrated the abundance of known and novel functional PGx variants that are potentially actionable clinically. Since existing large-scale PGx studies include mostly Caucasian subjects e.g. there are only 196 Asians in the eMERGE-PGRNseq study (2.2% of the cohort), this study addresses the unmet research gap and provides further insights into the PGx landscape of Chinese individuals.



## PgmNr 473: Utility of pharmacogenomic dosing intervention in children with chronic disease.

### Authors:

D.S. Cross <sup>1</sup>; K. Fulda <sup>2</sup>; S. Capasso <sup>4</sup>; W.P. Bowman <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Microbiology Immunology and Genetics, UNTHSC Graduate School of Biomedical Sciences, Fort Worth, Tx.; 2) Family Medicine, Texas College of Osteopathic Medicine UNTHSC, Fort Worth, Tx; 3) Pediatrics, Texas College of Osteopathic Medicine UNTHSC, Fort Worth, Tx.; 4) School of Education, Human and Health Sciences, Bay Path University, Longmeadow, MA.

---

**Problem Statement:** Utilizing pharmacogenomic testing (the use of genetics to guide pharmaceutical choice and dosing) in a pediatric patient population has the ability to reduce adverse drug reactions ADRs (unexpected reactions to a drug taken at the recommended dose). This type of implementation has the ability to increase health disparities if racial and ethnic differences in allelic variation, drug utilization, and patient preferences are not part of clinical utility models. Here we determined how racial, and ethnic differences affect clinical utility of a model pharmacogenomic implementation.

**Methods:** We developed a Markov model that simulates a 15 year timeline for a cohort of children, the perspective of the model was that of the individual payer. Children were assigned to a health state with three different levels of disease state, healthy, chronic disease, or complex chronic disease. We assume that genotyping for pharmacogenomic dosing will occur at the time of diagnosis with a chronic disease. The model was interrogated for the cohort with and without a pharmacogenomic intervention. The incremental cost effectiveness ratio (ICER) was calculated. Sensitivity analysis was performed to determine highly influential variables. We further varied the rate of ADRs and the likelihood of benefit for a pharmacogenomic intervention based on racial and ethnic likelihoods.

**Results:** The cohort eligible for genotyping was 5366 individuals. With the assumption that genotyping cost was \$250, approximately 22 QALY's would be saved through genotyping, with an incremental cost effectiveness ratio (ICER) of \$46,777.18. Parameters that are sensitive to racial and ethnic background such as the likelihood that genotyping would affect prescription dosing, the number of QALYs lost from an ADR and the likelihood to have an ADR and are highly influential model parameters (ranked 1st through 3rd) in influence in the sensitivity analysis. Therefore the cost effectiveness varied with race and ethnicity when estimates of the chance of an ADR and the likelihood that genotyping would reduce an ADR were varied by racial and ethnic background. Individuals who were Black or White non-Hispanic were most likely to benefit from a pharmacogenomic intervention. The highest number of scenarios where cost effectiveness was achieved (8 and 4 scenarios) while White- Hispanic and Asian groups demonstrated poor cost effectiveness (1 and 0 scenarios).

## PgmNr 474: The design and implementation of a cohort-level pharmacogenomics program at UnitedHealth Group.

### Authors:

D. Crosslin<sup>1</sup>; K. Tzeng<sup>2</sup>; R. Altman<sup>3</sup>; M. Barnes<sup>4</sup>; D. Dimmock<sup>5</sup>; J. Johnson<sup>6</sup>; B. Piening<sup>7</sup>; D. Voora<sup>8</sup>; B. Carleton<sup>9</sup>; M. Snyder<sup>10</sup>; C. Kennedy<sup>2</sup>; C. Mirando<sup>2</sup>; C. Miller<sup>2</sup>; C. Theis<sup>2</sup>; E. Preimesberger<sup>2</sup>; E. Seterra III<sup>2</sup>; E. Schlicht<sup>2</sup>; J. Crockett<sup>2</sup>; J. Staib<sup>2</sup>; J. Gardiner<sup>2</sup>; M. Truex<sup>2</sup>; M. Jarvis<sup>2</sup>; R. Mangel<sup>2</sup>; S. Pradhan<sup>2</sup>; S. Youssef<sup>2</sup>; S. Grossinger<sup>2</sup>; S. Smela<sup>2</sup>; T. Lifke<sup>2</sup>; E. Fultz<sup>2</sup>; B. Keating<sup>11</sup>; UnitedHealth Group Research and Development

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 2) UnitedHealth Group Research and Development, Minnetonka, MN; 3) Department of Bioengineering, Stanford University, Stanford, CA; 4) Queens Mary University of London, London, UK; 5) Rady Children's Institute for Genomic Medicine, San Diego, CA; 6) University of Florida College of Pharmacy Gainesville, FL; 7) Earle A. Chiles Research Institute, Providence Cancer Center, Portland, OR; 8) Duke Center for Applied Genomics & Precision Medicine, Durham, NC; 9) Department of Pediatrics, University of British Columbia, Vancouver, BC; 10) Department of Genetics, Stanford University, Stanford, CA; 11) Perelman School of Medicine at the University of Pennsylvania, Philadelphia PA

---

UnitedHealth Group (UHG) has assembled a panel of academic researchers to assist in the development of a cohort-level pharmacogenomics program for UHG member institutions. The panel includes a synergy of medical geneticist and pharmacogeneticists, charged with identifying drug-gene pairs with potential pharmacogenetic testing to improve drug efficacy as well as economic outcomes.

A total of six working groups were created based on panel member expertise and research interest, and paired with UHG data analysts and economists. The goal was to model and compare a pharmacogenomic test-informed path versus standard care path for each candidate gene-drug pair. The desired outcome was to select one or two candidates to pilot in a randomized clinical trial across multiple sites with future studies to follow. The models were informed by published allele frequencies, evidence-based outcome frequencies, and economic data derived from UHG claims. The models included evidence provided by Clinical Pharmacogenetics Implementation Consortium (CPIC) and the Pharmacogenomics Knowledgebase (PharmGKB).

The majority of gene-drug pair candidates contained genes encoding the cytochromes P450 family of enzymes, including *CYP2D6* (opiates), and *CYP2D6/CYP2C19* (anti-depressants) with the desired outcomes of pain relief and number of prescriptions, and Hamilton depression score changes and number of prescriptions, respectively. *CYP3A5* (tacrolimus) was nominated for immune suppression and less acute rejection of transplants. *TPMT* (thiopurines), and *UGT1A6*, *SLC28A3*, and *RARG* (anthracyclines) were nominated for oncology treatment and the reduction of adverse events. Finally, the *HLA* region was nominated for abacavir, carbamazepine, allopurinol, phenytoin, and various adverse drug reactions.

Besides economic modeling, others implementation factors were considered. For instance, an alternate therapy needed to be identified for an actionable allele. The panel also identified how quickly return of results were needed for an indication, and the range of physician compliance with standard of care. National impact was also a consideration. Details of these elements will be shared. The *CYP2D6* (opiates) project was selected as the first pilot, and trial design and site selection are currently ongoing. We hope our pharmacogenomics program collaboration between academia and industry will serve as a model, and provide a template for future genomic implementation initiatives.

## PgmNr 475: *HLA-B\*51:01* and *CYP2C9\*3* are risk factors for phenytoin-induced eruption in Japanese population: The analysis in the Biobank Japan Project.

### Authors:

K. Hikino <sup>1</sup>; T. Ozeki <sup>1</sup>; M. Koido <sup>1</sup>; C. Terao <sup>1</sup>; Y. Kamatani <sup>1</sup>; T. Mushiroda <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 2) Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan

---

Phenytoin is one of the most widely used anti-epileptic drugs, and therefore it is important to prevent adverse reactions due to phenytoin from happening in order to provide timely therapy. In East Asians, *CYP2C9\*3* and *HLA-B\*15:02* alleles were reported to be associated with phenytoin-induced eruption; however these results were not readily applicable for Japanese population given the small sample sizes used in those previous studies for *CYP2C9* and also the extremely low frequency of *HLA-B\*15:02* in Japanese population. In this study, we aimed to investigate the risk alleles in the Japanese population using the samples and data of BioBank Japan Project. A total of 747 patients were selected for the analyses, comprising 24 cases and 723 tolerant controls. Cases are the patients who developed, and tolerant controls are those who did not develop the phenytoin-induced eruption. Among the 24 patients that had a phenytoin-induced eruption one was with erythema multiforme, four were with erythema, one with erythroderma, one with fixed drug eruption, and two with maculopapular exanthema. We genotyped DNA samples of cases and tolerant controls for *CYP2C9\*3*, *CYP2C19\*2* and *CYP2C19\*3*, by multiplex PCR, which was followed by Invader assay. *HLA-B* four-digit allele genotyping was performed on a Luminex 200 system, using a WAKFlow HLA typing kit. Subsequently, case-control association tests were conducted on the genotype data for the *CYP2C9\*3*, *CYP2C19\*2*, *CYP2C19\*3* and *HLA-B* alleles. We have found that *CYP2C9\*3* carrier status was significantly associated with phenytoin-induced eruption ( $P=0.0022$ , odds ratio 7.1, five carriers in cases (21%) and 26 carriers in tolerant controls (3.6%)). Additionally, significant association was identified for the *HLA-B\*51:01* ( $P=0.010$ , odds ratio 3.2, nine carriers in cases (38%) and 114 carriers in tolerant controls (16%)). Lastly, the area under the receiver operating characteristic curve was improved by 10% by including the *HLA-B\*51:01* as well as *CYP2C9\*3* in the predictive model to predict the risk of eruption which was developed by using multiple logistic regression analysis. Thus, we conclude that checking the carrier status of *CYP2C9\*3* and *HLA-B\*51:01* alleles before starting phenytoin administration can potentially reduce cases of eruption in clinical practice. This work was supported by JSPS Core-to-Core Program, A. Advanced Research Networks.

## PgmNr 476: A comprehensive landscape of CYP2D6 variation across 30,000 individuals.

**Authors:**

S. Luo; R. Jiang; J. Dhaliwal; S. Jacobs; W. Lee

[View Session](#) [Add to Schedule](#)

**Affiliation:** Helix, San Carlos, CA.

---

More than one hundred CYP2D6 star alleles (haplotypes of the CYP2D6 gene) are listed in the Pharmacogene Variation Consortium (PharmVar) database, but commercial pharmacogenomic analyses only test for a minor fraction of these, often missing alleles that are common in certain populations. This is due, in part, to structural rearrangements in CYP2D6 (such as the CYP2D6\*36 and CYP2D6\*68 alleles) that require novel analytical tools and custom assay design to guarantee high confidence allele calls. These arrangements cannot typically be ascertained by off-the-shelf assays. This gap between alleles that are known to exist and alleles that are reported in tests can lead to mischaracterization of an individual's CYP2D6 genotype. In the clinical setting, this misclassification may result in prescriptions that are harmful to the patient. In the research setting, misclassification means poor quality input data for association tests. We have developed a clinically validated pipeline that calls 106 CYP2D6 star alleles listed in the PharmVar database, including CYP2D6\*36, CYP2D6\*59, and CYP2D6\*68, which occur with nontrivial frequency but are not typically reported. This pipeline is built on Helix's Exome+ assay and achieves over 99% accuracy without requiring orthogonal assay supplementation. Here we present the results of applying this pipeline to 30,000 exomes, providing an unprecedented characterization of CYP2D6 genotypic diversity in the US population.

## PgmNr 477: Analyses of core and extended pharmacogenes in sickle cell disease patients in Africa.

### Authors:

K. Mnika<sup>1</sup>; C.D. Bope<sup>1,3</sup>; G.K. Mazandu<sup>1,4</sup>; E.R. Chimusa<sup>1</sup>; V. Nembaware<sup>1</sup>; C. Dandara<sup>1,2</sup>; A. Wonkam<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Human Genetics, Department of Pathology; 2) Institute of Infectious Disease and Molecular Medicine (IDM), Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa; 3) University of Kinshasa, Faculty of Sciences, Departments of Mathematics and Computer Sciences, Kinshasa, Democratic Republic of Congo; 4) African Institute for Mathematical Sciences (AIMS), Cape Town, South Africa

---

Sickle cell disease (SCD) is a multisystem genetic disorder that results in organ damage, specifically in the cardiovascular system (stroke, acute chest syndrome, and kidney dysfunctions), and increased susceptibility to infections. Thus, most SCD patients are subjected to the long term administration of hydroxyurea (HU) and opioid therapy in addition to various other medications with variable responses to such therapies. High variability in individual responses to HU treatment and other therapies of importance to SCD have been reported. The extend to which polymorphisms in genes involved in drug absorption, distribution, metabolism and excretion (ADME) contribute in-part to the high heterogeneity of drug responses has not been explored in SCD patients from African populations.

This study focuses on 32 core and 267 extended pharmacogenes from the PharmacoScan platform to systematically investigate genetic variations in ADME genes in Cameroonian population. Specifically, we (1) assess minor allele differences in SCD patients versus ethnically matched Cameroonian controls (2) evaluate variant associations to a range of clinical factors: a “long survivor group” (age over 40 years), a “stroke group” (at least one episode of overt stroke), and a “random group” (patients younger than 40 years with no cerebrovascular disease) and (3) assess correlations between variants to endophenotypes of therapeutic importance (vaso-occlusive crises (VOC), HU, fetal hemoglobin (HbF) and stroke) to variants in the core and extended genes.

No statistically significant differences in allele frequencies were detected between SCD cases and controls, also no correlation between clinical variability and variants in the core genes was detected. Two SNPs in the pharmacogenes were significantly associated with VOC (*SLCO4A1*- rs118042746,  $p=1.21e-07$ ; *UGT1A10*, *UGT1A8*- rs10176426,  $p=1.22e-07$ ). To our knowledge this is the first study to report an association between these two SNPs and VOC. In particular no correlation was detected between variants in the pharmacogenes and key therapeutic indicators such as HbF and variability in response to HU. This finding highlights the lack of influence of core and extended pharmacogenes in ADME in African populations, including SCD patients. Therefore, it supports the need to invest in research exploring population-specific drug design, targeting and efficacy, for improved clinical management of patients of African descent.

## PgmNr 478: The clinical utility of self-identified race/ethnicity compared to genetic ancestry for predicting drug response in U.S. populations.

### Authors:

S.D. Nagar<sup>1,2,3</sup>; A.B. Conley<sup>1,2,3</sup>; E.T. Norris<sup>1,2,3</sup>; K.L. O'Neal<sup>1</sup>; W.N. Martin III<sup>1</sup>; I.K. Jordan<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA; 2) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA; 3) PanAmerican Bioinformatics Institute, Cali, Colombia

---

The practical utility of race and ethnicity in clinical settings remains a matter of considerable controversy. We explored this issue in the context of pharmacogenomic (PGx) variants, which can be used to predict how individual patients respond to drugs. Previous studies have shown that PGx variant frequencies differ substantially among genetic ancestry groups, suggesting group-specific response probabilities for a number of commonly prescribed medications. However, the extent to which self-identified race/ethnicity (SIRE) actually serves as proxy for genetic ancestry has been called into question, particularly for modern populations with high levels of admixture. The goal of this study was to evaluate the relative utility of SIRE versus genetic ancestry for predicting drug response in a modern, cosmopolitan population. We analyzed a cohort of 8,629 individuals from the United States (US), for whom we had both SIRE information and whole genome genotypes. We focused on the three largest SIRE groups in the US: White (5,927), Black (1,527), and Hispanic (1,174). Whole genome genotypes were used to characterize individuals' continental ancestry fractions – European, African, and Native American – and individuals were clustered into three groups according to either (1) their SIRE labels or (2) their continental ancestry profiles. We then explored the extent to which PGx variant frequencies differ among the SIRE groups compared to genetic ancestry groups. SIRE and continental ancestry are highly correlated for the major US race/ethnic groups; continental ancestry alone predicts individuals' SIRE with 96.24% accuracy. The relationship between SIRE and ancestry differs among groups, with the highest correlation seen for self-identified Whites and the lowest seen for Hispanics. There is substantial variation of PGx variant frequencies among both the SIRE and ancestry defined groups, consistent with the influence of population structure on drug response. Use of continental ancestry, an objective measure of genetic diversity, does not yield an overall increase in PGx variation compared to the more subjective SIRE labels. Nevertheless, the utility of SIRE as a PGx predictor is dependent on the specific group; racial groups carry more information than the pan-ethnic Hispanic label. Our results demonstrate that SIRE contains clinically relevant information on predicted responses to certain medications, which remains useful in the absence of patient genetic data.

## PgmNr 479: Enrollment of diverse populations in the INGENIOUS clinical trial.

### Authors:

E. ShahWilliams<sup>1</sup>; K.D. Levy<sup>2</sup>; Y. Zang<sup>3</sup>; A.M. Holmes<sup>4</sup>; C. Stoughton<sup>6</sup>; P. Dexter<sup>2,5</sup>; T.C. Skaar<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana, USA; 2) Department of Medicine, Indiana University, Indianapolis, Indiana, USA; 3) Department of Biostatistics, Indiana University, Indianapolis, Indiana, USA; 4) Fairbanks School of Public Health, Indiana University-Purdue University, Indianapolis, Indiana, USA; 5) Regenstrief Institute for Health Care, Indianapolis, Indiana, USA; 6) Department of Urology, IU Health, Indianapolis, Indiana, USA

---

Recruitment of diverse populations and patients living in Medically Underserved Areas and Populations (MUA/P's) into clinical trials is a considerable challenge. For example, representation of African Americans in pharmacogenetic trials are often low and insufficient to identify genetic variation within and between racial populations. Thus, less information is often available about genetic variation that impacts drug responses in these patients. To identify variables that predict enrollment in a diverse underserved population, we analyzed the recruitment data from the INGENIOUS pharmacogenomics implementation clinical trial. We used a predictive logistic model to test variables in patients from the Eskenazi Health (EH) and Indiana University Health (IUH) care systems that predict successful enrollment into the trial, after patients were successfully contacted by research recruiters. We also evaluated the reasons why clinical trial eligible patients refused enrollment. At EH, 1) African Americans were less likely to refuse the study than nonHispanic-whites, (OR 0.68, 95% CI: 0.53-0.86),  $p=0.002$ . 2) Other minorities were more likely to refuse the study than nonHispanic-whites, (OR 1.58, 95% CI: 1.02-2.45),  $p=0.04$ . 3) When age increased by 1 year, the odds of refusing the study increased by 2%, (OR 1.02, 95% CI: 1.02-1.03),  $p<.001$ . At IUH, 1) African Americans were also less likely to refuse the study than nonHispanic-white participants (OR 0.64, 95% CI: 0.53-0.76),  $p<.001$ . 2) As age increased by 1 year, the odds of refusing the study also increased by 2% (OR 1.02, 95% CI: 1.02-1.03),  $p<.001$ . 3) Females were less likely to refuse the study than males (OR 0.81, 95% CI 0.72-0.95),  $p=0.01$  and 4) those not living in MUA/P's were less likely to refuse the study than those living in MUA/P's (OR 0.81, 95% CI 0.70-0.94),  $p=0.007$ . Barriers to enrollment from prospective patients include a lack of interest, being too busy, transportation and illness, and rarely a lack of trust. The successful recruitment of diverse patients was likely due in part to the exposure to many research studies across the sites and recruitment by diverse and skilled clinical researchers. In conclusion, diverse and underserved populations can be effectively recruited to pharmacogenomic clinical trials and addressing barriers that prospective patients reported may increase participation across all populations, even more.



## **PgmNr 480: CC-CKAT: A fast and powerful method for SNP-set joint test of prognostic and predictive effects in pharmacogenomics studies.**

**Authors:**

J. Shen; H. Zhang; D. Mehrotra

[View Session](#) [Add to Schedule](#)

**Affiliation:** Merck & Co., Inc., Kenilworth, New Jersey.

---

Many statistical methods have been developed for SNP-set analysis in sequencing studies, mainly focusing on discovering prognostic genetic markers. However, a patient's clinical outcomes may be influenced by either prognostic, predictive or both factors in pharmacogenomics (PGx) scenarios. Very few statistical methods were developed to jointly test the SNP-set-based main genetic effect and the genotype by treatment interaction effect. Here we propose a Cauchy Combined Composite Kernel Association Test (CC-CKAT), a fast and powerful composite kernel machine regression method for the joint effect test by leveraging the efficient computation advantage of the recently developed Cauchy p-value combination method. First, an adaptive weighting method is proposed to construct the multiple composite kernels capturing the potentially unbalanced strength proportions of the main effect and interaction effects and then the individual composite-kernel-level p-values are combined using the Cauchy p-value combination method. The proposed CC-CKAT method is compared with the existing method rareGE and another method CKAT we have recently developed, which adopts the same composite kernel machine regression framework; but uses the minimal p-value (minp) approach to calculate final p-value. Through analysis of extensively simulated data and the whole exome sequencing data from a Merck PGx study with the *Clostridium difficile* Infection recurrence in patients treated with bezlotoxumab, we demonstrate that the proposed CC-CKAT provides better type I error control and power performance across multiple scenarios and runs much faster than the alternative methods.

## PgmNr 481: Genetic correlations and exploration of uterine fibroid clinical phenome in black and white women.

### Authors:

B. Mautz <sup>1,2</sup>; S.H. Jones <sup>3</sup>; E.S. Torstenson <sup>1,2</sup>; J.N. Hellwege <sup>2,4</sup>; T.L. Edwards <sup>1,2,3</sup>; D.R. Velez Edwards <sup>2,3,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN; 4) Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) Division of Quantitative Sciences, Department of Obstetrics and Gynecology,; 6) Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN

---

Uterine fibroids affect up to 70% of women by menopause, disproportionately impacting black females. Prior studies have investigated a limited set of clinical factors associated with fibroid risk in both groups. Moreover, the genetic relationship underpinning fibroids and other diagnoses remain unexamined. Electronic health records (EHR) and publicly available genome-wide association study (GWAS) data provide a unique opportunity for a comprehensive, agnostic investigation of the fibroid phenome and genetic correlations. Utilizing the Vanderbilt University Medical Center EHR database, cases and controls in black (N = 3,568 cases; 12,521 controls) and white women (N = 7,577; 60,296) were identified using a validated method verified by imaging. We conducted race-stratified phenome-wide association studies (PheWAS) to test for association of fibroid case/control status with diagnostic code-based outcomes in models adjusted for body mass index (BMI) and age at diagnosis. We grouped outcomes into standard categories and tested for over- and under-representation and effect direction using hypergeometric and sign tests. We identified publicly available GWAS for significant PheWAS associations and known fibroid risk factors such as BMI and hypertension. Using LD Score Regression and our own fibroid GWAS summary statistics, we estimated the genetic correlation between traits. We detected 228 and 444 diagnoses that were significantly associated with fibroids in blacks and whites, respectively. Associations were over-represented in the genitourinary category (blacks  $P = 1.03 \times 10^{-20}$ ; whites  $P = 1.03 \times 10^{-14}$ ). Odd ratios of significant diagnoses in dermatological, genitourinary, endocrine/metabolic, and musculoskeletal categories trended in the risk direction. Across racial groups the most significant PheWAS associations were known fibroid symptoms (e.g. excessive menstruation,  $P < 1.0 \times 10^{-274}$ ; dysmenorrhea,  $P < 2.58 \times 10^{-152}$ ). We also detected numerous novel associations including malignant and benign cancers, endometriosis, and coronary atherosclerosis ( $P < 2.56 \times 10^{-15}$  for blacks and whites). We did not detect any significant fibroid-trait genetic correlations in either race, though we observed a suggestive negative relationship of fibroids with type 2 diabetes in blacks ( $r_g = -1.21$ ,  $P = 0.09$ ). These results provide novel insight into the fibroid phenome. More thorough analyses of genetic correlations are underway.

## PgmNr 482: Comparison of genetic testing approaches for implementation of preemptive pharmacogenomics.

### Authors:

K. Crisamore <sup>1</sup>; S. Adams <sup>2</sup>; P. Empey <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Pittsburgh School of Pharmacy, Pittsburgh, PA; 2) Shenandoah University School of Pharmacy, Fairfax, VA

---

**Background:** Preemptive pharmacogenomics (PGx) leverages lifelong reusability of genomic data to guide downstream prescribing and individualization of medications. Advancements in technologies now offer varied approaches for testing of PGx. The present study sought to evaluate performance of whole genome sequencing (WGS) and probe-based testing by concordance of variant-level calls, particularly in clinically important genes, with a CLIA-certified reference lab.

**Methods:** In an IRB-approved volunteer study, blood and saliva samples were obtained. Control DNA samples were purchased (Coriell Institute, Camden, NJ) to test rare and structural variants. Samples were blinded and analyses were performed by a CLIA-certified reference lab (n=94, OpenArray, Indianapolis, IN), ThermoFisher Scientific (n=94, PharmacoScan, Santa Clara, CA), and UPMC Genome Center (n=72, WGS, Pittsburgh, PA). Raw data, VCF files generated through Axiom Analysis Suite v4.0, and genomic VCF files were compared between labs, respectively. Concordance was determined as number of matching variant-level calls divided by total calls, excluding undetermined, failed, and no calls.

**Results:** A total of 111 and 114 variants analyzed by PharmacoScan and WGS, respectively, were evaluated for concordance with the CLIA-certified reference lab's clinical array. Variants were represented in 10 clinically important genes with high levels of evidence. *CYP2D6* copy number variation (CNV) was evaluated between PharmacoScan and the CLIA reference lab. Overall variant-level concordance between PharmacoScan and the CLIA reference lab (n=91) was 99.7% and *CYP2D6* CNV reached 100% concordance. Paired blood and saliva samples (n=20) tested on PharmacoScan were 99.7% concordant. Variant-level concordance between WGS and the CLIA reference lab (n=72) reached 96.14% and differed primarily in *CYP2D6*.

**Conclusion:** Both testing approaches had high overall concordance between platforms. PharmacoScan was also highly concordant within platform in paired blood and saliva samples. Given clinical importance of *CYP2D6*, probe-based testing may be considered more favorable than short-read sequencing for PGx testing at this time. Integration of the PharmacoScan solution for clinical implementation of PGx is underway.

## PgmNr 483: Precision medicine 2.0: Integrate precision medicine into clinical practice.

**Authors:**

Z. Gu; G. Hu

[View Session](#) [Add to Schedule](#)**Affiliation:** Admera Health, South Plainfield, NJ

---

**Background:** Physicians are becoming increasingly interested in electronic clinical decision support (CDS) for precision medicine. Improved decision support delivered through use of real-time information that combines clinical and pharmacogenomics (PGx) data has the potential to improve physicians' ability to make personalized drug therapy decisions. CDS embedded in the Electronic Health Records (EHR) systems, that connected with Laboratory Information System (LIS), might provide a potential for delivering such support of interpreting patient specific genetic data for precision medicine.

**Methods:** A PGx CDS was developed and built to support personalized drug therapy. The PGx knowledgebase was built based on FDA, CPIC, PharmGKB data to provide accurate clinical interpretation. Multiple EHR/EMR systems were integrated with LIS utilizing HL7 messages, system API, etc. The prototype CDS was developed within Cerner environment. AWS instances and S3 buckets were used to build mobile application and data warehouse for a crisp, secure, HIPAA compliant APP which can be accessed by physicians and patients.

**Results:** An active CDS was implemented with alert messages in the Cerner's medication order entry context to provide support for prescribing decisions. For a patient who had Admera's PGxOne Plus test, the patient's personal results will be incorporated within Cerner's system, where a physician would receive an alert message if the prescribing medication has a non-normal clinical interpretation, such as consider alternatives or dosage change. A semi-active CDS was implemented with drug interaction warning messages in the Admera APP's Modify Current Medication menu. A physician could modify his/her patients' current medication list to check if there would be any drug to drug interaction, drug to food interaction, drug to alcohol interaction, etc. before finalizing.

**Discussion:** With the experience of prototype CDS within Cerner environment, an improved solution to host a data warehouse with Fast Healthcare Interoperability Resources (FHIR) API that contains clinical interpretation data from LIS has been proposed. This solution can provide standardized interpretation data to different systems with minimum compatibility requirement. This in turn will constitute the core component of integrating precision medicine into clinical practice.

## PgmNr 484: Genotyping On ALL patients (GOALL): Clinical implementation of high-throughput genotyping arrays.

### Authors:

J.G.J. Rooij; J.M.H. Verkerk; B. Sedaghati-khayat; L. Broer; J.H. de Vries; G.M. van Dijk; J.B.J. van Meurs; A.G. Uitterlinden

[View Session](#) [Add to Schedule](#)

**Affiliation:** Internal Medicine, Erasmus Medical Center, Rotterdam, Zuid-Holland, Netherlands

---

**Introduction:** Genetic testing is increasingly used in standard clinical practice for most clinical disciplines. However, genetic tests are expensive and/or must be rerun when additional clinically relevant variants/genes are identified. Dense genotyping arrays (such as GSA or PMDA) cover a large portion of genetic variation and provide a cost-effective, high-throughput and highly standardized alternative. In the GOALL project, we are determining the clinical utility and applications of genotyping arrays.

**Methods:** We investigate the clinical utility of the GSAMD\_v2 and PMDA arrays by: 1) technical validation of common (vs. WES) and rare variants (vs. clinically determined genotypes by different methods), 2) prospectively offering polygenic risk scores (PRS) as primary utility in daily clinical practice, 3) prospectively counseling and testing eligibility of feeding back secondary findings (pharmacogenomics, ACMG mutations) using array based genotyping. We are piloting these applications in collaboration with the departments of Oncology (breast cancer), Ophthalmology (age-related macular degeneration), Clinical Chemistry (pharmacogenomics), Clinical Genetics (counseling) and Public Health within the Erasmus Medical Centre.

**Results:** The GSAMD\_v2 array contains 12,755 clinical variants from the ClinVar database. Our preliminary results suggest that >90% of the clinical mutations and >80% of the pharmacogenetic markers routinely determined in our hospital are covered by the array design (e.g. *APOE*, *MTFHR*, *F5* and *HLA/CYP450* etcetera). We are currently selecting carriers of clinically relevant mutations to check genotyping concordance. Comparison of array vs. WES of 197 population samples shows an overall non-reference concordance of 98%, and 95% for singletons among ~30,000 directly genotyped variants and respectively 97% and 93% among ~150,000 imputed variants (QUAL>0.8).

**Discussion:** Our preliminary results suggest that a portion of current hospital-wide clinical genetic testing can cost-effectively be replaced by array-based genotyping. In addition to reporting highly concordant results, array-based genotyping allows additional reporting of several PRS and pharmacogenomics (*HLA/CYP450*). This data can easily be stored and re-analyzed when additional genetic loci or PRS are identified, or unblinded when a patient needs further genetic testing not part of the initial primary clinical question.

## PgmNr 485: Harmonizing clinical sequencing and interpretation for the eMERGE III Network.

### Authors:

E. Venner<sup>1</sup>; H. Zouk<sup>2</sup>; D. Muzny<sup>1</sup>; D. Murdock<sup>1</sup>; D.R. Crosslin<sup>4</sup>; A.S. Gordon<sup>5</sup>; M. Harden<sup>3</sup>; N. Lennon<sup>3</sup>; H.L. Rehm<sup>2</sup>; R.A. Gibbs<sup>1</sup>; eMERGE Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) HGSC - Baylor College of Medicine, Houston, Texas.; 2) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA; 3) The Broad Institute of MIT and Harvard, Cambridge, MA; 4) Departments of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 5) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL

---

As Precision Medicine moves closer to routine clinical practice, new methods to produce and deliver complex, heterogeneous data to physicians and patients are required. As part of the eMERGE III project, we have built a network of eleven participant collection sites and two clinical genetic testing laboratories, which has served as an ideal environment to implement, develop and evaluate such methods. DNA capture panels targeting 109 genes and 1500 SNP sites were used for testing DNA from 25,015 participants and the sample collection, data generation, variant interpretation, report generation, report delivery and data storage were each harmonized. A compliant and secure environment was provided for storage of clinical reports, which contain participant PHI, and we employed system of ongoing review and reconciliation of clinical variant interpretations. Variant information was also deposited into ClinVar. Throughout the project, a high level of communication and data sharing between clinicians and investigators provided feedback on these systems.

Two hundred and six participants had positive diagnostic findings relevant to the indication for testing while 1,276 had additional/secondary findings of medical significance deemed to be returnable, establishing data return rates for other testing endeavors. This study accomplished the creation of harmonized structured genomic results that is enabling efforts to integrate these results into multiple electronic health record systems, setting the stage for clinical decision support to enable genomic medicine. Clinical sites also evaluated pharmacogenomic data, with one opting to incorporate it into their medical records. The established processes enable different sequencing sites to harmonize technical and interpretive aspects of sequencing tests, a critical achievement towards global standardization of genomic testing. Many eMERGE protocols and tools are available for widespread dissemination.

## PgmNr 486: Mapping impact of plasma proteins on dental diseases: A Mendelian randomization study.

### Authors:

D. Shungin<sup>1,2</sup>; G. Hindy<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Institute of MIT and Harvard, Cambridge, Massachusetts; 2) Umea University, Sweden; 3) Lund University, Sweden.

**Background** Dental diseases account for a large morbidity burden and healthcare spending, while their proteomics characterization remains largely unknown despite that variation in multiple plasma proteins represents a variety of important biological mechanism and is a major source for drug targets. Recently published studies identified genetic basis of human plasma proteome, potentiating studies on causal inference between proteins and complex disease traits with a help of Mendelian Randomization approach (MR). Emergence of such studies opens new horizons to generate hypotheses about potential novel drug targets for complex diseases, including for key dental diseases. The aim of the study was to assess the causal effect of plasma proteins on oral health

**Methods** We used two-sample MR approach to estimate effects of proteins for the two major dental diseases: caries and periodontitis. We utilized summary statistics from meta-analysis of GWAS with clinical and self-reported data for periodontitis and caries (Shungin et al, 2019), and summary statistics from GWAS for proteins (Sun et al, 2018). We constructed instruments for each protein using common variants (MAF > 0.01) associated with the plasma protein levels at two significance thresholds ( $p < 0.01$  and  $p < 1e-05$ ) and located in the respective encoding genes or 300 kb flanking regions followed by pruning ( $r^2 < 0.1$ ). Two sample inverse-variance weighted (IVW) MR was used to test for causal associations between proteins and dental phenotypes. Sensitivity analyses using MR-Egger and the weighted median estimator were used to correct for possible pleiotropic bias.

**Results** Around 1100 proteins had 3 or more SNPs that were used as instruments in the MR analyses. We identified several proteins implicated in caries and periodontitis at the stringent significance level ( $p < 1e-05$ ). Among the prioritized causal associations, we observed targets that have been previously implicated in etiopathology of dental diseases, such as *LCT* ( $\beta_{IVW} = 0.014$ ,  $P = 1.55e-17$ ) and *AMY1* ( $\beta_{IVW} = 0.03$ ,  $P = 9.5e-11$ ) for caries, and *SIGLEC14* ( $\beta_{IVW} = -0.01$ ,  $P = 7.2e-06$ ) for periodontitis, confirming validity of the utilized prioritization approach for pQTLs. We have further identified previously not reported proteins implicated in caries and periodontitis that might serve as potential drug targets.

**Conclusion** This study highlighted causal roles of protein biomarkers for the two key dental diseases using MR analysis, pointing at potential drug targets.



## PgmNr 487: *GRK5* polymorphism affects $\beta$ 2-agonist response in preterm labor patients.

### Authors:

J. Chung <sup>1</sup>; S. Hwang <sup>1</sup>; J. Yee <sup>2</sup>; J. Park <sup>2</sup>; K. Lee <sup>3</sup>; Y. Kim <sup>4</sup>; H. Gwak <sup>2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) College of Pharmacy, Hanyang University, Ansan, Gyeonggi-do, 15588, Korea; 2) College of Pharmacy and Division of Life & Pharmaceutical Sciences, Ewha Womans University, 52 Ewhayeodae-gil Sedaemun-gu, Seoul 03760, Korea; 3) College of Pharmacy, Chungbuk National University, Osongsaengmyeong 1-ro, Osong-eup, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do 28160, Korea; 4) Department of Obstetrics and Gynecology, Ewha Womans University Mokdong Hospital, 1071 Anyangcheon-ro, Yangcheon-gu, Seoul 07985, Korea

---

**Purpose:** Ritodrine, a tocolytic  $\beta$ 2-agonist, has been used extensively in Europe and Asia despite its safety concerns. This study aimed to examine the contribution of most frequent single nucleotide polymorphisms of G-protein Kinase 5 (*GRK5*) to efficacy and adverse effect of ritodrine in preterm labor patients.

**Methods:** A total of 164 preterm labor patients were included in this study. Seven single nucleotide polymorphisms of the *GRK5* gene (rs915120, rs2230345, rs2230349, rs7923896, rs1020672, rs4752308, and rs4752292) were assessed. The primary endpoints were time to delivery and proportions of women who remained undelivered. The secondary endpoint was ritodrine-induced adverse drug events (ADEs). To predict the possible effects of given variants on splicing, *in silico* analysis was performed.

**Results:** The mutant type homozygote carriers of rs4752292 and rs1020672 had 0.6 times the hazard of delivery than those with the wild-type allele (95% CI, 0.41~0.98 and 0.38~0.97, respectively). Proportions of patients who remained undelivered at 7 days after initiation of ritodrine therapy were significantly higher in patients with variant homozygote carriers of rs4752292 and rs1020672 than the wild-type allele carriers. In addition, patients with variant homozygotes of rs4752292 and rs1020672 had 2.5- (95% CI, 1.15-5.28) and 2.3-fold (95% CI, 1.06~5.13) ADEs compared to those with the wild-type carriers. *In silico* analysis showed that both rs4752292 and rs1020672 had the potential to affect mRNA splicing by alteration of splicing motifs.

**Conclusion:** This study demonstrates that the pharmacodynamics of  $\beta$ 2-adrenergic receptor targeted therapy is associated with *GRK5* polymorphisms in patients with preterm labor.



## PgmNr 488: Leveraging tensor decomposition to extract biological processes underlying drug treatments.

### Authors:

Y. Zhong <sup>1</sup>; T. De <sup>1</sup>; Y. Xu <sup>2</sup>; Y. Luo <sup>3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Pharmacology, Northwestern University, Feinberg School of Medicine, Chicago, Illinois.; 2) University of Chicago; 3) Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, Illinois.

---

Decomposition of transcriptome matrixes has been useful to understand the biological processes underlying tumor heterogeneity, tissue specificity and responses to perturbations. Here we apply a Bayesian tensor decomposition approach to the transcriptome (three axes correspond to individual \* drugs \* genes) measured on 60 African American (AA) primary hepatocytes before and after treating with 6 different enzyme-inducing drugs in order to understand commonalities and specificities underlying transcriptomic changes to drug exposures.

We identified 5 robust latent components (C1-C5) through hierarchical clustering of the components across 10 runs. The clustering of the robust components confirms known drug similarity (Carbamazepine and Phenytoin have a common mechanism of action for the treatment of epilepsy). Omeprazole has the largest divergence from other drug expression profiles and is enriched in C5 which was highly weighted for differentially expressed drug metabolism genes, *CYP1A1* and *CYP1B2*. *CYP1A1* is responsible for the metabolism of estrogens, which are known to differ between populations. Altered enzyme activity has been linked to breast, colon and ovarian cancers, all of which show profound health disparities. C1 highly differentiates Omeprazole from phenytoin and was enriched in genes related to response to cytokines. C3 corresponds to gender differences which feature *XIST*. Interestingly, the weight of C3 is higher in drug treatment conditions compared with baseline, suggesting that gender difference in gene expression is enlarged upon drug treatment.

We also performed genome-wide association analysis using these derived components as the phenotype. We found associations between C1 and rs2518683 and the SNP is previously associated with platelet count ( $p=2.649e-07$ ). C1 has the highest component in phenytoin which is known to cause thrombocytopenia (decreased platelet count). Importantly, this cardiovascular phenotype is highly relevant to AAs.

Together, our tensor decomposition analysis, 1) illustrates the structure of hierarchical similarity of expression profiles with drug treatment 2) extracts rich features specific to individuals and drugs as well as clinically important underlying genes. These findings highlight the potential of applying tensor decomposition to multi-omic datasets such as GTEx and CMap to deconvolute the effects of the individual, condition and their interaction with respect to the transcriptome variation.

## PgmNr 489: Multi-ethnic single nucleotide polymorphism (SNP)-based screening of pharmacogenomic HLA alleles implicated in drug-induced hypersensitivity reaction.

### Authors:

P. Nicoletti <sup>1,2</sup>; Y. Seki <sup>1</sup>; L. Shi <sup>1,2</sup>; R. Wallsten <sup>1</sup>; A. Owusu Obeng <sup>2,3</sup>; N. Cody <sup>1,2</sup>; M. Smith <sup>2</sup>; A.J. Chen <sup>1</sup>; E.R. Scott <sup>2</sup>; R. Sebra <sup>1,2</sup>; R. Kornreich <sup>1,2</sup>; E.E. Schadt <sup>1,2</sup>; L. Edelmann <sup>1,2</sup>; S.A. Scott <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York.; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 3) Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029

Drug-induced hypersensitivity reactions are life-threatening adverse events that have been strongly associated with specific human leucocyte antigen (HLA) alleles. Among the reported pharmacogenomic *HLA* risk alleles, clinical practice guidelines and/or FDA/EMA testing recommendations are currently available for *HLA-B\*57:01* (abacavir), *HLA-B\*15:02* (carbamazepine, phenytoin) and *HLA-A\*31:01* (carbamazepine), and *HLA-B\*58:01* (allopurinol). However, routine *HLA* typing for pharmacogenomic risk prediction is not cost-effective due to the technical challenges with interrogating the polymorphic *HLA* gene region. To facilitate pharmacogenomic *HLA* allele screening, previous studies identified single nucleotide polymorphisms (SNPs) in linkage disequilibrium with the most relevant *HLA* risk alleles; however, these proxy-SNPs have inadequate sensitivity across multi-ethnic populations and poor specificity, which has limited their use in clinical practice. To identify a multi-ethnic panel of proxy-SNPs for *HLA-B\*57:01*, *HLA-B\*15:02*, *HLA-A\*31:01* and *HLA-B\*58:01*, publicly available whole genome sequencing and *HLA* typing data from the 1000 Genomes (1KG) Project were interrogated. The sensitivity and specificity of all SNPs in the MHC region ( $n=608,256$ ) were initially assessed across a multi-ethnic subset of 1KG subjects ( $n=1079$ ), which identified a panel of 18 proxy-SNPs that detected the four pharmacogenomic *HLA* risk alleles with a mean sensitivity and specificity of 99%. Selected false positive and negative 1KG samples were clarified by confirmatory genotyping (Agena Bioscience), which resulted in the proxy-SNP panel having sensitivity and specificity >99% for *HLA-B\*57:01* ( $N_{\text{carriers}}=69$ ;  $N_{\text{SNPs}}=8$ ), *HLA-B\*15:02* ( $N_{\text{carriers}}=65$ ;  $N_{\text{SNPs}}=3$ ) and *HLA-A\*31:01* ( $N_{\text{carriers}}=88$ ;  $N_{\text{SNPs}}=4$ ); and >95% for *HLA-B\*58:01* ( $N_{\text{carriers}}=77$ ;  $N_{\text{SNPs}}=3$ ). The validity of the proxy-SNP panel was confirmed in an independent 1KG subset ( $n=1425$ ), and in the combined cohort ( $n=2504$ ) the majority of proxy-SNPs had 100% sensitivity and 99% specificity for the four pharmacogenomic *HLA* alleles in the Asian, European and Hispanic populations. Importantly, the panel also included proxy-SNPs with 98% sensitivity and specificity for *HLA-A\*31:01* and *HLA-B\*58:01* in the African population. Taken together, these data indicate that the identified novel proxy-SNP panel could enable cost-effective and rapid genotype-based screening to predict pharmacogenomic *HLA* allele risk in diverse patient populations.

## PgmNr 490: Prioritizing drug targets from GWAS by training on Mendelian trait associations.

**Authors:**

E.A. King; J.W. Davis; J.F. Degner

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genomics Research Center, AbbVie, North Chicago, IL.

---

Drugs with therapeutic mechanisms supported by human genetic evidence are more likely to succeed in clinical development. We have recently shown that this increase in success depends on the strength of evidence for a specific causal gene, but most significant disease-associated variants from genome-wide association studies (GWAS) occur in noncoding regions of the genome and have uncertain causal genes. Candidate causal genes may be identified through physical distance to the variant, through follow-up experiments, or through leveraging publicly available resources on variant function. However, there remains a question of how to weight different evidence sources. We sought to develop models to estimate causal probabilities using the curated set of genetic associations from the Online Mendelian Inheritance in Man (OMIM) database. We posited that modest-effect regulatory GWAS variants would to a degree be phenocopied by more severe Mendelian mutations in the relevant causal gene. We take advantage of the EFO ontology to automatically generate a set of positive labels for GWAS candidate genes with OMIM genetic associations for similar traits. We train models to predict these labels from candidate gene link properties. We find such positive labels, though only applicable to a minority of significant GWAS associations, are significantly more frequent than predicted by chance, even when excluding OMIM entries for complex diseases. We estimate a decay in causal gene probability as a function of distance from the gene boundary that is similar to the relationship between distance and regulatory QTL position. We also find a significant, positive effect of colocalization of the GWAS signal with gene expression QTL and positive effects of coding variants in the gene. We show we can outperform models in which the closest gene is assumed to be causal and models in which the causal gene is randomly selected from a set of possible candidate genes when predicting OMIM positive labels. As further validation of our approach for utility in pharmaceutical applications, we use estimated causal gene probabilities derived from the model output to quantify human genetic evidence for target involvement in a set of drug indications. We find that accounting for human genetic evidence in this way improves our ability to distinguish drug targets from non-targets and increases the estimated effect of genetic evidence on target-indication pair approval probability.

## PgmNr 491: Efficient algorithm for choosing fewest number of reference samples required to validate clinical variants for return of results.

### Authors:

N. Rafaels <sup>1</sup>; T. Brunetti <sup>1</sup>; R.A. Mathias <sup>2,3</sup>; K.C. Barnes <sup>1</sup>; K. Crooks <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Colorado Center for Personalized Medicine, School of Medicine, University of Colorado, Aurora, Colorado.; 2) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA; 3) Department of Medicine, Johns Hopkins University, Baltimore, USA; 4) Department of Pathology, School of Medicine, University of Colorado, Aurora, CO

---

### Introduction

The Colorado Center for Personalized Medicine (CCPM), has worked with Illumina to design the Infinium CU Custom MEGA-Ex (CU-MEGA) genotyping array with 45,000 custom variants, including over 200 additional pharmacogenetic variants. CCPM proposes to return pharmacogenetic results directly from the CU-MEGA chip back to patients, which requires validation from control samples with previously determined genotype calls. We've developed an algorithm that selects the fewest number of reference samples required to accomplish this goal.

### Methods

We selected a set of variants of interest from CU-MEGA in common with 1000 genomes (TGP) Phase 3 data also in Coriell (2388 samples), and then split by sample. We then calculated genotype counts per sample per variant, such that each genotype count is either 0 or 1 using PLINK/Seq. Then for every subject, "Patient 0", we chose the sample with highest number of 1's discrepant with "Patient 0", replaced 0's from "Patient 0" with 1's from "Patient X", and repeated until the number of discrepancies is 0, and tallied the number of subjects required to reach 0. For each subject, we repeated 50 times to account for random ties. We then saved the "Patient 0" with the lowest tally for a list of Coriell samples to order. To reduce the number of files generated, we deleted files for "Patient 0" if the tally is greater than the lowest tally generated thus far. If there are multiple samples with the lowest tally, we further filtered by choosing a set of samples most unrelated with most ethnic diversity.

### Results

For this experiment, we chose 169 variants from PharmGkB in common with both CU-MEGA and TGP. The algorithm produced a 5-way tie with 78 samples minimally required to view all genotypes available in the Coriell and TGP databases. The most diverse dataset included 21 AFR, 6 AMR, 19 EAS, 17 EUR, and 15 SAS.

### Conclusions

This algorithm can be expanded to any set of variants, reference samples and/or genotyping panel. This algorithm cannot validate genotypes not included in the reference dataset. CCPM plans to utilize this algorithm iteratively as new genotyping platforms and new clinical information are released.

## PgmNr 492: Homology aware design facilitates microarray CNV detection within gene families.

### Authors:

K. Aull; C. Bruckner; R. Varma; C. Nelson; A. Roter; J. Schmidt; J. Gollub; Blood Transfusion Genomics Consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** Thermo Fisher Scientific, Santa Clara, California.

---

Detection of copy number variation (CNV) is challenging in members of highly homologous gene families. *RHD*, one of the two genes underlying the Rh blood group system, has very high homology to *RHCE*. The frequent deletion of *RHD* underlies the D-/- type, the most critical locus for red blood cell matching for transfusion after the *ABO* one. *CYP2D6*, involved in the metabolism of 25% of all prescribed drugs, is highly homologous to a pseudogene, *CYP2D7*.

Sequence similarity or identity between gene family members makes copy number calling very challenging. This is even more difficult when the off-target sequence match is to a region with its own copy number or SNP variation. *RHD* copy number calling suffers from the first problem. Although *RHCE* copy number changes are extremely rare, there are relatively few regions of *RHD* unique to this gene. *CYP2D6* copy number calling additionally suffers from the second problem, in that copy number variation in the *CYP2D7* pseudogene can be confused with copy number changes in *CYP2D6*. In both cases, careful probe selection improves results, while proactively eliminating less useful probes from the array yields better data while using less space.

Using data from the Applied Biosystems™ Axiom™ PMD Array, which has extensive coverage of both *RHD* and *CYP2D6*, we applied both empirical and analytical methods for CNV probe selection. A set of 360 samples from the HapMap and 1000 Genomes Project collections was studied. Samples were selected in part for *CYP2D6* copy number diversity, and *RHD* deletions were very common as well. Accuracy was evaluated by comparison to TaqMan assays and publicly available copy number data. We found that analytical methods that take into account the length of the genomic homology were predictive of CNV probe performance. Based on this and empirical performance metrics, we identified a reusable, space-efficient *RHD* CNV module with close to 100% accuracy in the samples studied. CNV calls in *CYP2D6* exon regions were also found to be highly accurate, validating this approach for further copy number detection-oriented arrays focused on complex gene families.

## **PgmNr 493: Genetics for pre-health undergraduate students in the era of advanced genomics and personalized medicine.**

**Authors:**

K. Makky

[View Session](#) [Add to Schedule](#)**Affiliation:** Biomedical Sciences, SC446, Marquette Univ, Milwaukee, Wisconsin.

---

In modern medicine and patient-centered treatment, clinical applications of genetics have evolved to be more than just understanding the cause of rare inherited diseases. With advancements in the field of genomics and genetic testing, many believe that future healthcare providers should also have the ability to understand and analyze personalized information based on a patient's genetic profile. This has prompted a shift in medical education, starting with pre-med pedagogy, with genetics curriculum being the center of this shift.

In the Department of Biomedical Sciences at Marquette University, we offer a human medical genetics course that follows the medical knowledge competencies proposed by the Association of Professors of Human and Medical Genetics (APHMG). Here, we discuss and share the curriculum offered by our department, which starts with an introduction to the human genomic structure and includes topics such as epigenetics, Mendelian disorders, complex genetic disorders, and pharmacogenetics. High impact teaching practices such as case studies are used, some of which will be shared. The objective is to start a conversation to solicit feedback and information from other clinicians, geneticists, genetic counselors, and educators by asking the following questions: What can biology educators do to prepare undergraduate students for their future professional career as health care providers? What educational practices will best help our pre-health undergraduate students? Do we have to reform genetics classes at the college level to better prepare our pre-health students?

## PgmNr 494: Pharmacogenetics beyond the drug label, with NIH's Medical Genetics Summaries (MGS).

### Authors:

A. Malheiro; L. Dean; M. Hoepfner; B. Kattman

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Center for Biotechnology Information, NLM, NIH, Bethesda, MD

---

Pharmacogenetic testing is an integral part of precision medicine. By acting on the results of a pharmacogenetics test, a clinician can identify responders and non-responders prior to or during therapy, avoid adverse events, and optimize the initial drug dose.

Currently, 262 drugs have pharmacogenetic information in the FDA label. However, this information is not provided in a standard format and may not be actionable (when to test may not be addressed, and dosing alterations may not be provided).

Because it is often at the clinician's discretion whether to test a patient before starting therapy, it is important to have access to a peer-reviewed, up-to-date summary of the latest literature, a quick review of the drug, and how to use the patient's genotype to guide therapy.

Medical Genetics Summaries (MGS, <https://www.ncbi.nlm.nih.gov/books/NBK61999>) unites the latest drug and gene information with dosing recommendations from the FDA and professional and medical societies. All 46 summaries include information from the FDA label, 18 have dosing recommendations from the Dutch Pharmacogenetics Working Group (DWPG), 14 have recommendations from the Clinical Pharmacogenetics Implementation Consortium (CPIC), and 3 have recommendations from the American Society of Clinical Oncology (ASCO).

By centralizing dosing guidelines, MGS provides supplemental information when the drug label lacks pharmacogenetic dosing recommendations. For example, for the chemotherapy drug capecitabine, both the FDA label and CPIC guidance agree there is no safe dose for individuals who lack DPD activity. However, for individuals who have partial DPD activity, CPIC provides dose recommendations whereas the FDA states there is insufficient data for dose alterations.

MGS simplifies the implementation of precision medicine by explaining the influence of genetic variant(s) on the metabolism and action of a drug, and linking to relevant tests in the NIH Genetic Testing Registry (GTR, <https://www.ncbi.nlm.nih.gov/gtr/>).

The GTR currently has 59,597 orderable genetic tests, voluntarily submitted by testing labs from across the world. Of these, 407 are pharmacogenetic tests, testing for at least one drug response, with some tests analyzing a panel of genetic variants, accounting for over 100 drug responses. Approximately 99 drugs from the FDA list have tests in GTR. There are also tests for 50 drugs not in the FDA list.

MGS and GTR are free, easy to use, and can help you optimize your care delivery.

## PgmNr 495: Family history vs. polygenic risk scores as the predictors of genetic risk.

### Authors:

R. Mägi; K. Läll; T. Laisk

[View Session](#) [Add to Schedule](#)

**Affiliation:** Estonian Genome Center, Institute of Genomics, Univ Tartu, Tartu, Estonia

---

Genome-wide association studies have discovered thousands of novel genetic associations for a wide variety of diseases. Even though most of the association effect sizes are small, genetic risk scores (GRS) combining genome-wide information into single risk estimators have shown to describe big enough proportions of genetic variability of diseases to be considered as tools for clinical practice. Despite of their prognostic abilities, GRS are still treated with skepticism in medical community and in many cases disease family history is considered to be the best possible tool for disease prediction. Also the prognostic value of GRS is disputed in the light of their low specificity.

Using Estonian Biobank data (N=150,000) which contains approximately 18,000 parent-child duos and several example disease risk scores (type 2 diabetes, coronary heart disease, early menopause), we show that even though parental GRS is a good predictor of offsprings GRS, there are cases where the offspring of high-risk parents has low genetic risk of the disease and vice versa. We studied if the offspring's genetic risks are lower compared to their parents (potentially due to prenatal purifying selection) but found no statistically significant difference. Additionally we compared the predictive values of GRSs with other risk factors already implemented into clinical practice (BMI, smoking status, concurrent diseases) and found that the predictive value of GRS is comparable with some phenotypic factors already in use.

As a conclusion we emphasize the potential of GRS in clinical practice, especially in the light of new era of precision medicine. We consider it as a cost effective tool for medical doctors for gaining additional insights into patient's possible health outcomes.



## PgmNr 496: Nashville Biosciences: Driving precision medicine with BioVU®.

**Authors:**

J. Wenke; E. Sundaram; L. Sundararajan; Y. Wu; J. Stover; L. Ekstrom

[View Session](#) [Add to Schedule](#)

**Affiliation:** Nashville Biosciences, Nashville, Tennessee.

---

The availability of genomic data has enabled the development of targeted therapies and personalized medicine. BioVU® is a biobank created by VUMC consisting of approximately 250K de-identified DNA samples linked to longitudinal electronic medical records (EMRs). BioVU® was created to enable exploration of the relationships among genetic variation, disease susceptibility, and drug responses. VUMC also has a database of approximately 3M de-identified EMRs for diverse VUMC patients called the Synthetic Derivative. Nashville Biosciences was created to expand the use of these VUMC resources to help industry advance drug discovery and development. Here, we describe three representative pharma client projects entitled: (1) Phenome wide association study (PheWASTM) drug target prioritization, (2) A synthetic control arm clinical trial, and (3) Identification of variants associated with protection from a rare disease using WES. (1) Using PheWASTM-based studies, we prioritized gene targets for early stage clinical development by identifying novel disease associations. A proprietary scoring system consisting of scientific and clinical metrics was implemented to select the strongest disease phenotype associations using a collaborative and iterative approach. We consulted published literature to develop a mechanistic hypothesis and recommend experimental next steps. (2) We created a cohort of patients with a rare lymphoproliferative disorder as a part of a synthetic placebo control arm for a phase III clinical trial. We reviewed EMRs and extracted detailed patient data including demographics, medical history, treatments, lab values, and survival data to enable analysis of treatment response, health resource utilization, and survival time. Designing a synthetic control arm for clinical studies from a rich database can reduce trial clinical costs and the time-period significantly. (3) We generated a virtual cohort of patients using clinical parameters including diagnosis codes for disease risk factors, findings from pathology and radiology reports, and lab values to sort patients into matched case and control groups for the disease of interest. Variants associated with protection from the chronic disease of interest were identified using WES. We also provided genotyping data for the de-identified patients to enable a GWAS. By leveraging real-world clinical and genomic data, we are advancing the discovery and development of personalized therapeutics and diagnostics.

## **PgmNr 497: The Taiwan Precision Medicine Initiative (TPMI): Incorporating pharmacogenetics in clinical practice.**

### **Authors:**

C.Y. Wei <sup>1</sup>; J.H. Yang <sup>1</sup>; M.F. Tsai <sup>1</sup>; C.C. Chen <sup>1</sup>; J.Y. Wu <sup>1</sup>; Y.T. Chen <sup>1</sup>; P.Y. Kwok <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute of Biomedical Science, Academia Sinica, Taipei, Taiwan; 2) Cardiovascular Research Institute, Institute for Human Genetics, and Department of Dermatology, University of California, San Francisco, San Francisco, 94143, CA, USA

---

Advances in pharmacogenetic research have linked numerous variants to drug responses and the USFDA have required pharmacogenomic (PG) information in the labeling of over 200 medications. However, only a small number of drug labels provide clear instruction for clinical implementation. As part of the Taiwan Precision Medicine Initiative (TPMI), a pilot project to generate genetic profiles for 1 million people in the country and incorporate genetic information in clinical practice, we have translated relevant pharmacogenetic results into guidelines for medication usage and dosing for real time implementation in the clinic.

A custom-designed Taiwan Precision Medicine (TPM) SNP array based on the Applied Biosystems Axiom platform that includes probes for 3,865 PG markers was validated on >50,000 Taiwan Biobank samples and the population-specific allele frequencies of the PG markers established. The accuracy of the actionable PG panel was extremely high, with sensitivity and specificity of individual tests >99%. Our preliminary data show that more than 80% of participants have risk alleles affecting drug response. For 16 medications with clear clinical guidelines, 56% of participants harbor actionable PG variants requiring change in dosage or switching to a different medication.

The PG genotyping results become part of the patients' electronic health record, and, when a medication affected by the patient's genotype is prescribed, an alert will appear on the physician's drug ordering screen based on the guidelines from the Clinical Pharmacogenetics Implementation Consortium (CPIC) and USFDA for that medication. Implementation of pre-emptive genotyping for pharmacogenomics in a large population will improve drug safety and efficacy. Early experience based on the use of pharmacogenetics in 70,000 patients in the TPMI will be presented.

## PgmNr 498: Tissue-specific gene expression, gene expression regulation, and genetic association phenotypes improves prediction of clinical trial drug side effects.

### Authors:

A. Duffy<sup>1,2</sup>; M. Verbanck<sup>1,2</sup>; A. Dobbyn<sup>1,2,3</sup>; H.H. Won<sup>4</sup>; J. Rein<sup>5</sup>; G. Nadkarni<sup>2,5</sup>; G. Rocheleau<sup>1,2</sup>; R. Do<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 2) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 3) Pamela Sklar Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 4) Samsung Advanced Institute for Health Sciences and Technology (SAIHST), Sungkyunkwan University, Seoul, South Korea; 5) Division of Nephrology, Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

---

Adverse side effects can account for the high failure rates associated with drug therapeutic development in clinical trials. Recent studies have shown that side effects of drugs can be predicted using genetic association loci identified from genome-wide association studies. However, these studies have been restricted due to a small number of studied phenotypes, inadequate approaches to infer causal genes and/or limited ability to identify biologically relevant tissues. Here, we overcome these challenges by utilizing co-localization of genetic association summary statistics data from a phenome-wide association study (PheWAS) of 1,435 phenotypes in 366,756 individuals from UK Biobank, and gene expression data from 48 tissues from the Genotype Tissue-Expression Project (GTEx). First, we evaluated how either testing of additional phenotypes from a PheWAS or testing of multiple tissues - which allows for identification of tissue-specific expression quantitative trait loci - can increase levels of genetic support for side effects of drugs in clinical trials. We observed that the percentage of side effects supported by a genetic association phenotype increased after sampling 10 to 440 phenotypes, and sampling 1 to 48 tissues from the GTEx project. Next, we identified five genetic features that are predictive of side effects, including: 1) Mendelian loci, 2) tissue-level effects and 3) phenotype-level effects of gene expression regulation, 4) genetic constraint, and 5) tissue-specificity of gene expression. By examining these features together, we observed that genes with all five genetic features had up to three-fold increased risk of side effect, compared to genes with no genetic support. Finally, we investigated whether side effects may arise due to a drug being delivered to non-relevant tissues. Using gene expression regulation and a genetic association phenotype as a proxy for a drug therapeutic causing a side effect, we observed that gene expression regulation found in a high number of tissues also tended to have many genetic association phenotypes. This correlation was driven by instances where the tissue and genetic association phenotype matched in terms of biological significance. These results imply that exposure of drugs to multiple tissues can induce several side effects. In conclusion, we demonstrate the utility of tissue-specific gene expression data for drug side effect prediction and the need for tissue-specific delivery of drug therapeutics.

## PgmNr 499: Predicting nevirapine concentrations using pharmacogenetics in Ghanaian HIV-infected children.

### Authors:

T. Langaee<sup>1</sup>; M. Al-Shaer<sup>1</sup>; Y. Gong<sup>1</sup>; E. Lima<sup>2</sup>; S. Antwi<sup>3,4</sup>; A. Enimif<sup>3,4</sup>; A. Dompheh<sup>3</sup>; H. Yang<sup>5</sup>; W. Alghamdi<sup>1,6</sup>; L. Wiesner<sup>7</sup>; C. Peloquin<sup>1</sup>; A. Kwara<sup>8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pharmacotherapy and Translational Research, Center for Pharmacogenomics and Precision Medicine, College of Pharmacy, Univ Florida, Gainesville, Florida; 2) Ohio State University, Columbus, Ohio, United States; 3) Directorate of Child Health, Komfo Anokye Teaching Hospital, Kumasi, Ghana; 4) Department of Child Health, School of Medical Sciences, Kwame Nkrumah; 5) Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY, United States; 6) Department of Clinical Pharmacy, College of Pharmacy, King Khalid University, Abha, Saudi Arabia; 7) Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa; 8) Department of Medicine, College of Medicine and Emerging Pathogens Institute, University of Florida, Gainesville, FL, United States

**Background:** Nevirapine (NVP) is a nonnucleoside reverse transcriptase inhibitor used in the treatment of human immunodeficiency virus (HIV) in young children, and is metabolized by CYP3A4 and CYP2B6 enzymes. We investigated the association of certain SNPs with NVP drug concentrations in HIV-infected children with and without tuberculosis (TB). **Methods:** This pharmacokinetic (PK)/pharmacogenetics study was performed in Ghana between 2012 and 2017. NVP-naïve patients aged 3-35 months old or weighed < 10 kg were enrolled. Patients with opportunistic infections other than TB were excluded. All parents/guardians provided written informed consent. Patients received NVP plus zidovudine and lamivudine or abacavir and lamivudine twice daily in addition to the first-line anti-TB therapy if they had TB. PK sampling was performed after at least 4 weeks of ART. Blood samples were collected at 0, 2, 6, and 12 hours after NVP dose. Drug concentration was measured using validated assay at the University of Cape Town. Minimum concentration ( $C_{min}$ ), area under the concentration-time curve from time 0 to 12 hours ( $AUC_{0-12}$ ), and apparent clearance (CL/F) were calculated using non-compartmental analysis on Phoenix v8.0. Genotyping for *rs3745274*, *rs28399499*, *rs4803419*, *rs3211371*, *rs28399433*, *rs776746*, *rs10264272*, *rs1045642*, *rs2472677*, *rs2307424*, *rs2502815*, *rs3003596*, *rs3732356*, and *rs6785049* SNPs was performed by TaqMan® allelic discrimination method. Statistical analysis was performed using JMP Pro v14.1. **Results:** A total of 53 patients were included in this study. Median (range) age and weight were 1.6 years (0.3-3.6) and 8.3 kg (4.5-12.5). Fifty eight percent (n=31) were male and 43% (n=23) had HIV/TB co-infection. The median (range) NVP dose received was 10 mg/kg (7-14). For PK parameters, the patients had median (range)  $C_{min}$  of 3.3 mcg/mL (0.0-14.0) and  $AUC_{0-12}$  56.0 mcg.hr/mL (16.7-202.6). Table below shows the SNPs which predicted the PK parameters while adjusting for age, gender, TB, and dose. Using step-wise regression, *rs3745274* and *rs6785049* predicted  $AUC_{0-12}$ ,  $C_{min}$ , and CL/F; *rs28399499* and *rs776746* predicted  $AUC_{0-12}$  and  $C_{min}$ ; and *rs28399433* predicted  $AUC_{0-12}$  only. **Conclusions:** In addition to CYP2B6, CYP3A5, and CYP2A6 SNPs, *rs6785049* (*NR1I2*) predicted NVP PK parameters.

Gene	SNP rs #	$AUC_{0-12}$	$C_{min}$	CL/F
------	----------	--------------	-----------	------

		Beta	p-value	Beta	p-value	Beta	p-value
<i>CYP2B6</i>	<i>rs3745274</i>	-26.1	0.005	-1.64	0.020	0.45	0.023
<i>CYP2B6</i>	<i>rs28399499</i>	28.10	0.150	2.87	0.047	-0.36	0.380
<i>NR1I2</i>	<i>rs6785049</i>	90.12	0.005	6.35	0.008	-1.75	0.009

## PgmNr 500: Using electronic health record data to define adverse drug events for large scale pharmacogenomic studies.

### Authors:

J.E. Olson; J.L. St. Sauver; L. Wang; N.B. Larson; D.J. Jacobson; M.E. McGree; R. Jiang; S.J. Bielinski

[View Session](#) [Add to Schedule](#)

**Affiliation:** Health Sciences Research, Mayo Clinic, Rochester, Minnesota.

---

**Background:** Adverse drug events (ADEs) cause significant morbidity and mortality each year, but are highly underreported. Electronic health record (EHR) data hold a wealth of clinical information related to ADEs, but much of the relevant information is in the clinical narrative and is challenging to extract on a large scale. For example, when an ADE is suspected, often clinicians engage in a rule out trial (e.g. stop one drug at a time) to determine which drug may be responsible for symptoms. This process is documented in the clinical narrative over time and therefore necessitates manual abstraction which can be prohibitively time consuming. We outline a hybrid process of text extraction by natural language processing (NLP) followed by manual review of the extracted text to more efficiently assign the probability of an ADE for a large scale PGx study.

**Methods:** This study included 10,082 participants in the Right Drug, Right Dose, Right Time (RIGHT) study who received a new prescription between January 1, 2005 and December 31, 2017 for at least one of 20 drugs with clinically relevant pharmacogenomic interactions. ADEs associated with the drugs of interest were identified through review of EHR data for 1 year following long-term prescription medications (i.e. simvastatin) and 3 months for short-term drugs (i.e. codeine and tramadol). Using rule-based NLP methodologies we extracted all text that included the drug name (generic and trade) and any disorder term that appeared in the same sentence or adjacent sentences. Text for each medication was reviewed and assigned a 0 (no indication of problem), 1 (AE), or 2 (possible AE), with little or no additional medical record review.

**Results:** Of the 10,082 participants, we identified over 15,000 instances where a study participant had a first-time exposure to the 20 key drugs. On average, 7.6 clinical notes were identified by NLP for each patient that had both the drug mention and a MedDRA disorder term in the same sentence or adjacent sentences. Manual review of extracted sentences is currently ongoing. Clopidogrel review is complete: 562 patients were exposed, 3993 sentences were reviewed, and 9 patients had ADEs.

**Conclusions:** We have identified an NLP trigger methodology that combines NLP techniques and manual abstraction to increase the efficiency by which ADEs are identified in a large scale study of pharmacogenomics.

## **PgmNr 501: Identification and functional classification of novel variants in important pharmacogenomic genes using whole-genome sequencing data.**

**Authors:**

S. Scherer; X. Qin; D. Muzny; R. Gibbs

[View Session](#) [Add to Schedule](#)

**Affiliation:** HGSC/Mol & Human Gen, Baylor Col Med, Houston, Texas.

---

Next-generation sequencing provides a more thorough assay of variants for pharmacogenetic testing in both research and clinical settings. Pharmacogenetic alleles can be assigned using available software tools without the limitation of the pre-selected sites used in array assays. Next-generation sequencing can also reveal novel variants that have not been previously identified or have not yet been included in known pharmacogenetic alleles (star alleles). In this study, we examined the novel variants in 17 important pharmacogenomic genes using 70 Illumina whole-genome sequencing samples, classified the possible functions of these novel variants, and discuss ways to report novel variants in pharmacogenetic testing. In addition, since many pharmacogenetic alleles are defined by more than one variant and the relationship among variants identified from short read sequencing is mostly unknown, we compared the pharmacogenetic diplotype calls using phased variants from trios data to allele calls without phasing to assess the accuracy of allele calls with and without phasing.

## PgmNr 502: A landscape of risk predictions and drug metabolism in Penn Medicine patient population.

### Authors:

S. Verma <sup>1</sup>; A. Verma <sup>1</sup>; M. Risman <sup>1</sup>; A. Lucas <sup>1</sup>; S. Damrauer <sup>1</sup>; D. Rader <sup>1</sup>; K. Sangkuhl <sup>2</sup>; M. Whirl-Carrillo <sup>2</sup>; T. Klein <sup>2</sup>; S. Tuteja <sup>3</sup>; M. Ritchie <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of Pennsylvania, Philadelphia, Pennsylvania.; 2) Stanford University, Stanford, California; 3) Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

---

Pharmacogenomics aims at identifying and applying effective and safe ways to utilize a patient's genetic data in guiding drug selection and dosing. The Clinical Pharmacogenetics Implementation Consortium (CPIC) provides guidelines with strong evidence for genes affecting more than 50 therapeutic areas. Penn Medicine BioBank (PMBB) consists of a diverse group of ~20,000 individuals whose electronic health record (EHR) data is linked to genome-wide genotype data. A retrospective data mining of the Penn Medicine EHR, which includes ~3.6M patients showed that ~30,000 patients have been prescribed more than two CPIC Level A drugs every year from 2011- 2016. Our analyses give us a snapshot view into the distribution of important CPIC drugs in the Penn Medicine EHR. We evaluated 11 CPIC genes using PharmCAT, which can annotate genetic variants in a VCF file to identify samples with functional PGx phenotypes. Genetic analyses of 19,106 samples from PMBB, stratified by ancestry, reflect the differences in PGx phenotypes by ancestry. Among the genotyped individuals, approximately 98% of samples have at least one non-reference CPIC level A variant allele. PharmCAT annotation also identified 63% of samples are carriers of one or more PGx actionable variants. PGx actionable variants refer to genetic variations that leads to recommendations for drug dosage or alternative therapies to minimize adverse drug reactions. Linking genetic data back to EHR referenced that less than 5% of participants are prescribed the affected drugs. We identified 76.6% EA and 19.8% AA carry alleles associated with warfarin sensitivity and 390 samples that contained polymorphisms for lack of clopidogrel efficacy leading to conditions such as acute coronary syndrome, coronary artery disease, and myocardial infarction. Lastly, polygenic risk scores (PRS) calculated in PMBB population to identify the risk of coronary artery disease (CAD) also reflected 10% (1916) samples were more likely to be prescribed statins. Among these 1916, 1137 samples are already prescribed statins supporting the prediction suggested by PRS. PRS<sub>CAD</sub> was also highly associated with prescription of atorvastatin (p-value=5.2e-06). This analysis provides an overview of the distribution of carriers of PGx associated genetic variants and also identifies a target population for considering pre-emptive genotyping for PGx implementation. PRS may be useful to guide preemptive PGx interventions in the setting of preventive care.



## PgmNr 503: Polygenic risk distribution across common complex diseases.

### Authors:

N. Sharapova <sup>1</sup>; J.M. Maxwell <sup>1</sup>; K. Glanville <sup>1</sup>; S.P. Hagenaars <sup>1,2</sup>; J.R.I. Coleman <sup>1,2</sup>; G. Breen <sup>1,2</sup>; P.F. O'Reilly <sup>3</sup>; C.M. Lewis <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Social, Genetic and Developmental Psychiatry Centre, King's College London; 2) NIHR Maudsley Biomedical Research Centre, London; 3) Icahn School of Medicine at Mount Sinai, New York City

---

Polygenic risk scores (PRS) aggregate risk for complex traits/diseases from many common variants across the genome into a single number indicative of an individual's risk of developing the phenotype in question. For any particular phenotype, there is a distribution of PRS values with those in the top percentiles at increased risk relative to those in the bottom. This information could inform screening programs, improving the accuracy of individual stratification by risk and peak risk period. While most individuals will be in the middle of the distribution for most diseases, they will likely be at extremely high or low risk for at least one disease. We used the UK Biobank (UKB) to investigate individuals in the tails of PRS distribution across common complex diseases that are a serious burden on affected individuals and the healthcare system: coronary artery disease, stroke, type 2 diabetes, rheumatoid arthritis and Alzheimer's disease. External genome-wide association studies (GWAS) were used to predict into corresponding phenotypes in the UKB. The optimal P-value threshold explaining maximum phenotypic variance was identified for each. We then built PRS scores in unrelated UKB participants of European ancestry (n=385 780) at these optimal thresholds and adjusted for batch, assessment centre and 6 principal components. The adjusted PRS were split into quantiles and different PRS distribution patterns were investigated to identify individuals at increased risk for at least one disease. Finally, we explored comorbidities between the diseases studied and stratified risk by family history. We show that 40.8%, 22.6% and 4.9% of the population studied were in the tail 10%, 5% and 1% for at least one disease. Furthermore, 9.2%, 2.4% and 0.1% of the population were in the tail 10%, 5% and 1% PRS for at least two disorders, respectively. This indicates a slight increase in the enrichment of multiple high PRS compared to chance, reflecting a moderate positive correlation between the five PRS. Individuals in the top tail of PRS distribution were more likely to have family history of the disease, excepting stroke where no strong relationship was observed. While the proportion of individuals at a level of clinically elevated risk is low for each disease alone, risk distribution across these common complex diseases could be used to inform the optimization of prevention, screening and treatment, with implications on a population scale.

## PgmNr 504: Integrating longitudinal omics for detection of adverse events in deep space missions.

### Authors:

G. Mias<sup>1,2</sup>; S. Domanskyi<sup>3</sup>; S. Xue<sup>1,3</sup>; C. Piermarocchi<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan.; 2) Institute for Quantitative Health Science and Engineering, Michigan State University, East Lansing MI; 3) Physics Department, Michigan State University, East Lansing, Michigan.

Previous large scale investigations, have demonstrated the feasibility of multi-omics data generation, and their utility towards precision medicine. We will present work on the necessary next steps towards monitoring individual astronaut health in deep space missions: the development of novel algorithms that can monitoring wellness departures from individualized personnel baselines.

Our methods characterize longitudinal multi-omics signals (transcriptomes, proteomes, and physiological measures), using novel statistical algorithms based on topological properties of time series. Our methodology first analyzes individual signals based on their frequencies, accounting for underlying noise, uneven sampling rates, missing data (without imputation), and sampling over multiple time-scales for different data modalities. Our software package pyiomica uses spectral analysis to transform time signals to periodograms and classify them based on their autocorrelations. This allows the identification of temporal trends, including periodic events, constant changes, and sudden changes compared to baseline, that can correspond to adverse health events during flight. Clusters of multiple omics components are identified within each temporal class, and these collective signals are then transformed into visibility graphs, that are networks representing intensities at different timepoints of multiple omics. The network characteristics are then used to identify modules and abnormal events over time, relating to changes in physiological state of the personnel, as assessed by biological evaluation for enrichment of particular pathways. We will demonstrate the application of the algorithms using saliva data from 100 timepoints (transcriptomics, proteomics and extracellular vesicle content), as an evaluation of the utility of saliva for non-invasive monitoring of generalized health - particularly convenient for in-flight sampling of astronauts.

Personalized dynamic omics have high utility for monitoring astronauts, and our approach is an extensible prototype that can incorporate new dynamic measurements as measured during spaceflight with different monitors.

Research reported is supported by Project Number T0412 grant by the Translational Research Institute for Space Health, funded under NASA Cooperative Agreement NNX16A069A. The content is solely the responsibility of the authors and does not necessarily represent the official views of the supporting funding agencies.

# PgmNr 505: Genetic evidence for prediction of severe on-target adverse effects of common drugs: A proof of principle.

## Authors:

D.J. Wright; M. Zwierzyzna; N. Lewis; J. Dunbar; J.D. Holbrook; D.I. Swerdlow

[View Session](#) [Add to Schedule](#)

**Affiliation:** BenevolentAI, London, United Kingdom

---

## Background

Drug targets supported by human genetic evidence for efficacy are more likely to be approved. Robust prediction of target-mediated adverse effects (AEs) early in drug development would be a valuable de-risking strategy, but the utility of genetics in doing so has not been evaluated. Using a set of curated gene-AE pairs, we assess whether common genetic evidence can retrospectively predict on-target AEs.

## Methods

We constructed a benchmark of target-encoding genes and serious AEs for which European GWAS summary statistics for the AE trait/a clinical correlate were publicly available. For each trait, we identified genetic signals ( $p \leq 5 \times 10^{-6}$ ) within 500kb of each target start site, with a subanalysis requiring  $p \leq 5 \times 10^{-8}$ . To assess whether the signal at each locus was driven by *cis* gene expression, we ran pairwise (approximate Bayes factor) colocalisation analyses against tissue-specific eQTL summary statistics for each target-tissue-trait combination (GTExV7) using COLOC. Colocalisation was declared where the posterior probability (PP) of colocalised signals (H4) was the highest of all posteriors for the five hypotheses evaluated by COLOC, with sub-analyses requiring a  $PP \geq 0.8$ . We declared genetic evidence linking target to trait when colocalisation was observed in  $\geq 1$  tissue.

## Results

Metrics of precision and sensitivity were used to assess error (enrichment analysis of genetic evidence amongst AE-annotated genes vs. controls was not feasible due to low cell counts). Of the 15 trait-geneset groups analysed,  $\geq 1$  annotated gene was correctly identified in two cardiac rhythm traits: atrial fibrillation (AF)/other cardiac arrhythmia (*NR3C1*, *ATP1B2*, *KCNN3*) and heart rate (*CALCR1*). Main analyses ( $p \leq 5 \times 10^{-6}$ , highest ranked PP) for AF showed 50% precision and 33% sensitivity, and HR 14.3% precision and 100% sensitivity. The most rigorous subanalyses ( $p \leq 5 \times 10^{-8}$ ,  $PP \geq 0.8$ ) optimised both metrics in both AF (100% precision, 40% sensitivity) and HR (50% precision, 100% sensitivity).

## Conclusion

These analyses demonstrate that drug targets known to be associated with severe AEs can be retrospectively identified using colocalisation between trait- and expression-associated variants.

Whilst results varied highly between traits, and null results cannot be interpreted as evidence in support of the absence of an association, we highlight the potential of this method to prospectively flag concerning targets.

## **PgmNr 506: Successful application of preimplantation genetic diagnosis for oculocutaneous albinism type 1 merging with preimplantation genetic testing for aneuploidy by a combination of quantitative polymerase chain reaction approaches.**

### **Authors:**

S.P. Chang <sup>1</sup>; A. Yan <sup>1</sup>; G.C. Ma <sup>1</sup>; F.P. Tsai <sup>2</sup>; M. Chen <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Genomic Medicine, Changhua Christian Hospital, Changhua City, Taiwan, Taiwan; 2) Poyuan Women's Clinic, IVF Center, Changhua, Taiwan; 3) Department of Obstetrics and Gynecology and Department of Medical Genetics, College of Medicine and Hospital, National Taiwan University, Taipei, Taiwan

---

Autosomal recessive oculocutaneous albinism type 1 (MIM# 203100; OCA1) is one of the thousands of monogenic diseases characterized by absence of pigment in hair, skin, and eyes. Mutations of the tyrosinase (*TYR*, MIM\* 606933) gene are clinically associated with the disorder, and so far, more than 300 mutations have been described. Preimplantation genetic diagnosis (PGT-M) has been widely utilized for the affected families who wish to prevent the transmission of genetic disorders. Preimplantation genetic testing for aneuploidy (PGT-A) is also employed to improve live birth rates in IVF (In vitro fertilization) cycles for couples with advanced maternal age, repeated implantation failure, and recurrent miscarriage. More and more at-risk couples have the need to combine PGT-M and PGT-A technologies during IVF treatment. In the present study, a Taiwanese couple with a maternal age of 38 years was enrolled. Their first child with OCA1, suffering from eye problems, caused by heterozygous *TYR* mutants of paternal c.1366+3A>T and maternal c.895C>A (p.R299S). We developed and evaluated an in-house duplex-nested ARMS-qPCR technique for PGT-M, and compared it to direct sequencing. The optimized protocol was then applied clinically to detect the disease-causing mutations and to screen simultaneously for aneuploidy by a qPCR-based PGT-A. After ovarian stimulation, a total of 14 embryos with good morphology were selected and biopsied from blastocysts on day 5 to detect PGT-A and PGT-M simultaneously. The PGT-A results showed that two embryos were aneuploid and 12 were euploid, and those of PGT-M showed that three wild-type embryos without the familial mutants, three affected embryos with two heterozygous mutants, four maternal-carrier embryos, and four paternal-carrier embryos. Finally, two euploidy embryos were selected (one wild-type and one maternal carrier) for transfer at Day 6, achieving a successful singleton-pregnancy. After 38 weeks of uneventful gestation, a healthy boy was born weighing 3088g. Prenatal and postnatal genotyping confirmed the baby to be an unaffected maternal carrier. We here report a successful application for the strategy combined PGT-M and PGT-A by a combination of quantitative polymerase chain reaction approaches, resulting in the healthy boy birth. The results show that such a combination of technologies is a fast and valid screening tool and can be provided as an optimized solution for high-risk families.

## PgmNr 507: Targeted sequencing and RNA assay reveal a noncanonical *JAG1* splicing variant causing Alagille syndrome.

### Authors:

Y.Y. Chen <sup>1,2</sup>; J.Y. Zhang <sup>1,2</sup>; X.L. Liu <sup>1,2</sup>; H.J. Fei <sup>1,2</sup>; S.C. Chen <sup>1,2</sup>; C.M. Xu <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) International Peace Maternal and Child Health Hospital, Shanghai Jiao Tong University, Shanghai; 2) Key Laboratory of Embryo-Fetal Original Adult Disease, Shanghai

---

**BACKGROUND:** Alagille syndrome (ALGS), as known as congenital arteriohepatic dysplasia, is a rare autosomal dominant multi-systemic disorder. Mutations in *JAG1* or more rarely *NOTCH2* have been revealed as the cause of ALGS. In this study, a 5-year old girl with typical ALGS' feature and her pregnant mother came to our reproductive genetics clinic for counseling.

**AIM:** We aimed to clarify the genetic diagnosis and provide prenatal genetic diagnosis for the pregnant.

**METHODS:** Next generation sequencing (NGS)-based multigene panel was used to identify pathogenic variant of the proband. The candidate variant was verified by using Sanger sequencing. RNA assay was performed to clarify splicing effect of the candidate variant. Amniocentesis, karyotyping, and Sanger sequencing were performed for prenatal testing.

**RESULTS:** We identified a novel *de novo* noncanonical *JAG1* splicing variant (c.2917-8C>A) in the proband. Peripheral blood RNA assay indicated that the mutant transcript might escape nonsense-mediated mRNA decay (NMD) and encode a C-terminal truncated protein. Information of the variant led to success prenatal diagnosis of the fetus.

**CONCLUSION:** In conclusion, we clarified the genetic diagnosis of an ALGS patient and ensured utility of prenatal genetic testing.

## PgmNr 508: Mutation of *RONIN* (*THAP11*) results in a combined vitamin B<sub>12</sub> and ribosome biogenesis deficiency syndrome.

### Authors:

A. Achilleos<sup>1</sup>; X. Tong<sup>1</sup>; M.C. Hill<sup>1</sup>; M. Zhang<sup>1</sup>; L. Wong<sup>1</sup>; A. Chaudhury<sup>1</sup>; J.R. Neilson<sup>1</sup>; J. Sun<sup>1</sup>; R.S. Ray<sup>1</sup>; D. Watkins<sup>2</sup>; D.S. Rosenblatt<sup>2</sup>; J.F. Martin<sup>1,3</sup>; R.A. Poché<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX.; 2) Department of Human Genetics, McGill University, Montreal, CA; 3) Texas Heart Institute, Houston, TX

Combined methylmalonic acidemia and homocystinuria (*cb/C* type), an inherited disorder of vitamin B<sub>12</sub> metabolism, is a rare metabolic and multi-systemic disease caused by mutations in *MMACHC*. Patients with *cb/C* can have severe neurodevelopmental defects including microcephaly, hydrocephaly, and seizures as well as renal, cardiac and hematological defects. Recently, an X-linked variant of *cb/C* was discovered and termed *cb/X*. Rather than being due to mutations in *MMACHC*, *cb/X* was shown to result from hemizygous mutations in the transcription cofactor *HCFC1*, which is known to be an obligatory partner for the transcription factor *RONIN* (*THAP11*). Intriguingly, a single patient exhibiting *cb/X*-like pathology has been found to carry homozygous mutations in *RONIN*. Patients with either *HCFC1* or *RONIN* mutations were shown to have a dramatic reduction in *MMACHC* transcription. We have previously shown that the *Hcfc1*/*Ronin* transcriptional complex directly regulates mouse *Mmachc* expression. These findings suggest that *cb/X* and the new *RONIN* (*THAP11*) disorder comprise a novel family of rare and severe *cb/C*-like disorders that are transcriptional in nature. To better understand the cellular and molecular mechanisms underlying the pathophysiology of these devastating neurodevelopmental diseases, we have generated a mouse model that carries the human mutation in *Ronin* (*Ronin*<sup>F80L</sup>). Here we report the generation of the first mouse model, *Ronin*<sup>F80L</sup> along with its phenotypic and molecular characterization. *Ronin*<sup>F80L/F80L</sup> homozygous mice die soon after birth and exhibit severe brain developmental defects that recapitulate those observed in the human *cb/X* and *RONIN*-deficient patients. Moreover, consistent with an inherited vitamin B<sub>12</sub> disorder, cells from *Ronin*<sup>F80L/F80L</sup> homozygous embryos exhibit defects in cobalamin metabolism. Surprisingly however, RNA-seq and ChIP-seq analyses also revealed a role for *Ronin* in directly regulating ribosome biogenesis. Furthermore, we were able to show that there is also a functional deficit in ribosome biogenesis and hence protein translation. This identifies for the first time a role for *RONIN* in ribosome biogenesis. Together, the phenotypic and molecular data confirm that the *Ronin*<sup>F80L/F80L</sup> mouse model will serve as a powerful tool to further uncover the pathophysiology of this complex family of diseases.

## PgmNr 509: Copy number variations of the PKA catalytic subunits *PRKACA* and *PRKACB* are present in isolated micronodular adrenocortical disease (iMAD).

### Authors:

N. Settas; L. Drougat; A. Berthon; C. Kamilaris; P. Salpea; F.R. Faucz; C.A. Stratakis

[View Session](#) [Add to Schedule](#)

**Affiliation:** Section on Endocrinology & Genetics (SEGEN), NICHD, NIH, Bethesda, Maryland

---

Genetic defects that cause abnormal c-AMP dependent protein kinase A (PKA) signaling constitute the leading cause of adrenal tumors and Cushing syndrome (CS). Inactivating *PRKAR1A* mutations are the main genetic cause of PKA pathway activation identified so far in bilateral adrenocortical hyperplasia (BAH). However, activating mutations and amplification of the PKA catalytic subunit gene *PRKACA* were recently described in cortisol-producing adrenal tumors and BAH, respectively. Herein, we studied copy number variations (CNVs) of the PKA catalytic subunits *PRKACA* and *PRKACB* in 77 patients with isolated micronodular adrenocortical disease (iMAD) that are negative for *PRKAR1A* mutations. We performed TaqMan copy number assay through ddPCR for *PRKACA* and *PRKACB* genes. We screened 69 germline samples and unilateral or bilateral adrenal tissues from 34 patients. The *PRKACA* gene was duplicated in 10 out of 69 germline samples (14.4%). We detected two heterozygous deletions of *PRKACB* gene at the germline level (2.8%); one of the two patients had also duplication of the *PRKACA* gene. At the somatic level we detected duplication or amplification (more than 3 copies) of the *PRKACA* gene in 16 of the studied tissues (47%) and one had an homozygous deletion (2.9%). In five of those 16 cases (14.7%) we detected the *PRKACA* CNVs in both adrenal glands, while absent in the germline, indicating potential mosaicism. Regarding the *PRKACB* CNVs, we found that two patients had a heterozygous deletion of the *PRKACB* gene in the germline (2.8%) one of whom had also the *PRKACA* duplicated; unfortunately, adrenal tissue was not available from this patient. Moreover, at the somatic level we found that 9 out of 34 adrenal tissues (26.4%) had deletion of the *PRKACB* gene, while 7 of them (20.5%) had also duplication or amplification of the *PRKACA* gene. These data suggest that *PRKACA* and *PRKACB* copy number variations may happen concurrently in the early stages of adrenocortical tumor development in BAH. Although the exact way these defects participate in tumor growth remains unclear, we can speculate that almost certainly they are involved in the increased PKA activation that is the hallmark of BAH and most benign cortisol-producing adrenal tumors.



## PgmNr 510: Analysis of ultrasonic manifestation of fetus with 17q12 microdeletion.

### Authors:

L. Zhang; Y. Li; E. Yuan; H. Li; P. Yang; G. Li; Y. Xu; H. Zhang; Z. Zhang; S. Cui

[View Session](#) [Add to Schedule](#)

**Affiliation:** The 3rd affiliated hosp. of Zhengzhou Univ., Zhengzhou, China

### Objectives

In our study, we firstly analyzed the ultrasonic manifestations of 25 fetus diagnosed as 17q12 microdeletion by prenatal chromosome microarray analysis (CMA).

### Methods

Cells from fetuses in amniotic fluid were collected by centrifugal method, and cell DNA was extracted for CMA. The Affymatrix CytoScan 750K chip (Thermo, USA) was used for CMA detection after confirming no maternal blood contamination by short tandem repeat (STR) locus comparison.

### Results

Among the 3,901 pregnant women, a total of 25 fetuses were detected with 17q12 microdeletion, accounting for 0.6% (25/3,901). There were 340 fetuses (8.7%, 340/3,901) showed kidney abnormality by prenatal ultrasonography, and included all the 25 fetuses with 17q12 microdeletion, accounting for 7.4% (25/340). Among these 340 fetuses with kidney abnormalities, 48 fetuses were classified as pathogenic (14.1%, 48/340), among which 17q12 microdeletion accounted for 52.1% (25/48) and chromosome aneuploidy disease accounted for 27.1% (13/48), the other 20.8% (10/48) were pathogenic chromosome microdeletion and microduplication.

88.0% (22/25) fetuses showed hyperechogenic kidney by prenatal ultrasonography in 17q12 microdeletion fetuses. Among the 22 fetuses with hyperechogenic kidney, 24.0% (6/22) had no other abnormalities except hyperechogenic kidney. Collecting system separation (28.0%, 7/25), polycystic dysplastic kidney (20.0%, 5/25), renal cysts (8.0%, 2/25) and hydronephrosis (8.0%, 2/25) were included in other fetuses. 28.0% (7/25) fetuses with only kidney abnormalities by prenatal ultrasonography, 16.0% (4/25) were combined with increased amniotic fluid, 24.0% (6/25) were combined with other system abnormalities, and the most common abnormalities was nervous system. In addition, there were hyperechogenic bowel, single umbilical artery and growth retardation. There were ten pregnant women did the parental CMA verification, four of which were inherited from mother and the other six fetuses were newly mutation. Follow-up study showed that five (20.0%, 5/25) fetuses had a family history of diabetes and there was one pregnant woman's father with renal cysts.

### Conclusion

In conclusion, hyperechogenic kidney was a common prenatal ultrasonic manifestation of 17q12 microdeletion syndrome. Hyperechogenic kidney was found at prenatal, 17q12 microdeletion syndrome should be considered and prenatal chromosome microarray analysis should be performed.

## PgmNr 511: Whole exome sequencing identifies candidate gene variants in patients with Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome.

### Authors:

M.A. Morton <sup>1</sup>; S. Brakta <sup>1</sup>; L.P. Chorich <sup>1</sup>; K. Berwick Tam <sup>1</sup>; H.G. Kim <sup>2</sup>; A.C. Lossie <sup>3</sup>; M. Friez <sup>4</sup>; J.A.III Phillips <sup>5</sup>; L.C. Layman <sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Section of Reproductive Endocrinology, Infertility, & Genetics, Dept Ob/Gyn, Medical College of Georgia, Augusta University, Augusta, Georgia.; 2) Neurological Disorders Research Center, Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Doha, Qatar; 3) Beautiful You MRKH Foundation, Silver Spring, MD; 4) Greenwood Genetics Center, Greenwood, SC; 5) Division of Medical Genetics and Genomic Medicine, Dept Pediatrics, Vanderbilt University, Nashville, TN; 6) Dept Neuroscience and Regenerative Medicine, Medical College of Georgia, Augusta University, Augusta, GA LCL was supported by NCID R01 HD092505

---

Mayer-Rokitansky-Küster-Hauser (MRKH) Syndrome is a phenotypically complex condition characterized by incomplete development of the uterus and vaginal canal, and can be accompanied by unilateral renal agenesis, skeletal abnormalities, hearing impairment, and cardiac anomalies. While some large copy number variants and several mutations in *HNF1B* and *WNT4* have been identified in a small number of patients, the genetic basis remains poorly understood for most patients. To better understand genetic causes of MRKH, we performed whole exome sequencing on DNA from 111 women with MRKH to identify likely pathogenic variants (frameshift, splice site, or nonsense) and missense variants of undetermined significance (VUS) with an allele frequency of <0.01. We identified and prioritized candidate variants with the following characteristics: 1) those in murine genes known to impair müllerian development; 2) those in genes with putative disease causing variants in humans, but lacking definitive evidence; and 3) those near the breakpoints of derivative chromosomes from a patient with a t(3;16)(3p22.3;16p13.3) translocation. Among 14 human orthologs from mouse models (*EMX2*, *HOXA9*, *HOXA10*, *HOXA11*, *HOXA13*, *LHX1*, *PAX2*, *PBX2*, *RARA*, *TP63*, *WNT4*, *WNT5A*, *WNT7A*, and *WNT9B*), we found 3 heterozygous likely pathogenic variants in *EMX2*, but in none of the other genes. Among genes previously studied, but not confirmed in humans (*AMH*, *AMHR2*, *CFTR*, *CTNNB1*, *GALT*, *HOXA7*, *LAMC1*, *LRP10*, *MMP14*, *PBX1*, *RARG*, *RXRA*, *SHOX*, *TBX6*, *WT1*, and *ZNHIT3*), we identified heterozygous likely pathogenic variants in *ZNHIT3* (n = 4) and *CFTR* (n = 2). For genes located near the breakpoints on derivative chromosomes 3p22.3 (*CCR4*, *CMTM6*, *CMTM7*, *CMTM8*, *CNOT10*, *CRTAP*, *DYNC1I1*, *GLB1*, *TMPPE*, and *TRIM71*) and 16p13.3 (*IL32*, *MEFV*, *OR1F1*, *TIGD7*, *ZNF200*, *ZNF205*, *ZNF213*, *ZNF263*, *ZNF75A*, and *ZSCAN10*), *TIGD7* (n=1) and *MEFV* (n=1) had heterozygous likely pathogenic variants. We also identified heterozygous VUS with a CADD score >20 in most of the other genes, including 13 variants in 6 murine orthologs, 43 variants in 15 putative human genes, and 23 variants in 11 genes at the breakpoints of the t(3;16) translocation. Sanger sequencing is ongoing for confirmation, and *in vitro* confirmation and segregation analyses will be performed for selected variants. Our preliminary data suggests that low frequency deleterious alleles in *EMX2*, *ZNHIT3*, *TIGD7*, and perhaps *CFTR*, could play a role in the pathogenesis of MRKH.

## PgmNr 512: Utilizing CRISPR/Cas9 to study SOX9 haploinsufficiency in human iPSC-derived chondrocytes.

### Authors:

T.Y. Ha; K.K. Miu; W.Y. Chan

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Biomedical Sciences, Faculty of Medicine, the Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR

---

SOX9 is an important gene in development which actively involves in the differentiation of 3 germ layers and it is the master regulatory gene in chondrocyte development. Moreover, haploinsufficiency (HI) in SOX9 will result in severe developmental disorder called Campomelic Dysplasia (CD). Patients with CD suffer from bent long bones, club feet, small thoracic cavity etc. CD patients usually die shortly after birth due to respiratory failure and only very few of them could survive into adolescent years. Our study aims at elucidating the effect of different pathways which are important for SOX9 HI on chondrocytes.

To generate our desired SOX9 HI model, we utilized CRISPR/Cas9 to make a single nucleotide polymorphism (SNP) mutation at the end of SOX9 exon 2 splice site in human induced pluripotent stem cells (hiPSCs). Single clones were screened by Sanger sequencing of the PCR fragments. The SOX9 HI hiPSCs were differentiated to chondrocytes according to a well-established 14-day protocol. To validate our desired SOX9 HI mutant, RT-qPCR, immunofluorescence and Safranin O staining were performed on day 14. Our results from these experiment consistently showed that the SOX9 HI chondrocytes have decreased expression of SOX9 in terms of mRNA and protein levels.

Transcriptome analysis were then performed. By comparing wild type chondrocytes and SOX9 HI chondrocytes, we identified around 700 differentially expressed genes for each phenotype and from gene ontology analysis, binding of around 1000 genes were being affected in SOX9 HI chondrocytes. From the KEGG analysis results, notch signaling, p53 and ribosomal pathways have the highest rich factor scores.

To conclude, we had successfully generated SOX9 HI hiPSCs and differentiated them to chondrocytes. From the transcriptome analysis results, several pathways were being affected in SOX9 HI chondrocytes. Our further research would be to inhibit these pathways one by one and try to rescue the SOX9 HI phenotype with different approaches.

## PgmNr 513: The role of TWIST1 phosphorylation and miR10 in regulating neural crest cell fate and migration.

### Authors:

W.D. Fakhouri <sup>1,2</sup>; J. Wildgrube Bertol <sup>1</sup>; V.K. Xie <sup>1</sup>; K.M. Hubka <sup>1,3</sup>; M.C. Farach-Carson <sup>1,3</sup>; R. Keuls <sup>4</sup>; R.J. Parchem <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Craniofacial Research, Department of Diagnostic and Biomedical Sciences, School of Dentistry, University of Texas Health Science Center at Houston, TX, USA; 2) Department of Pediatrics, McGovern Medical School, University of Texas Health Science Center at Houston, TX, USA;; 3) Department of Bioengineering, Rice University, Houston, TX, USA; 4) Department of Neuroscience, Stem Cells and Regenerative Medicine Center, Baylor College of Medicine, Houston, TX, USA

---

A wide variety of common birth defects and pediatric cancers arise from aberrations in the development of neural crest (NC) cells originating from the neural plate borders. The NC cells switch their epithelial to mesenchymal phenotype with the acquisition of migratory and plasticity properties towards the pharyngeal arches during facial and peripheral nervous system formation. This neurulation system provides an excellent *in vivo* model to uncover the mechanism(s) governing how transcription factors and microRNAs regulate NC cell specification, epithelial to mesenchymal transition (EMT), delamination, and migratory processes. The transcription factor TWIST1 is a master regulator of EMT and plays a critical role during tumor invasiveness in cancer metastasis. TWIST1 phosphoregulation in cancer cell line models showed that several critical phospho-residues regulate both protein activity and transcription of target genes including miR-10 family members. Despite its importance in NC cells, the function of Twist1 phosphoregulation and expression of the miR-10 family has not been investigated *in vivo*. Previous developmental studies suggested that conditional knockout (cko) of *Twist1* in cranial neural crest (CNC) cells increased cell death of NC cells localized to pharyngeal arches, while complete loss of *Twist1* led to failure of neural tube closure and early embryonic lethality. However, our recent quantitative data using dual fluorescent cell tracing markers and neural tube explants showed that *Twist1*cko in CNC cells significantly reduced cell delamination and disrupted the EMT process. The detached *Twist1*null CNC cells retained their epithelial signatures and migrated over a significantly less distance, providing an explanation for observed increased cell death and craniofacial abnormalities. Notably, TWIST1 is highly phosphorylated in CNC cells, while miR10a, miR10b and miR99b are temporally regulated during neurulation. To investigate the importance of TWIST1 phosphorylation in CNC cell fate, *Twist1*phospho-incompetent mice for multiple serine residues were recently generated. Our initial results showed that Tw1-S68A phospho-mutant mice have craniofacial defects and hemorrhage in the forebrain. To further investigate the importance of TWIST1 phosphorylation, EMT and migration of CNC cells are being analyzed in S18/20A and S68A mutant mice. These studies will lead to a better mechanistic understanding of the role of *Twist1* and miR10 in NC cell fate.

## PgmNr 514: *HMGA2* disruption in monozygotic twins with Silver-Russell syndrome.

### Authors:

A.C.G. Gonçalves <sup>1</sup>; H. Dornelles-Wawruk <sup>1</sup>; M.M. Mehrjouy <sup>2</sup>; M. Bak <sup>2</sup>; M. Santos <sup>3</sup>; N. Tommerup <sup>2</sup>; J.F. Mazzeu <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Programa de Pós-graduação em Ciências da Saúde, Universidade de Brasília, Brasília, Brazil; 2) Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark; 3) Hospital Universitário de Brasília, Brasília, Brazil; 4) Faculdade de Medicina, Universidade de Brasília, Brasília, Brazil

---

Silver-Russell Syndrome (SRS) is a disorder characterized by prenatal and postnatal growth retardation, macrocephaly, a prominent forehead, body asymmetry and facial dysmorphisms. Epimutations at chromosome 11 and maternal uniparental disomy of chromosome 7 together explain about 60% of SRS cases. However, patients clinically diagnosed with SRS sometimes present negative results in routine genetic tests for the syndrome. This reinforces the idea that other genes may be involved in the etiology of the syndrome. We studied monozygotic twin sisters clinically diagnosed with Silver-Russell syndrome carrying a balanced translocation between chromosome 3 and chromosome 12 (46,XX,t(3;12)(q29;q14)). Chromosome microarray analysis did not identify any imbalances. Mate-pair sequencing was then performed to identify the translocation breakpoints. The breakpoint on chromosome 3 interrupts the *NAALADL2* gene. This gene has an important role in the central nervous system acting in the release of glutamate in the synaptic cleft. The breakpoint on chromosome 12 disrupts *HMGA2*. *HMGA2* gene encodes an architectural transcription factor highly expressed in embryonic development involved in cell growth and proliferation. *HMGA2* regulates expression of *IGF2*, a causative gene for SRS. Loss-of function and large deletions in *HMGA2* have been reported in a few patients with a Silver-Russell phenotype. Therefore, this study provides further evidence of the role of *HMGA2* in the etiology of Silver-Russell syndrome.

## PgmNr 515: Phenotypic and molecular characterization of orofacial development in *Mecom* loss-of-function mutants.

### Authors:

B.G. Doolittle <sup>1</sup>; J.C. Jones <sup>1</sup>; A.S. Perkins <sup>2</sup>; D. Anand <sup>3</sup>; S.A. Lachke <sup>3</sup>; I. Saadi <sup>4</sup>; B.C. Bjork <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Biochemistry and Master of Biomedical Sciences Program, Northwestern University, Downers Grove, IL; 2) Dept Pathology and Laboratory Medicine, Univ of Rochester Med Center, Rochester, NY; 3) Dept Biological Sciences, Univ of Delaware, Newark, DE; 4) Dept Anatomy and Cell Biology, Kansas Univ Med Center, Kansas City, KS

---

Human orofacial clefting is one of the most common human birth defects with a complex multifactorial etiology. Isolated cleft palate (CP) can arise due to abnormal growth, migration or fusion of the palate shelves caused by perturbations intrinsic to the palate shelves or as a secondary consequence of extrinsic factors. Pierre Robin sequence (PRS) is a form of CP in which individuals exhibit micrognathia and glossoptosis that leads to failed palate shelf elevation and fusion. *Prdm16* null mutant mice exhibit PRS-like cleft palate. A role for the *Prdm16* paralog, *Mecom*, in orofacial development has not been reported; however, we observe CP in the *Mecom* loss-of-function mutant, *Evi1*<sup>ko3</sup>. In contrast to *Prdm16* mutants in which palate shelf elevation is obstructed, *Evi1*<sup>ko3</sup> mutant palate shelves elevate but fail to meet and fuse at the midline. The requirement for *Mecom* appears to be intrinsic to the palate. Both genes encode zinc-finger transcription factors that regulate gene expression through both DNA-dependent and DNA-independent mechanisms. Both *Mecom* and *Prdm16* are robustly expressed in the developing craniofacial structures with partially overlapping patterns throughout the embryo. This study focuses on a targeted *Mecom* deletion allele, *Evi1*<sup>ko3</sup>, in which exon 3 of *Evi1* is removed to prevent expression of two transcripts expressed from the *Mecom* locus (*Mds1-Evi1* and *Evi1*) but not *Mds1*. We outcrossed the allele 10 generations onto the FVB/NJ strain and observed ~40% CP penetrance. We examined non-penetrant *Evi1*<sup>ko3</sup> mutants for potential “subclinical” phenotypes using SEM and histology. Embryonic day (E)16.5 mutants without CP exhibit palate rugae defects and abnormal palate morphology at the midline that warrants further characterization to identify potential submucous CP or other “subclinical” phenotypes. To assess the molecular consequences of palate-intrinsic loss of *Mecom*, we carried out transcriptional profiling via RNAseq analysis of total RNA from pooled embryonic day E13.5 *Evi1*<sup>ko3</sup> mutant and wild type palate shelves. Validation of differentially expressed genes is ongoing. These data address the hypothesis that loss of *Mecom* results in CP due to altered developmental and molecular pathways intrinsic to the palate shelves and may contribute to a greater understanding of normal palatogenesis and suggest additional contributors to the etiology of CP in mice and human.

## PgmNr 516: Involvement of transcription factors that control embryonic morphogenesis in the repair of DNA-DSBs after fetus radiation exposure.

### Authors:

A. Noda; K. Hamasaki; Y. Kodama; N. Nakamura

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Molecular Biosciences, Radiation Effects Research Foundation, Hiroshima, Japan

---

In the study of Hiroshima and Nagasaki atomic bomb survivors exposed in utero, it has been noted that exposure-related small head size and intellectual impairment had increased. Thus developing central nervous system seems to be radiation sensitive. In a process of searching for specific proteins that react with unrepairable, persistent DSBs induced by radiation exposure, we have detected an ATM dependent phosphorylation of transcription factor CHD7 (chromodomain containing DNA helicase 7) which is known as morphogenesis factor of brain and neurosensory organ. Antibodies raised against the phosphorylated peptide sequence, KLEDDDKSEESpSQPEAGA, which was a putative ATM kinase target in CHD7, strongly reacted with DSBs formed by radiation exposure. The phosphorylated form of CHD7 co-localized with γH2AX, 53BP1 and p-ATM, making DSB repair-foci. Co-immunoprecipitation experiments indicated the antibody cross-reactivity with 53BP1, since both proteins share the ATM kinase target, SpSQP sequence. Thus we reconfirmed the specificity of ATM-dependent CHD7 phosphorylation under the 53BP1 null background. CHD7 is known as a transcription factor that controls cell differentiation and morphogenesis during embryonic development especially with regards to formation of brain, heart and sensory organs (eyes, ears) from the neural crest via non-canonical WNT signaling. In human, heterozygous mutation of *CHD7* causes malformation (Charge Syndrome) and results in embryonic lethality in the case of null (homozygous) mutation. Mouse fetus exposure resulted in the strong accumulation of p-CHD7 proteins at the DSB-repair foci in developing neurosensory organs. Thus we conclude that there exists molecular mechanism of morphogenesis-coupled DSB repair in the early development. This mechanism seems to be universal since (1) PTIP (Pax2 Transcription factor Interacting Protein) also participate in DSB-repair foci formation, where Pax2 is a transcription factor that control kidney morphogenesis. (2) COBRA1, a BRCA1 interacting protein which controls mammary gland morphogenesis also participates in DSB-repair foci formation, and (3) MCPH1/BRIT1, whose mutation causes microcephaly, has three BRCT domains and strongly accumulates at DSB-repair foci after radiation exposure.



## PgmNr 517: Integrated transcriptome and network analysis reveals spatiotemporal dynamics in dysmorphic skull development.

### Authors:

N. Lu <sup>1,9</sup>; G. Holmes <sup>1,9</sup>; A.S. Gonzalez-Reiche <sup>1,2,9</sup>; X. Zhou <sup>1,9</sup>; J. Rivera <sup>1,8</sup>; D. Kriti <sup>1,2</sup>; R. Sebra <sup>1,2</sup>; A.A. Williams <sup>1</sup>; M.J. Donovan <sup>3</sup>; S.S. Potter <sup>4</sup>; D. Pinto <sup>1,2,5,6</sup>; B. Zhang <sup>1,2</sup>; H. van Bakel <sup>1,2</sup>; E.W. Jabs <sup>1,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Pathology, Icahn School of Medicine at Mount Sinai, New York; 4) Division of Developmental Biology, Cincinnati Children's Medical Center, Cincinnati, OH; 5) Department of Psychiatry, and Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY; 6) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Department of Cell, Developmental and Regenerative Biology and Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY; 8) Present address: Center for Personalized Cancer Therapy, University of Massachusetts, Boston, MA; 9) These authors contributed equally

---

Craniofacial dysostosis including craniosynostosis is a common birth defect that occurs in one in at least 2,000 newborns. The metopic suture, one of the major craniofacial sutures, is prematurely fused in 25% of all craniosynostosis cases. Gene mutations affecting suture growth are a significant source of calvarial pathology, but no more than 30 genes and few mouse models are known to be associated with metopic abnormalities. The metopic suture is abnormally widened in Apert and Saethre-Chotzen craniosynostosis syndromes. To characterize the transcriptome of the murine frontal suture (homologous to the human metopic suture) during embryonic development in mutants compared to wildtype (WT), we applied bulk and single-cell RNA-seq and network analysis. Bulk RNA-seq analysis was performed on laser capture microdissected tissue of 60 suture mesenchyme (SM) and osteogenic front (OF) subregions at embryonic day (E)16.5 and E18.5 of Apert *Fgfr2*<sup>+/S252W</sup> and Saethre-Chotzen *Twist1*<sup>+/-</sup> mutants, with widened frontal sutures at E18.5, and WT mice. In WT mice across both ages there were a combined 4,282 differentially expressed genes (DEGs) between SM and OF subregions. Mesenchymal gene expression signatures suggested functional roles as a mechanoresponsive connective tissue and developing stem cell niche. Across both subregions, there were 2,480 DEGs between E16.5 and E18.5, and changes were consistent with a decrease in proliferation and increase in maturation in both the SM and OFs from E16.5 to E18.5. In the *Twist1*<sup>+/-</sup> and *Fgfr2*<sup>+/S252W</sup> mice across both ages, there were 115 and 127 DEGs, respectively, compared to WT. Transcriptional changes affecting genes involved in vasculogenesis and ribogenesis distinguished the *Twist1*<sup>+/-</sup> and *Fgfr2*<sup>+/S252W</sup> mice from WT, respectively. Co-expression network analysis characterizing the transcriptional organization of the SM and OFs identified a novel SM gene expression module, from which we validated key driver genes regulating osteoblast differentiation. Single-cell RNA-seq identified at least five suture subpopulations that suggested an osteogenic differentiation trajectory within the suture. Altogether, our rich transcriptomic data with network biology analyses provided a model of frontal suturogenesis. Supported by grant NIH U01 DE024448.



## PgmNr 518: Are fetal endothelial nitric oxide synthase gene haplotypes risk factors in chorioamnionitis.

### Authors:

K. Yanamandra; D. Napper; J.A. Bocchini Jr.

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pediatrics, LSU Health Science Center, Shreveport, Louisiana.

### Background:

Chorioamnionitis is an infection of chorion and amnion and has a multifactorial etiology leading to growth restriction of the fetus in utero, sepsis, and premature rupture of membranes. It could risk severe prematurity, small for gestation, and fetal distress leading to respiratory failure after birth. Animal studies suggest that NO in placental vasculature is required to facilitate fetal growth. Earlier studies showed that preterm infants with severe respiratory failure respond favorably to inhaled nitric oxide. These infants may have a deficiency in the inflammatory response, including a defect in nitric oxide generation in airspaces. We have previously reported that the mutant eNOS genotypes, leading to decreased Nitric oxide, increased the risk of respiratory distress, PPRM, and IUGR, in premature infants.

### Objective:

To study the potential association of infant eNOS genotypes and amnionitis.

### Methods:

We analyzed eNOS gene haplotypes (T786C, G894T, intron 4 VNTR b/a genotypes), in 54 preterm infants (6 amnionitis cases and 48 Non-amnionitis controls), using microplate PCR genotyping methods.

### Conclusions:

Our data on eNOS double marker (T786C/Intron4 VNTR b/a; T786C/G894T; G894T/Intron4 VNTR b/a) analysis revealed that amnionitis cases showed significantly higher 50% C-a, 38% C-T, 29% T-a mutant haplotypes compared to 19%, 19%, 14% without amnionitis history respectively. Data on triple marker (T786C, Intron VNTR b/a, and G894T) analysis revealed the mutant haplotype C-a-T significantly higher 39% vs. 17% in cases vs. controls respectively. These data suggest a significant association of eNOS mutant haplotypes in double (C-a, C-T, T-a); and (C-a-T) triple markers with the amnionitis disease condition. Infants C-T, T-a); markers and (C-a-T) triple marker with the amnionitis disease condition. Infants with amnionitis have higher frequencies of mutant haplotypes, potentially reducing the endogenous NO availability thus leading to the disease. This is the first report to suggest a role for eNOS gene haplotypes in the etiology of amnionitis in premature newborn infants.

## PgmNr 519: *In utero* endotoxin exposure decreases subsequent inflammatory bowel disease severity in mice.

### Authors:

E. Banfield <sup>1</sup>; W. Fulton <sup>2</sup>; I. Burd <sup>3</sup>; C. Sodhi <sup>2</sup>; D. Hackam <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland.; 2) Department of Surgery, Johns Hopkins University, Baltimore, Maryland.; 3) Department of Gynecology and Obstetrics, Johns Hopkins University, Baltimore, Maryland.

---

Inflammatory bowel disease (IBD) affects 3 million adults in the US. Though its exact etiology remains unknown, IBD is characterized by excessive inflammation of the intestine, leading to fever, pain, and severe diarrhea. Current dogma implies that this inflammation is due to an aggressive immune response to gut microbes in a genetically susceptible host and specific environmental factors including antibiotic exposure. Though genes including *CARD15* have been implicated in IBD, the role of the pre-natal environment in onset of disease remains largely unknown. We hypothesize that in utero exposure to inflammation will increase responses to subsequent IBD triggers in genetically susceptible adults by altering expression of related IBD genes, protecting them from excessive inflammation and pain.

To mimic in utero inflammation, we injected 2 mg/kg of lipopolysaccharide into the uterine horns of pregnant CD-1 mice. At 5 weeks, the mice underwent a 6-day exposure to 2% dextran sodium sulfate (DSS), a chemical mimic of IBD. Analyses included body weight and colon length measurements and qPCR for IBD-associated genes. Relevant controls (e.g. no DSS) were used at both time points.

Colon lengths in controls that did not receive pre-natal LPS but received DSS were significantly shorter than in those that received neither exposure ( $p < 0.0001$ ), reflective of increased disease severity. Colon length was similar between mice that received both LPS and DSS exposure and controls that received only LPS exposure. Body weight measurements showed that mice exposed to LPS in utero gained an average of 1.8g during the model, regardless of DSS exposure status, revealing reduced disease. However, body weights of mice that did not receive LPS but received DSS decreased by 1.3g and those that received neither exposure gained 2.1g. We showed via qPCR that *Atg16l1* expression was significantly increased among mice that only received DSS exposure ( $p = 0.002$ ,  $p = 0.031$ ,  $p = 0.002$ ) compared to all other groups. *Card15* expression was significantly higher among mice that did not receive LPS but received DSS compared to two other groups ( $p = 0.01$ ,  $p = 0.004$ ).

Counterintuitively, these results demonstrate that pre-natal exposure to inflammation leads to a reduced development of IBD later in life, evidenced by a lack of physical and gene expression changes associated with DSS exposure in mice. This suggests that in utero inflammation exposure protects the gut from excessive inflammation into adulthood.

## PgmNr 520: Misregulation of growth-differentiation factor 11 (*GDF11*) in MeCP2-related disorders.

**Authors:**

S. Bajikar; H. Zoghbi

[View Session](#) [Add to Schedule](#)**Affiliation:** Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.

---

Loss-of-function mutations or duplication in methyl-CpG binding protein 2 (*MECP2*) cause two severe neurodevelopmental disorders: Rett syndrome and *MECP2*-duplication syndrome, respectively. While the genetic cause of these disorders is known, the mechanism by which disruption in MeCP2 levels leads to pathogenesis unknown. Disruption in normal levels of MeCP2 causes a mild change in expression of thousands of genes. The objective of this study is to identify robustly misregulated transcripts in mouse models of MeCP2-disorders to reveal candidate mediators of pathogenesis. I employed a meta-analysis of transcriptome profiles collected from *Mecp2*-null mouse models and identified Growth differentiation factor 11 (*Gdf11*) as a gene sensitive to MeCP2-levels. *Gdf11* is downregulated in *Mecp2*-null models, while upregulated in *MECP2*-transgenic models. Additionally, *Gdf11* expression levels are returned to wild-type levels when MeCP2 levels are normalized. The opposing regulation of *Gdf11* by MeCP2 implicates GDF11 misregulation as one potential driver of some phenotypes in Rett and *MECP2* duplication syndromes. GDF11 is a transforming growth factor beta family ligand that is a critical patterning morphogen for the skeletal system. However, despite broad expression throughout the brain, the role of GDF11 in brain development and homeostasis is unknown. On going and future work will test if modulating GDF11 levels specifically in the brain using either will on its own cause aberrant neurological phenotypes. Next, I will test if normalization of GDF11 levels rescues neurological phenotypes seen in mouse models of *MECP2* disorders. In sum, I have found that GDF11 is misregulated in *MECP2*-related disorders, and I will test its potential as a therapeutic modifier of these disorders.

## PgmNr 521: Human and mouse essentiality screens as a resource for disease gene discovery.

### Authors:

P. Cacheiro <sup>1</sup>; V. Muñoz-Fuentes <sup>2</sup>; T.F. Meehan <sup>2</sup>; D. Smedley <sup>1</sup>; Genomics England Research Consortium, International Mouse Phenotyping Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Pharmacology, William Harvey Research Institute, Queen Mary University of London, London, United Kingdom; 2) European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, United Kingdom

---

Although genomic sequencing has been transformative in the study of rare genetic diseases, identifying causal variants remains a considerable challenge that can be addressed in part by new gene-specific knowledge.

In the present study, we integrate measures of how essential a gene is to supporting life, as inferred from the comprehensive viability and phenotyping screens performed on knockout mice by the International Mouse Phenotyping Consortium (IMPC) and from human cell line essentiality screens. We introduce a novel, cross-species gene classification across the Full Spectrum of Intolerance to Loss-of-function (FUSIL) and demonstrate that genes in five mutually exclusive FUSIL categories have differing characteristics in the biological processes they regulate, tissue expression levels and human mutation rates. Most notably, Mendelian disease genes are highly overrepresented (2.6 fold-increase odds) in the developmental lethal category, representing genes not essential for cell survival but required for organism development. This enrichment is partially driven by an association with early onset, multisystemic, autosomal dominant disorders. Consistent with this, we find an even higher overrepresentation (3.3 fold-increase) of genes associated with developmental disorders. By combining this information with intolerance to variation scores derived from human population sequencing studies, we provide a framework to prioritise novel developmental disorder associated genes. To validate our approach, we screened 163 developmental lethal genes with no previous association to rare disease and highly intolerant to variation for candidate variants in unsolved developmental disorder cases from three independent disease sequencing consortia and identified potentially pathogenic, *de novo* variants shared in different patients. We describe findings for around half the genes and, in particular, highlight the evidence for a novel role for *VPS4A* and *TMEM63B* in intellectual disability, with patients exhibiting similar neurological and movement phenotypes that are partially modelled by the IMPC mouse.

In summary, in this work we demonstrate how by combining different resources, as diverse as human cell proliferation scores, mouse viability data and intolerance to variation scores derived from large sequencing programs, we can identify a set of genes more likely to be associated to disease and therefore propose FUSIL as an efficient resource for disease gene discovery.

## **PgmNr 522: When one tissue is not enough: Child with developmental delays and mosaic marker chromosome duplicating 11p12-11p11.2, present in amniocytes and buccal cells but not in blood cells.**

### **Authors:**

C. Melder; C. Motter

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pediatrics, Div Gen, Akron Children's Hosp, Akron, Ohio.

---

We report the case of an 8 year old girl with developmental delays, who first came to our service's attention prenatally in 2010. An amniocentesis karyotype done for advanced maternal age was read as 47,XX, +mar. To further define the marker chromosome, a prenatal oligonucleotide array was performed which showed mosaicism for a 5.3 Mb gain from 11p12 to 11p11.2 in 70% of the cells studied (using FISH BAC clone from this region). Parental FISH was negative. Delivery was unremarkable, and no further testing was done after birth. The patient returned to our clinic in October 2018; her medical history is unremarkable with the exception of headaches but she has significant expressive speech delay, learning difficulty in all areas and ADHD. A SNP array on blood in our laboratory was normal, without any sign of the 11p duplication. As we suspected there might be mosaicism in other tissues, a microarray was done on buccal cells. This was positive for mosaic (approximately 57%) heterozygous 7.67 Mb duplication at 11p11.12-p12 including many genes. Although this exact duplication has not, to our knowledge, been previously reported, the size of the duplication and the number of genes involved support that it is likely pathogenic. This phenomenon of mosaic chromosomal rearrangements found in other tissues but absent in blood is described for a few other regions, but we have not been able to find cases involving this particular region. It is interesting to note that if the prenatal testing had not been done, chromosomal testing would likely have stopped with the normal study on blood and we might have gone on to single-gene testing. This raises the question of when it is necessary to obtain samples from other tissues for clinically significant mosaicism.

## PgmNr 523: Report of the first patient with a homozygous *OTUD7A* variant responsible for epileptic encephalopathy.

### Authors:

P. Garret <sup>1,2,3</sup>; G. Delplancq <sup>1,2,4</sup>; F. Ebstein <sup>5</sup>; B. Dozieres-Puyravel <sup>6</sup>; A. Boughalem <sup>3</sup>; S. Auvin <sup>6,7</sup>; Y. Duffourd <sup>1,2</sup>; B. Zieba <sup>5</sup>; S. Mahmoudi <sup>8</sup>; K. Singh <sup>9</sup>; C. Thauvin-Robinet <sup>1,2,4</sup>; J.M. Costa <sup>3</sup>; E. Krüger <sup>5</sup>; D. Trost <sup>3</sup>; A. Verloes <sup>7,10</sup>; L. Faivre <sup>1,11</sup>; A. Vitobello <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UMR1231 GAD Inserm – Université Bourgogne-Franche Comté – Dijon – France; 2) Unité Fonctionnelle Innovation en Diagnostic génomique des maladies rares – FHU-TRANSLAD – CHU Dijon Bourgogne – Dijon – France; 3) Laboratoire CERBA – Saint-Ouen l’Aumône – France; 4) Centre de Référence maladies rares « déficience intellectuelle » – centre de génétique – FHU-TRANSLAD – CHU Dijon Bourgogne – Dijon – France; 5) Universitätsmedizin Greifswald – Institut für Medizinische Biochemie und Molekularbiologie – 17475 Greifswald – Germany; 6) AP-HP, Hôpital Robert Debré – Service de Neurologie pédiatrique – Paris – France; 7) UMR1141 INSERM – Université Paris Diderot – Sorbonne Paris Cité – Paris – France; 8) Service de Pédiatrie – Centre Hospitalier René-Dubos – Pontoise – France; 9) Stem Cell and Cancer Research Institute – Department of Biochemistry and Biomedical Sciences – McMaster University – Hamilton L8S 4L8 – Canada; 10) Genetics Department – AP-HP, Robert-Debré University Hospital – Paris – France; 11) Centre de Référence maladies rares « Anomalies du développement et syndromes malformatifs » – centre de génétique – FHU-TRANSLAD – CHU Dijon Bourgogne – Dijon – France

---

Chromosome 15q13.3 microdeletion syndrome (*MIM: 612001*), is characterized by mental retardation, epilepsy, and variable dysmorphism of the face and digits. For years, *CHRNA7* (*MIM: 118511*), encoding for the  $\alpha 7$  nicotinic acetylcholine receptor (nAChR), has been considered as the candidate gene accounting for the neuropsychiatric manifestations associated to this syndrome. However, recent studies in mouse models have shown that *OTUD7A/CEZANNE2* (*MIM: 612024*), which encodes for an ovarian tumor (OTU) deubiquitinase, should be considered as the critical gene, within the 15q13.3 region, responsible for brain dysfunction.

We performed exome sequencing in a patient presenting with severe global developmental delay, language and cortical visual impairment, infantile spasms replaced by generalized myoclonic seizures, and abnormal MRI findings.

Trio analysis allowed us to identify a homozygous *OTUD7A* missense variant NM\_130901.2:c.697C>T. The proband is the second born from consanguineous heterozygous parents, both presenting with learning disability. The brother is also heterozygous for the variant and presents school difficulties. The variant is predicted to alter an ultra-conserved amino acid, p.(Leu233Phe), lying within the OTU catalytic domain.

We are performing functional analyses on this variant which seems to be a strong candidate to explain the neuropsychiatric manifestations in this patient.

## PgmNr 524: Large-scale neuroanatomical study uncovers 198 gene associations in mouse brain morphogenesis.

### Authors:

B. Yalcin<sup>1,2,3,4,5,6</sup>; A. Mikhaleva<sup>6</sup>; K. Vrcelj<sup>7</sup>; V. Vancollie<sup>8</sup>; N. Demeure<sup>1,2,3,4</sup>; C. Wagner<sup>1,2,3,4</sup>; M. Kannan<sup>1,2,3,4</sup>; H. Whitley<sup>1,2,3,4</sup>; R. Balz<sup>6</sup>; L. Anthony<sup>8</sup>; A. Edwards<sup>9</sup>; H. Moine<sup>1,2,3,4</sup>; J. White<sup>8</sup>; D. Adams<sup>8</sup>; A. Reymond<sup>6</sup>; C. Lelliott<sup>8</sup>; C. Webber<sup>7,10</sup>; S. Collins<sup>1,2,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67404 Illkirch, France; 2) Centre National de la Recherche Scientifique, UMR7104, 67404 Illkirch, France; 3) Institut National de la Santé et de la Recherche Médicale, U964, 67404 Illkirch, France; 4) Université de Strasbourg, 67404 Illkirch, France; 5) Inserm UMR1231 GAD, University of Bourgogne Franche-Comté, 21000 Dijon, France; 6) Center for Integrative Genomics, University of Lausanne, CH-1015 Lausanne, Switzerland; 7) Department of Physiology, Anatomy and Genetics, University of Oxford, OX1 3PT Oxford, United Kingdom; 8) Wellcome Trust Sanger Institute, Hinxton, CB10 1SA Cambridge, United Kingdom; 9) Woodland view Hospital, NHS Ayrshire and Arran, Irvine KA128SS, Scotland; 10) UK Dementia Research Institute, University of Cardiff, CF24 2HQ Cardiff, United Kingdom

---

Brain morphogenesis is an important process contributing to higher-order cognition, however our knowledge about its biological basis is largely incomplete. Here we analyzed 118 neuroanatomical parameters in 1,566 mutant mouse lines and identified 198 genes whose disruptions yield NeuroAnatomical Phenotypes (NAP), mostly affecting structures implicated in brain connectivity. Groups of functionally similar NAP genes participate in pathways involving the cytoskeleton, the cell cycle and the synapse, display distinct fetal and postnatal brain expression dynamics and importantly, their disruption can yield convergent phenotypic patterns. 17% of human unique orthologues of mouse NAP genes are known loci for cognitive dysfunction. The remaining 83% constitute a vast pool of genes newly implicated in brain architecture, providing the largest study of mouse NAP genes and pathways. This offers a complementary resource to human genetic studies and predict that many more genes could be involved in mammalian brain morphogenesis.



## PgmNr 525: Placental epigenetic and gene expression changes associated with maternal dyslipidemia in early pregnancy.

### Authors:

M. Ouidir; P. Mendola; X. Zeng; S. Workalemahu; D. Shrestha; K.L. Grantz; C. Zhang; F. Tekola-Ayele

[View Session](#) [Add to Schedule](#)

**Affiliation:** Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

Although maternal lipid profiles influence placental function which may lead to early fetal programming of adult health and diseases, the underlying mechanisms remain unclear. We performed an epigenome-wide association study to identify placental DNA methylation and mRNA gene expression signatures of maternal dyslipidemia in early pregnancy.

Placental samples from the NICHD Fetal Growth Studies – Singleton cohort were profiled for DNA methylation using Illumina’s Infinium Human Methylation450 Beadchip (n=262) and RNA sequence (n=64). Maternal lipid levels were measured using blood samples collected in early pregnancy and dichotomized as high total cholesterol (TChol,  $\geq 200$  mg/dl), low high-density lipoprotein cholesterol (HDLc,  $\leq 50$  mg/dl), high low-density lipoprotein cholesterol (LDLc,  $\geq 100$  mg/dl) and high triglycerides (TG,  $\geq 150$  mg/dl). For each maternal lipid trait, we conducted epigenome-wide association analyses with covariate-adjusted robust linear regression models. Genes harboring top differentially methylated CpGs (FDR-adjusted  $P$ -value  $< 0.05$ ) were evaluated for placental mRNA gene expression.

High TChol was associated with hypermethylation of cg20637199 (*STK11*) and cg10695180 (*MOGAT2*). High LDLc was associated with hypomethylation of cg02047211 (*RPTOR*). High TG was associated with hypomethylation of cg07255925 (*HECTD2*), cg02785814 (*ALX4*) and cg04985482 (*MICA*) and hypermethylation of cg09326404 (*FAAH*). Genes annotating the top epigenome-wide significant CpGs were enriched in disease and functional pathways including cell death of mesenchymal cells ( $P$ -value =  $8.5 \times 10^{-5}$ ) and dyslipidemia ( $P$ -value =  $1.4 \times 10^{-4}$ ). High TChol, high TG and low HDLc were associated with decreased expression of *ALX4* in placenta (FDR-adjusted  $P$ -values were  $1.5 \times 10^{-3}$ ,  $1.9 \times 10^{-2}$  and  $1.3 \times 10^{-11}$ , respectively). Functional annotation found five genetic loci that regulate methylation levels (*cis*-meQTL) of cg02785814 (*ALX4*) in blood. The variants were also *cis*-eQTL with *ALX4* and *EXT2*, which are essential for fetal structure formation and angiogenesis. *ALX4* and *EXT2* exhibit the highest expression in placenta and female reproductive organs, and harbor loci associated with type-2 diabetes, hypertension, and LDLc in adults.

Our study revealed that maternal dyslipidemia in early pregnancy was significantly related to several novel placental epigenetic signatures that have been implicated in vascular and structural development and the etiology of cardiometabolic diseases in later life.



## PgmNr 526: Clustering of co-morbid conditions among *FMR1* premutation carrier women.

### Authors:

E.G. Allen <sup>1</sup>; K. Charen <sup>1</sup>; H. Hipp <sup>2</sup>; L. Shubeck <sup>1</sup>; A. Amin <sup>1</sup>; W. He <sup>1</sup>; S.L. Sherman <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University, Atlanta, Georgia.; 2) Department of Obstetrics and Gynecology, Emory University, Atlanta, GA

---

Fragile X-associated primary ovarian insufficiency (FXPOI) is seen in about 20% of women who carry an *FMR1* premutation (PM) allele (55-200 CGG repeats). These women develop hypergonadotropic hypogonadism with amenorrhea before the age of 40. A non-linear association with repeat size in the PM range and risk for FXPOI has been documented: those with a mid-range repeats are at highest risk (~70-100 repeats). Importantly, not all female carriers with 70-100 repeats experience FXPOI. Studies have also described an increased reporting of co-morbid conditions by PM women, including depression, anxiety, headaches, sleep problems, peripheral neuropathy, hypothyroidism, and many others. Furthermore, PM women are at risk for developing a late-onset fragile X-associated tremor-ataxia syndrome (FXTAS), typically after age 50. The risk for FXTAS increases linearly with repeat size, indicating a different molecular mechanism. We have collected general medical history information along with reproductive histories on 355 women with a PM, including 87 with FXPOI and 168 without FXPOI (still cycling or menopause after age 40), to more fully understand their complex health histories. Notably, anxiety, depression, tension headaches, and migraines were reported by more than 30% of all PM women. Using survival analysis to examine the age of diagnosis, adjusting for the age at evaluation, women with FXPOI compared with women without FXPOI had an earlier age at onset for anxiety, depression, tension headaches, hypothyroidism, fibromyalgia, and osteoporosis. We used cluster analysis to identify co-morbid conditions that grouped together across all PM women, and our final model consisted of eight clusters. Importantly, the majority of PM women fell into the three categories that were primarily defined by only one or a few medical issues: "Minimal conditions" (N=123), "Headaches" (N=33), and "FXPOI with minimal other conditions" (N=67). Also of note, a single cluster defined women with symptoms of FXTAS (e.g., tremor, ataxia and neuropathy), and none of these women met the criteria for FXPOI. Taking all the results together, it appears that many PM women are at risk for only minimal co-morbid conditions, and the most frequent conditions are depression, anxiety, and headaches. Further, women with symptoms of FXTAS appear to be distinct from women with symptoms of FXPOI, confirming theories that these two PM-associated phenotypes may be the result different molecular mechanisms.

## PgmNr 527: Genetic signals for vasomotor symptoms during menopause suggest potential drug targets.

### Authors:

A. Murray; R.N. Beaumont; J. Tyrrell; A.R. Wood; S.E. Jones; M.N. Weedon; T.M. Frayling; K.S. Ruth

[View Session](#) [Add to Schedule](#)

**Affiliation:** UEMS, University of Exeter, Exeter, United Kingdom

---

### BACKGROUND

Vasomotor symptoms such as hot flushes and night sweats are experienced by 80% of women around menopause and can significantly affect quality of life. On average they last around 5 years with 10% of women experiencing symptoms for >10 years. Changes in hormone levels, particularly oestrogen are key factors and hormone replacement therapy (HRT) is an effective treatment. However, the publication of health risks associated with HRT in 2002 meant a substantial reduction in uptake of the treatment in the UK – from 49% of women pre-2002 to 14% of women post-2002.

### METHODS

We have used the change in HRT use since 2002 to identify genetic variants associated with vasomotor symptoms in 150,000 post-menopausal women from UK Biobank. We compared genetic variants from a GWAS of HRT use in women taking HRT before and after 2002.

### RESULTS

We identified 15 independent signals associated with any HRT use. Eleven of the 15 signals were associated with menopause timing and of these, nine were associated with raised HRT use in women taking HRT before 2002 but not after. This implies that HRT prior to 2002 was often prescribed because a woman was menopausal, rather than for specific symptoms.

In contrast, one variant was associated with raised odds of using HRT in those taking HRT before and after 2002 (OR=1.17, 95%CI=1.13,1.21 pre 2002; OR=1.20, 95%CI=1.12,1.30 post-2002). This common variant with allele frequency of 95% was in *TACR3*, the receptor for neurokinin ( $P=1\times10^{-26}$  in all women). The locus was also identified in a previous genome-wide analysis of vasomotor symptoms ever versus never (Crandall et al 2017). Furthermore, an antagonist of the neurokinin B receptor has been shown to reduce vasomotor symptoms in a randomised control trial (Prague et al, 2017).

### CONCLUSIONS

Our study uses changes in health practice to identify a genetic variant associated with vasomotor symptoms during the menopause transition. The findings strengthen the evidence that targeting neurokinin B could be an effective treatment for the millions of women who experience menopausal symptoms, without the potential side effects associated with hormonal treatments.

## PgmNr 528: The landscape of variations of embryonic mitochondrial genome.

### Authors:

S. Madjunkova <sup>1</sup>; A. Aggarwal <sup>1</sup>; R. Abramov <sup>1</sup>; B. Wyse <sup>2</sup>; C. Librach <sup>3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Reproductive Genetics, Create Fertility Centre, Toronto, Ontario, Canada; 2) Research Department, Create Fertility Centre, Toronto, Ontario, Canada; 3) Departments of Physiology, Obstetrics and Gynaecology, and Institute of Medical Science, University of Toronto, Ontario, Canada; 4) Department of Gynaecology, Women's College Hospital, Ontario, Canada; 5) Create Fertility Centre, Toronto, Ontario, Canada

---

**Introduction:** Maternal transmission, variant selection during gametogenesis, and propensity to accumulate mutations highlight the role of mtDNA integrity in gamete maturation, fertilization, and embryonic development. Studying variations in mtDNA and evaluating effects of these on embryo development is essential in understanding the biology and outcomes of in vitro fertilization. Sparse data available on human blastocyst mtDNA variation suggests a higher rate of de-novo mutations and occurrence of heteroplasmy as a mechanism for mt dysfunction. The purpose of this study was to evaluate sequence variation of human embryonic mtDNA, and correlate this with different infertility parameters, to study their impact on implantation and embryonic development.

**Methods:** We performed a retrospective study on known outcome selected blastocysts after single euploid embryo transfer (96 from fertile-FP and 106 from infertile patients-IFP) (92-no implantation-NI, 56-ongoing pregnancy-OP, and 54-miscarried-AB). Whole genome amplified DNA from trophectoderm biopsies of blastocysts that had undergone PGT-A were utilized for targeted mtDNA library prep, using PrecisionIDmtDNA kit on the IonTorrent PGM platform. Variant calling and data analysis was performed using IonReporter, MitoMap and Partek software. Regions with >20x mtDNA coverage were included in the analysis.

**Results:** Mt genome was sequenced at av. depth of 400x and 41.6% coverage for all embryos. Mutation frequency was similar in FP and IFP: D-loop (48.2%, 41.1%), RNR1 (22.6%, 23.2%), RNR2 (14.5%, 9.8%), ND3 (6.9%, 11%), ND1 (4.8%, 3.2%), ND4L and CO1 (1.49%). Average variant load per embryo (VPE) in FP and IFP was  $6.3 \pm 7.2$  and  $7.3 \pm 8.4$  respectively, of which disease causing mutations (DCM) were  $0.73 \pm 0.81$  and  $0.75 \pm 1.1$ , respectively. 70.4% of them were homoplasmic (>80% mutant allele), OR 8.66 [95%CI: 8.64-8.67]. Embryos that resulted in OP had lower frequency of DCM compared to embryos that AB ( $p=0.0004$ ) and NI ( $p=0.03$ ). DCM in RNR1, CO1 and ND4L genes were present only in embryos that NI or AB ( $p=0.001$ ). There was no correlation between embryonic mtDNA variant load and maternal age.

**Conclusion:** Embryonic mtDNA has a high frequency of sequence variants, D-loop being the most common. Maternal age did not impact the variation load. Integrity of RNR1, CO1 and ND4L genes determines the outcome after euploid embryo transfer, however larger sample size and a validation study are necessary to confirm their predictive value.

## **PgmNr 529: 3- Hydroxyisobutyryl-Co A hydrolase (HIBCH) deficiency: An early infantile onset leigh-like presentation with two novel mutations with c.763C>T and c.428C>A on *HIBCH* gene.**

### **Authors:**

S.C. Chong <sup>1,2</sup>; S. Ho <sup>3</sup>; E. Law <sup>4</sup>; Y.P. Yuen <sup>4</sup>; H.M. Cheung <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Paediatrics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong SAR, China; 2) Joint Baylor College of Medicine-The Chinese University of Hong Kong Center of Medical Genetics, Hong Kong SAR, China; 3) Department of Paediatrics and Adolescent Medicine, Alice Ho Miu Ling Nethersole Hospital, Hong Kong SAR, China; 4) Department of Chemical Pathology, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong SAR, China

---

3-Hydroxyisobutyryl-CoA hydrolase (HIBCH) deficiency is a rare inborn error of valine metabolism which inherited autosomal recessively involving the HIBCH gene on chromosome 2q32. HIBCH is an enzyme that catalyzes the hydrolysis of 3-hydroxyisobutyryl-CoA into 3-hydroxyisobutyric acid in the mitochondria. With a mutation in both copies of HIBCH gene, catabolism of valine would be affected, leading to accumulation of methacrylyl-CoA, a highly toxic intermediate metabolite suspected to reduce mitochondrial enzyme activities. There is by far no cure for the disease and treatment remained supportive. We present a 8 years old patient who was born at 38 weeks and 6 days with birth weight 2.46kg. She presented at the age of 3 months old with clinical features resembling Leigh syndrome, failure to thrive, developmental delay, microcephaly, hypotonia, and poor visual fixation. Whole exome sequencing found two novel mutation with a heterozygous c.763C>T(p.R255X) pathogenic variant in the HIBCH gene and a heterozygous c.428C>A (p.T143K) variant of unknown significance in the HIBCH gene. Our patient developed into a phase of progressive neurodegeneration with deteriorated swallowing function, motor function, and became ventilator dependent at around 9 months old. Echocardiogram showed hypertrophic cardiomyopathy. Magnetic resonance imaging of the brain showed T2 hyperintense and T1 hypointense signal in bilateral basal ganglia, thalami, cerebral peduncles, hippocampal regions, brainstem and deep gray matter with cavitary lesion. Acylcarnitine profile showed elevation of C4-carnitine in one sample only at the time of acute neural deterioration, and soon became normalized. Gas chromatographic analysis of urinary organic acids also showed elevated lactic acid and branched-chain ketoacid. Metabolomic profiling was performed after she had the neuroregression phase, there was no valine metabolism pathway related metabolites detected. Later, she also has autonomic dysfunction with thermolability. Valine restriction diet does not change the clinical progress of the patient. The study describes two novel mutation and the natural history of HIBCH deficiency, it expands the clinical phenotype related to HIBCH deficiency.

## PgmNr 530: Newborn screening for lysosomal storage disorders in Oregon from results to clinical findings: A 6 month retrospective.

### Authors:

A. Yang <sup>1</sup>; A. Dennis <sup>1</sup>; D. Koeller <sup>1</sup>; S. Denniston <sup>2</sup>; L. Flint <sup>2</sup>; C. Biggs <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Medical Genetics, Oregon Health and Science University, Portland, Oregon.; 2) Northwest Newborn Screening Program, OHA, Hillsboro, Oregon

---

Newborn screening for lysosomal storage disorders (LSD) has been implemented in a number of states in the U.S. However, publications about their experience from initial laboratory testing through follow-up with DNA results and clinical outcomes are rare. We describe Oregon's experience during the first six months of screening for 4 LSDs from initial testing of the dried blood spot (DBS) through clinical follow-up and outcome. From October 1, 2018 through April 30, 2019, 24,209 babies were adequately screened for enzyme activity for Pompe, mucopolysaccharidosis (MPS1), Fabry, and Gaucher using DBS on the Baebies microfluidics platform. Samples with an enzyme activity below the cut-off were sent for second tier DNA testing. Of these, 5 babies were confirmed to have disease: 2 with MPS1, 2 with Gaucher, and 1 with Fabry. The approximate incidence of disease was 1 in 12,000 for both MPS1 and Gaucher and 1 in 24,000 for Fabry. Notably, both MPS1 patients have been determined to have non-Hurler disease and the Fabry patient has a late-onset allele. None of the confirmed cases are clinically indicated at this time to be treated. Two babies without current clinical signs or symptoms are being followed: one infant with pathogenic allele and pseudodeficiency allele in conjunction with a variant of uncertain significance (VUS) in the *GAA* gene, and one infant with a VUS in *trans* with a pathogenic variant in *IDUA*, who had normal urine mucopolysaccharides screen at 9 weeks of age. Finally, accounting for all known possible pseudodeficiency alleles, the false positive rates were 0.57%, 0.04%, 0.09%, and 0.80% for Pompe, MPS1, Gaucher, and Fabry disease, respectively.

## **PgmNr 531: Cytogenetic abnormalities found in 675 Mexican patients with diagnosis associated with abnormal sexual development.**

### **Authors:**

C. Alonso Muñoz; K.N. Gaytan Nares; E. E. Alonso Muñoz; C. Cortes Penagos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Cytogenetics, Laboratorios Mendel, Morelia, Michoacan, Mexico

---

Cytogenetic abnormalities found in 675 mexican patients, with a clinical diagnosis of disorders in sexual development, are evaluated. The diagnosis of each case was classified based on the sections of congenital malformations of the genital organs and adrenogenital disorders of the International Statistical Classification of Diseases and Related Health Problems of the World Health Organization.

We included 675 patients from April 2004 to March 2019with. The samples were processed according to the specifications established by The AGT Cytogenetics Laboratory Manual for obtaining metaphases stained with the GTG bands methodology. The karyotype analysis was carried out using the GEN ASIs system (Applied Spectral Imaging®).

The population was predominantly female, representing 62% of the cases, of the rest 34% corresponded to the male sex and the remaining 4% did not have data that allowed to corroborate sex. Most patients are children (0-5 years), which corresponds to 51.1% followed by adolescent (12-17 years) with 19.3%. The results of the karyotype showed that 39% of the cases present numerical and/or structural chromosomal abnormalities. According to the classification of the World Health Organization, 48.9% corresponds to Turner Syndrome, followed by indeterminate sex and pseudo hermaphroditism with 13.9%; These data coincide with that reported by other authors. The chromosomal abnormalities with greater frequency were those that presented only aneuploidies (79% of the cases) followed by aneuploidies with additional structural abnormalities (16% of the cases), those that only presented structural abnormalities represented 11% of the study population.

The sex chromosomes X and Y, presented the highest number of numerical and/or structural abnormalities. The autosome chromosomes 18, 5 and 21 showed more abnormalities than the rest of them. Of the structural abnormalities, the isochromosomes were the most frequent with 54.7% , ring chromosomes with 17%, followed by deletions and derivative chromosomes with the same number of cases represented by 9.4% for each anomaly.

Cytogenetic findings in the mexican population allow to identify the frequency of the abnormalities presented on new borns, children and young adults referred to our laboratory. The data obtained show the markers associated to the disorders of the sexual development. The information acquired has relevance of much importance for the pre and post natal medical care in our community.

## PgmNr 532: Clinical exome sequencing as a tool in prenatal diagnosis and genetic counseling: Assessment in 25 cases.

### Authors:

S. Singh; S. Thakur

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical Genetics and Fetal Medicine, FORTIS HOSPITAL and RESEARCH CENTER, NEW DELHI, India

---

**Background-** Prenatal diagnostic testing for rare genetic disorders is a rapidly advancing field. Genetic disorders resulting in prenatal or neonatal death are rare and genetically heterogeneous. An accurate diagnosis of the index case is essential for accurate prenatal diagnosis, but testing is often limited by the availability of fetal DNA. Without a potential prenatal test, couples are left in a dilemma for future pregnancies. To help parents to make an informed decision relating to the pregnancy. We describe our novel method to use exome sequencing parental DNA samples and on the affected fetus, to identify recessive monogenic disorders in a study of 25 samples referred. We demonstrate the utility of exome sequencing in reaching an accurate diagnosis and thus helping in prenatal diagnosis and genetic counseling. **Methods-** Exome sequencing was carried out in parents or fetus affected with a lethal or prenatal-onset disorder. Heterozygous rare variants in the same gene in both parents were selected for analysis. Likely, disease-causing variants were tested in fetal DNA to confirm co-segregation. **Results-** A total of 25 parents or fetus underwent whole genome sequencing or partial exome sequencing after ante-natal genetic counseling. Consanguinity was seen among 6 (24%) of the couples. Fetal DNA samples were obtained either from amniotic fluid (n=8, 40%) or by chorionic villus sampling (n=12, 60%). Whole genome sequencing (WGS) of either or both of the parents was done in 12 (48%) couples for identification of genetic illnesses. An index genetic disease case in the family was present in 16/25 (64%) couples screened and when prior fetus or child is affected, a genetic diagnosis was obtained in 69% (n=11) of the current fetus. Two couples, one each with homozygous carrier state for retinoblastoma gene and Leber's congenital amaurosis were identified during prenatal genetic counseling and thus WGS helped in prenatal counseling for risk assessment in a future pregnancy. **Conclusion-** Genome sequencing is an important tool in diagnosing genetic illnesses in a fetus who show changes in fetal ultrasonography. It is especially important for risk assessment if a history of genetic illness is present in previous pregnancies. Whole genome sequencing is important in prenatal counseling for parents.



## PgmNr 533: Clinical utility of noninvasive prenatal testing for chromosome disease syndromes in 94,000 subjects.

**Authors:**

L. Wu; D. Liang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China

---

**Purpose:** To assess the clinical performance of an expanded non-invasive prenatal test (NIPT) (named as “NIPT-Plus”) for detection of both aneuploidy and genome-wide microdeletion/microduplication syndromes (MMS).

**Methods:** A total of 94,085 women with a singleton pregnancy were prospectively enrolled in the study. The cell-free plasma DNA was directly sequenced without intermediate amplification and fetal abnormalities identified using an improved copy number variation (CNV) calling algorithm.

**Results:** A total of 1128 pregnancies (1.2%) were scored positive for clinically significant fetal chromosome abnormalities. This comprised 965 aneuploidies (1.026%) and 163 (0.174%) MMS. From follow up tests, the positive predictive values (PPVs) for T21, T18, T13, rare trisomies and sex chromosome aneuploidies were calculated as 95%, 82%, 46%, 29% and 47%, respectively. For known MMS (n=32), PPVs were 93% (DiGeorge), 68% (22q11.22 microduplication), 75% (Prada-Willi/Angleman) and 50% (Cri Du Chat). For the remaining genome-wide MMS (n=88), combined PPVs were 32% (CNVs > 10 Mb) and 19% (CNVs < 10 Mb).

**Conclusions:** NIPT-Plus yielded high PPVs for common aneuploidies and DiGeorge Syndrome, and moderate PPVs for other MMS. Our results present compelling evidence that NIPT Plus can be used as a first-tier pregnancy screening method to improve detection rates of clinically significant fetal chromosome abnormalities.



## PgmNr 534: Sleep disturbances in school age children with PTHS, SMS, and MAND mirror sleep abnormalities in autism spectrum disorder.

### Authors:

A. Gandhi <sup>1,2</sup>; D. Zhou <sup>1</sup>; J. Alaimo <sup>1</sup>; E. Chon <sup>1</sup>; M.D. Fountain <sup>1</sup>; S.H. Elsea <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 2) Rice University, Houston, TX 77005

---

Sleep disturbance in school-age children has been studied to some extent in children with Smith-Magenis syndrome (SMS). However, for neurodevelopmental disorders such as Pitt-Hopkins syndrome (PTHS) and MBD5-associated neurodevelopmental disorder (MAND), the existence of sleep disturbance in this population is not yet defined or well characterized. We surveyed caregivers of school-age children with these 3 disorders using the Children's Sleep Health Questionnaire (CSHQ) to assess how sleep disturbance differed between groups and identify disorder-specific sleep problems. We also compared our data to existing data regarding sleep disturbance in school-aged children with autism spectrum disorders (ASD). The CSHQ provided evidence for sleep disturbance in each condition (total score > 41); however, while there were similarities, specific concerns differed. The mean total CSHQ score for MAND ( $55.33 \pm 13.26$ ) was significantly different from that of ASD ( $48.83 \pm 9.68$ ), while PTHS ( $44.67 \pm 7.91$ ) and SMS ( $51.81 \pm 8.46$ ) were not. Further, mean nighttime sleep duration in hours for PTHS ( $10.18 \pm 1.52$ ) differed significantly from SMS ( $7.84 \pm 1.52$ ), MAND ( $8.41 \pm 2.2$ ), and ASD ( $8.78 \pm 1.55$ ). Overall, data indicate that sleep disturbance in PTHS is less severe than that observed in SMS and MAND. Sleep disturbance in SMS and MAND do not significantly differ from each other based on total scores on the CSHQ. Individuals with PTHS had significantly lower scores for daytime sleepiness than the ASD group; however, MAND > ASD > SMS on the daytime sleepiness subscale. Additionally, while the PTHS group did not differ from ASD for night wakings, the SMS and MAND groups scored significantly higher. All 3 groups showed significantly higher parasomnias than the ASD group. Restless sleep and bedwetting were notably present across all disorders. Overall, only parasomnias & daytime sleepiness differed significantly between PTHS and ASD, parasomnias & night wakings between SMS and ASD, and night wakings & parasomnias between MAND and ASD, suggesting that, while different aspects of sleep disturbance in each disorder mirror the wide range found in the general ASD population, certain aspects such as night wakings and parasomnias are shared and do differ significantly from ASD. These data further illustrate the complexity of these conditions and the challenges of underlying sleep disturbance in each population, indicating the need for more support, education, and treatment for these individuals.

## PgmNr 535: *FKBP8* rare variants are risk factors for human neural tube defects.

### Authors:

T. Tian <sup>1,2</sup>; X.Y. Cao <sup>1</sup>; S.E. Kim <sup>3</sup>; Y.L. Lin <sup>1</sup>; B.J. Wlodarczyk <sup>1</sup>; Y.P. Lei <sup>1</sup>; R.H. Finnell <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Environmental Health, Departments of Molecular and Cellular, Houston, TEXAS; 2) Institute of Reproductive & Child Health, Ministry of Health Key Laboratory of Reproductive Health, Department of Epidemiology & Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, PR China.; 3) Department of Pediatrics, Dell Medical School, The University of Texas at Austin, Austin, TX, USA.

---

**Background:** Neural tube defects (NTDs) are a group of serious congenital defects that occur due to a failure of neural tube closure during early embryonic development. The genetic etiology of NTDs remains poorly understood. *FKBP8*, an immunophilin family member that binds to FK506, is critical for proper neural tube ventralization through antagonism of the sonic hedgehog (Shh) signaling pathway during development. *FKBP8* also inhibits apoptosis by anchoring *BCL2* and *BCLXL* to mitochondrial membranes. Loss of *Fkbp8* results in increased apoptosis in the posterior neural tube. *Fkbp8*<sup>-/-</sup> mouse embryos showed NTDs consistent with a diagnosis of spina bifida. To date, no publication demonstrates any possible association between *FKBP8* and human NTD risk. Thus, this study provides the first evidence for a relationship between *FKBP8* and spina bifida in humans.

**Methods and Results:** Sanger sequencing was performed on 384 spina bifida cases. One loss of function variant (p.E140X) and four nonsynonymous rare mutations (p.S3L, p.K315N, p.A292S, p.A251G) which were predicted to be deleterious were identified. GFP-*FKBP8* was created by subcloning of *FKBP8* CDS region into pEGFPC1. Identified variants were introduced into GFP-*FKBP8* plasmids by site direction mutagenesis. Subcellular localization assay performed on HeLa cells showed that the variant p.E140X affected *FKBP8* localization from the mitochondrion in the cytoplasm into the nucleus. The other variants had no obvious effect on *FKBP8* protein localization. The *FKBP8* protein expression of variants p.S3L, p.K315N, p.A292S was decreased, and p.E140X produced a truncated form of *FKBP8* protein in 293T cells. Furthermore, *FKBP8* variants had enhanced Shh signaling activity as determined by Western Blot in 293T cells, indicating those variants may decrease the antagonism of *FKBP8* on the Shh pathway.

**Conclusion** Our study indicates that rare deleterious variants of *FKBP8* are risk factors for human NTDs. Rare variants of *FKBP8* could contribute to spina bifida via dysregulation of the Shh signaling pathway.

## PgmNr 536: Nephro-cutaneous syndrome in 21 Roma children with congenital EGFR mutation.

### Authors:

S. Mazurová <sup>1</sup>; V. Stranecky <sup>1</sup>; M. Tesarova <sup>1</sup>; A. Vondrackova <sup>1</sup>; H. Hansikova <sup>1</sup>; M. Giertlova <sup>2</sup>; A. Baxova <sup>3</sup>; J. Lastuvkova <sup>4</sup>; V. Chovanova <sup>5</sup>; J. Zeman <sup>1</sup>; T. Honzik <sup>1</sup>; M. Magner <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic; 2) Laboratory of Molecular Genetics, Medirex a.s., Kosice, Slovak Republic; 3) Institute of Biology and Medical Genetics, First Medical Faculty, Charles University in Prague, Czech Republic; 4) Department of Clinical Genetics, Masaryk Hospital, Usti and Labem, Czech Republic; 5) Department of Neonatology, Pediatric University Hospital, Kosice, Slovak Republic

Background: Epidermal growth factor receptor (EGFR), a transmembrane glycoprotein, mediates many cellular growth and differentiation processes, through its tyrosine-kinase activity. Acquired activating mutations are common in malignant transformation processes of various tumors. Unique loss of function, homozygous mutation is congenital and causes a distinct phenotype.

Cohort study: We report on the largest cohort of 21 Roma children with EGFR deficiency. Perinatal course involved polyhydramnios, intrauterine growth restriction and significant prematurity, followed by the development of specific skin condition initially with thin and fragile skin, later with persistent desquamative ichtiosiforme dermatitis prone to inflammation. Some children had progeroid facial features. Early postnatally recurrent infectious complications (presented as sepsis in 47 % of the cases) appeared. The prognosis was poor, all children beside one died before 6 months of age. Additionally we present the longest surviving, 13 year old boy, who developed previously unreported renal impairment with Fanconi-like syndrome and secondary hyperaldosteronism.

Results: Homozygous c.1283G>A (p.Gly428Asp) mutation in *EGFR* was confirmed in 16 of the 21 reported cases due to retrospective patient analysis.

Discussion: Severely underdiagnosed EGFR deficiency, prevalent in the Roma population, is a crippling disorder, presenting with repeated infections and ichtiosiforme dermatitis. Apart from the largest reported cohort of 20 children, we present a long-term surviving patient, who, apart from typical epidermal features, suffers from previously undescribed, severe combined renal glomerular and tubular dysfunction, which turned up to be survive limiting since infancy and may contribute to chronical growth retardation.

This work was institutionally supported by: AZV16-31932A, RVO-VFN 64165/2012, SVV 260367, UNCE 204064.

## PgmNr 537: The mortality and morbidity of very low birth weight infants with trisomy 21 in Japan.

### Authors:

H. Kawasaki<sup>1,2</sup>; T. Yamada<sup>1,2</sup>; T. Wada<sup>1,2</sup>; Y. Takahashi<sup>3</sup>; T. Nakayama<sup>3</sup>; S. Kosugi<sup>1,2</sup>; Neonatal Research Network of Japan

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Ethics and Medical Genetics, Kyoto University School of Public Health, Kyoto, Japan; 2) Clinical Genetics Unit, Kyoto University Hospital, Kyoto, Japan; 3) Health Informatics, Kyoto University School of Public Health, Kyoto, Japan

**Objectives:** Trisomy 21 (T21) is one of the major concerns in prenatal genetic testing. In the past few decades, the overall mortality and morbidity of T21 have improved through intensive management, but little is known about prognosis of very low birth weight (VLBW) infants with T21. Our purpose was to investigate mortality and morbidity of VLBW infants with T21 in Japan, compared with those of VLBWI infants with no birth defects (BD-).

**Method:** Maternal and neonatal data were collected prospectively for infants weighing less than 1,501 grams who admitted to centers of the Neonatal Research Network of Japan (NRN-J) during the period of January 2003 to December 2016. Statistical significance for unadjusted comparisons was determined using Fisher's exact test, Pearson's chi-square test, Wilcoxon test and logistic regression analysis. This study was approved by Ethics Committee of Kyoto University Graduate School and Faculty of Medicine.

**Results:** Of 60,136 VLBW infants, number of T21 was 328 (0.55 %). The frequency of maternal complications and the use of antenatal steroid was significantly low in T21 compared with BD-, except for diabetes mellitus. Cesarean section, male contribution, outborn and gestational age were significantly high in T21 compared with BD-. About two third of T21 VLBW infants were small for gestational age. Although non-reassuring fetal status was significantly high in T21, there was no significant difference in Apgar Score either at 1 minute or 5 minutes. The rates of respiratory distress syndrome, chronic lung disease and retinopathy of prematurity were significantly low in T21 compared with BD-. Mortality rate in NICU was 22.9 % for T21, compared with 5.6 % for BD-, respectively. Mean length of hospital stay for survivors was 123.1 days for T21, compared with 93.8 days for BD-.

**Conclusion:** NRN-J database contains about two third of VLBW infants born in Japan and our data is the first national survey for VLBW infants with T21. Although there is limitation that no data exists about the timing of diagnosis, which may affect the plan for care, our data showed that more than three forth of VLBW infants with T21 could achieve to go home. This will be a meaningful data for prenatal counseling, perinatal management and treatment.

## PgmNr 538: Genome sequencing analysis of fetuses with increased nuchal translucency.

### Authors:

R. Choy<sup>1,2,3</sup>; H. Wang<sup>4</sup>; M. Shi<sup>1</sup>; J. Chen<sup>5</sup>; Z. Yang<sup>1</sup>; R. Zhang<sup>4</sup>; H. Yan<sup>5</sup>; Y. Wang<sup>4</sup>; S. Chen<sup>4</sup>; M. Chau<sup>1</sup>; Y. Cao<sup>6</sup>; O.Y.M. Chan<sup>1</sup>; Y.K. Kwok<sup>1</sup>; Y. Zhu<sup>4</sup>; M. Chen<sup>5</sup>; T.Y. Leung<sup>1,2,3</sup>; Z. Dong<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Obstetrics & Gynaecology, Chinese University of Hong Kong, Hong Kong, China; 2) Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; 3) The Chinese University of Hong Kong-Baylor College of Medicine Joint Center for Medical Genetics, Hong Kong, China; 4) Department of Central Laboratory, Bao'an Maternity and Child Healthcare Hospital Affiliated to Jinan University School of Medicine, Key Laboratory of Birth Defects Research, Birth Defects Prevention Research and Transformation Team, Shenzhen, China; 5) Department of Obstetrics and Gynecology, Key Laboratory of Reproduction and Genetics of Guangdong Higher Education Institutes, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China; 6) Department of Molecular and Human Genetics, Baylor College of Medicine Houston, Texas, USA

---

**Background:** Increased Nuchal Translucency (NT) is an important biomarker which is associated with higher risk of fetal structural anomalies. It is known to be contributed by different genetic etiologies ranging from single-base mutations to those affecting millions of base-pairs. Currently, prenatal diagnosis is routinely performed by karyotyping and chromosomal microarray analysis (CMA), however, both of them are limited in resolution. The diversity of the genetic etiologies warrants an integrated assay such as genome sequencing (GS) for comprehensive detection of genomic variants. Herein, we aim to evaluate the feasibility of applying GS in prenatal diagnosis for the fetuses with increased NT.

**Methods:** We retrospectively applied GS (>30-fold) for the fetuses with increased NT ( $\geq 3.5$ -mm), which were undergone routine prenatal diagnosis. Detection of single-nucleotide variants, copy-number variants and structural rearrangements were performed simultaneously and the results were integrated for interpretation in accordance with the guidelines of American College of Medical Genetics and Genomics. Pathogenic or likely pathogenic (P/LP) variants were selected for validation and parental confirmation, if available.

**Results:** Overall, 50 fetuses were enrolled, including 34 isolated and 16 with reported fetal structural malformations. Routine CMA/karyotyping reported eight P/LP CNVs, yielding a diagnostic rate of 16.0% (8/50). In comparison, GS provided a 2-fold increased diagnostic yield (32.0%, 16/50), including one mosaic turner syndrome, eight cases with microdeletions/microduplications and seven cases with P/LP point mutations. Moreover, GS identified two cryptic insertions and two inversions. Follow-up study further demonstrated the potential pathogenicity of a balanced insertion disrupting an OMIM autosomal dominant disease-causing gene at the inserted site.

**Conclusions:** Our study demonstrates that applying GS in fetuses with increased NT is able to comprehensively detect and delineate the spectrum of genomic variants contributory to the diseases. Importantly, prenatal diagnosis by GS doubled the diagnostic yield compared with routine protocols. Given a comparable turn-around-time and less DNA required, our study provides a strong evidence to facilitate GS in prenatal diagnosis, particularly in fetuses with increased NT.

## PgmNr 539: HCFC1: A new transcriptional regulator coupling vitamin B12 metabolism to neural crest development.

**Authors:**

T. Chern; X. Tong; A. Achilleos; R. Poché

[View Session](#) [Add to Schedule](#)

**Affiliation:** Baylor College of Medicine, Houston, Texas.

---

Recently, patients with *cbIX*, an X-linked variant of the *cbIC* cobalamin (vitamin B12) deficiency syndrome, were discovered. While *cbIX* phenocopies most of the clinical symptoms observed in *cbIC*, the disease is caused by mutations in the transcription cofactor *HCFC1* (*Host Cell Factor C1*). Previous research has determined that *HCFC1*, along with the transcription factor *RONIN* (*THAP11*), directly regulates the expression of *MMACHC*, which encodes an enzyme essential for cobalamin metabolism and is mutated in *cbIC* patients. Thus, the origin of the *cbIX* phenotype is likely in part due to transcriptional reduction of *MMACHC* during development.

In order to address this possibility and to define the exact pathophysiology of *cbIX*, our group has generated the *Hcfc1 A115V* mouse model, which carries the most frequent missense mutation observed in *cbIX* patients. Unexpectedly, the *A115V* hemizygous mice exhibit a fully penetrant skin hypopigmentation phenotype and craniofacial dysmorphia, indicating a potential defect in neural crest cell (NCC) development. These data have led us to hypothesize that *HCFC1* functions in a transcription factor complex to regulate NCC development through regulating *Mmachc* and cobalamin metabolism.

We have determined that, analogous to *cbIX* patients, *A115V* hemizygous mice have a dramatic reduction of *Mmachc* levels in embryonic tissue. Moreover, the hypopigmentation phenotype is rescued by over-expressing *Mmachc* in these mice. These data suggest that we have uncovered a new transcriptional program influencing NCC development and raises the intriguing possibility that disrupted cobalamin metabolism may also play a role. Current efforts are aimed at determining the specific requirement for cobalamin during NCC development.

## PgmNr 540: Placental transcriptome analysis using RNA sequencing in monochorionic twin pregnancies complicated by selective growth restriction.

**Authors:**

T. Leung

[View Session](#) [Add to Schedule](#)

**Affiliation:** Obstetrics & Gynaecology, The Chinese University of Hong Kong, Hong Kong, China

---

**Objective:** to investigate if there any difference in placental gene expression between the selective intrauterine growth (sIUGR) twin and the normally growing co-twin in monochorionic diamniotic (MCDA) twin pregnancies

**Method:** This was a prospective study on differential gene expression in MCDA twin placenta complicated by sIUGR. The normally growing twin of MCDA pairs with sIUGR was used as a reference for control genetic factors of the fetus and maternal factors of the mother. This study is divided into 2 stages: (1) Gene discovery phase: Placental tissues from 5 sIUGR and 2 control pairs were sampled for transcriptome profiling. (2) Validation phase: Placental tissues from 13 sIUGR and 6 control pairs were collected for RNA and protein validation. Placental biopsies were obtained from women with sIUGR and matched controls in the delivery ward. In the gene discovery phase, the transcriptome profiling was discovered by RNA sequencing (RNA-seq). In the validation phase, RNA and protein expression level of candidate genes were determined by quantitative real-time PCR (qRT-PCR) and immunohistochemistry (IHC) staining.

**Result:** A total of 2455 transcripts were differentially expressed in placenta of all sIUGR-MCDA pairs, where 1043 were up-regulated and 1412 down-regulated. OS9 and M6PR were expressed in the trophoblasts of chorions. RNA and protein expression were significantly upregulated in sIUGR ( $p < 0.05$ ).

**Conclusion:** Endoplasmic Reticulum Lectin (OS9) and Mannose-6-Phosphate Receptor (M6PR) were increased in placenta of sIUGR-MCDA twin, but not in their normal growth co-twin.



## **PgmNr 541: Exome sequencing analysis, based on bioinformatics scores for prioritizing variants before fetal reverse phenotyping, optimize diagnosis and identify novel candidate genes in fetuses with multiple congenital abnormalities.**

### **Authors:**

A. Bruel <sup>1,2</sup>; M. Lefebvre <sup>1</sup>; E. Tisserant <sup>1</sup>; N. Bourgon <sup>1</sup>; Y. Duffourd <sup>1</sup>; M. Assoum <sup>1</sup>; P. Kuentz <sup>1</sup>; S. Nambot <sup>1,3</sup>; D. Lehalle <sup>1,3</sup>; S. Moutton <sup>1,3</sup>; N. Marle <sup>4</sup>; C. Philippe <sup>1,2</sup>; F. Tran Mau-Them <sup>1,2</sup>; J. Thevenon <sup>1,3</sup>; L. Faivre <sup>1,3</sup>; C. Thauvin-Robinet <sup>1,3</sup>; A. Vitobello <sup>1,3</sup>; FOETEX Physicians' Group

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Inserm - UB UMR 1231 GAD « Génétique des Anomalies du Développement », FHU-TRANSLAD, Dijon, France; 2) Unité Fonctionnelle d'Innovation diagnostique des maladies rares, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon France; 3) Centre de Référence « Anomalies du Développement et syndromes malformatifs », Centre de Génétique, FHU-TRANSLAD, CHU Dijon Bourgogne, Dijon, France; 4) Laboratoire de génétique chromosomique et moléculaire , CHU de Dijon, France

---

Singleton exome sequencing (sES) harbors a diagnostic yield in fetuses with multiple congenital abnormalities (MCA) almost 10% lower than in live birth individuals. Such difference is largely explained by a poor knowledge of the fetal clinical hallmarks associated with pathogenic variants in morbid genes, which hinders the ability to establish an exact genotype-phenotype correlation. We gathered a large prospective multicentric cohort of 95 MCA undiagnosed fetuses, and deployed a multistep sES analysis approach to facilitate the identification of causative variants and of novel candidate genes. In the first step, we prioritized variants making use of a genotype-first approach based on bioinformatics scores and variant annotation databases, blind to fetal phenotypes. We selected in OMIM genes associated with developmental anomalies (OMIM-da): (a) high-confidence (likely)pathogenic annotation in ClinVAR, compatible with the expected inheritance model; (b) truncating variants in genes being intolerant to loss-of-function (LoF-I) and autosomal dominant inheritance, (c) predicted damaging recessive variants in autosomal recessive or X-linked OMIM-ad. Causality was confirmed by reverse phenotyping on autopsic findings. In the second step, the remaining negative fetuses were analysed according current clinical practices based on phenotype-genotype correlations. Lastly, we analysed LoF variants in all OMIM genes not related to human diseases so far, prioritizing: (1) LoF variants in LoF-I genes; (2) recessive LoF (homozygous of compound heterozygous) variants. Family segregation of the selected variants was ascertained by Sanger sequencing. The first step resolved 18/95 patients, while the second step allowed the identification of 5 additional cases (global yield 24%). Variants of Unknown clinical Significance (VUS) in OMIM-ad genes were retained as candidates for 18/95 additional patients (18.9%). Among these, 2 strong candidate variants could not be classified as causal because of the poor fetal phenotype-genotype correlation. The third step identified eight novel candidate genes (8.4%). Our approach shows that a genotype-first approach followed by reverse phenotyping seems to be a valid alternative to a phenotype-driven variant hunting in sES data prioritization, in the context of low or partial genotype-phenotype correlation., This approach owns the potential to fill in the gap between fetal and



postnatal diagnostic yield when trio analysis is not available.

## PgmNr 542: Fetal phenotype of Rubinstein-Taybi syndrome caused by *CREBBP* mutations.

### Authors:

J. Van Gils <sup>1</sup>; S. Naudion <sup>1</sup>; J. Toutain <sup>1</sup>; G. Lancelot <sup>1</sup>; T. Attié-Bitach <sup>2</sup>; S. Blesson <sup>3</sup>; B. Deemer <sup>4</sup>; B. Doray <sup>5</sup>; M. Gonzales <sup>6</sup>; J. Martinovic <sup>2</sup>; S. Whalen <sup>7</sup>; L. Taine <sup>1</sup>; B. Arveiler <sup>1,8</sup>; D. Lacombe <sup>1,8</sup>; P. Fergelot <sup>1,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University Hospital, Bordeaux, Bordeaux , France; 2) Service d'Histologie, Embryologie, Cytogénétique, GH Necker-Enfants Malades, APHP, Paris, France.; 3) Foetopathologie, Service de Génétique, CHU de Tours, Tours, France.; 4) Génétique Clinique, Hôpital Nord, CHU Amiens Picardie, Amiens, France.; 5) Fédération de Génétique, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.; 6) Service de Génétique et d'Embryologie Médicales, Hôpital Armand Trousseau, APHP, Paris, France.; 7) Département de Génétique, GH Pitié-Salpêtrière, APHP, Paris, France.; 8) Laboratoire MRGM, INSERM U1211, Université de Bordeaux, Bordeaux, France.

---

Rubinstein-Taybi syndrome (RSTS; OMIM 180849) is an autosomal dominant developmental disorder characterized by facial dysmorphism, broad thumbs and halluces associated with intellectual disability. RSTS is caused by alterations in *CREBBP* (about 60%) and *EP300* genes (8%). RSTS is often diagnosed at birth or during early childhood but generally not suspected during antenatal period. We report nine cases of well-documented fetal RSTS. Two cases were examined after death in utero at 18 and 35 weeks of gestation and seven cases after identification of ultrasound abnormalities and termination of pregnancy. On prenatal sonography, a large gallbladder was detected in two cases, and brain malformations were noted in four cases, especially cerebellar hypoplasia. However, the diagnosis of RSTS has not been suggested during pregnancy. Fetal autopsy showed that all fetuses had large thumbs and/or suggestive facial dysmorphism. A *CREBBP* gene anomaly was identified in all cases. Alterations were similar to those found in typical RSTS children. This report will contribute to a better knowledge of the fetal phenotype to consider the hypothesis of RSTS during pregnancy. Genotyping allows reassuring genetic counseling.

## PgmNr 543: Clinical utility of a 12-gene Noonan spectrum disorders panel in prenatal and postnatal cohorts.

### Authors:

Y. Jiang <sup>1</sup>; H. Dai <sup>2</sup>; L. Wang <sup>1</sup>; S. Peacock <sup>1</sup>; E. Schmitt <sup>1</sup>; G. Wang <sup>1</sup>; S. Chen <sup>1</sup>; Y. Feng <sup>1</sup>; E. Gorman <sup>1</sup>; C. Chen <sup>1</sup>; H. Mei <sup>2</sup>; Z. Chen <sup>2</sup>; J. Zhang <sup>2</sup>; L. Wong <sup>2</sup>; C. Eng <sup>2</sup>; Y. Wang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor Genetics Laboratories, Houston, Texas, United States; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, United States

---

Noonan spectrum disorders are a clinically overlapping group of autosomal dominant diseases caused by germline mutations in genes involved in the RAS-MAPK signaling pathway with a prevalence at 1/1000-2500. This study evaluates the clinical utility of a 12-gene Noonan spectrum disorders panel in prenatal and postnatal cohorts. Two hundred thirty-four fetal samples and 113 postnatal specimens were sequenced in our laboratory between 2014 and 2019. In the prenatal cohort, the average gestational age at testing was 16.2 weeks (range: 11.1-25.3). Increased nuchal translucency (NT) was the most common prenatal ultrasound finding, present in 71.4% of the fetuses, followed by cystic hygroma (13.2%). Pathogenic or likely pathogenic variants were detected in four genes in the prenatal cohort: *PTPN11* (n=4), *RIT1* (n=2), *SOS1* (n=1) and *RAF1* (n=1). Among these eight positive cases, three fetuses showed solely cystic hygroma, two showed increased NT, one showed cardiac defect and facial dysmorphism, and two had unspecific ultrasound findings. In the postnatal cohort, 55 patients harbored a pathogenic or likely pathogenic variant spread amongst eight genes: *PTPN11* (n=29), *SOS1* (n=11), *BRAF* (n=4), *RIT1* (n=4), *HRAS* (n=3), *SHOC2* (n=2), *KRAS* (n=1) and *RAF1* (n=1). The average age at testing for the patients with positive results was 4.8 yrs (range: 0-18.8). Facial dysmorphism was the most common clinical phenotype, present in 61.8% of patients with positive results and specified clinical indications, followed by cardiac defects (47.1%), developmental delay (44.1%) and short stature (35.3%). The overall diagnostic yield was 3.4% (n=8/234) for the prenatal cohort and 48.7% (n=55/113) postnatally. Three genes, *PTPN11*, *SOS1* and *RIT1*, comprised 87.5% (n=7/8) of total disease-causing variants in the prenatal cohort and 80.0% (n=44/55) in the postnatal cohort, suggesting that the pattern of specific gene involvement is similar between the two groups. Of note, our diagnostic yield of 48.7% (n=55/113) in the postnatal cohort was significantly higher than that in two recent reported postnatal cohorts: 21.3% (n=87/409) and 21.7% (46/212). Different patient cohorts may explain the higher diagnostic yield in our test. Another contributing factor might be the inclusion of the *RIT1* gene, as it was not present in the NGS panel in one study and only tested in a very limited number of samples in the other.

## PgmNr 544: Tandem whole genome analysis enables comprehensive family planning.

**Authors:**

N. Neerman; R. Jyoti; A. Kaplun; J. Hogan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Variantyx Inc., Framingham, Massachusetts.

---

**Background:**

A variety of carrier screening tests are offered with the promise of helping young couples with family planning. Available tests generally focus on a small subset of specific conditions and genes. While these tests are able to detect the most common causes of selected recessive conditions, they are unable to provide information about the large number of rare disorders that are less common. On the surface, when looked at on an individual basis, these disorders are deceptively rare. It's only when the set of more than 5,000 diseases with a genetic basis are considered as a whole, it becomes apparent that the impact is significant. Whole genome sequencing (WGS) provides the ability to detect variants that cause these conditions, as well as the ability to identify complex variants that cause the more commonly screened conditions but go undetected with other tests. Tandem analysis of reproductive partner genomes via WGS has the potential to provide a comprehensive assessment of shared carrier risks.

**Method:**

To test the tandem analysis approach, we started with our clinical WGS pipeline which has been validated for detection of small sequence changes, mitochondrial variants, structural variants and more than 20 tandem repeat expansions. This is the same pipeline that is currently used for the clinical WGS tests performed in our lab, including our Genomic Inform™ test for healthy individuals - a physician-ordered clinical test that includes carrier status analysis and reporting. We then adapted the annotation filters and variant interpretation workflow to additionally process and analyze whole genomes of potential reproductive partners in tandem, evaluating more than 50 paired genomes.

**Result:**

Identified variants included small sequence changes in commonly (i.e. CFTR: cystic fibrosis) and uncommonly (i.e. SEPN1: multiminicore disease; congenital myopathy with fiber-type disproportion) screened genes, large structural variants as well as premutation alleles of tandem repeat genes like FMR1 which is responsible for Fragile X syndrome. The tandem analysis workflow successfully identified dual carrier genome pairings with the potential to produce offspring inheriting two copies of the same deleterious variant. It also identified dual carrier pairings that involved the potential to inherit compound heterozygous variants in the same gene, including examples of structural variants paired with small sequence changes. We will present examples for each case type.

## PgmNr 545: Haploinsufficiency of *PRR12* causes a congenital multiple-malformation neurodevelopmental syndrome with a wide phenotypic spectrum.

### Authors:

W. Bi <sup>1,2</sup>; L. Wang <sup>2</sup>; S. Lalani <sup>1</sup>; S. Heide <sup>3</sup>; B. Keren <sup>3</sup>; C. Mignot <sup>3</sup>; M.A. Magriñá <sup>4</sup>; G. Morley <sup>5</sup>; J. Walia <sup>6</sup>; P. Wheeler <sup>7</sup>; J. Lemons <sup>8</sup>; D. Rodriguez-Buritica <sup>8</sup>; E. Riberi <sup>9</sup>; E. Biamino <sup>9</sup>; K. Schatz <sup>10</sup>; M. Gunay-Aygun <sup>10</sup>; A. Wiesener <sup>11</sup>; C. Zweier <sup>11</sup>; O. Caluseriu <sup>12</sup>; Z. Sedláček <sup>13</sup>; M. McDonald <sup>14</sup>; S. Montgomery <sup>14</sup>; I. Wentzensen <sup>15</sup>; V.M. Siu <sup>16</sup>; T.B. Balci <sup>16</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor College of Medicine, Houston, TX; 2) Baylor Genetics Laboratories, Houston, TX; 3) GH Pitié-Salpêtrière, Paris, France; 4) Medical Specialties Unit From City Hall São José dos Campos, São Paulo, Brazil; 5) Mercyhealth Hospital, Rockford, IL; 6) Queen's University, Kingston, ON, Canada; 7) Arnold Palmer Hospital for Children, Orlando, FL; 8) University of Texas Health Science Center McGovern Medical School, Houston, TX; 9) University of Turin, Turin, Italy; 10) Johns Hopkins University School of Medicine, McKusick-Nathans Institute of Genetic Medicine, Baltimore, MD; 11) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 12) Department of Medical Genetics, University of Alberta, Edmonton, AB, Canada; 13) Department of Biology and Medical Genetics, 2nd Faculty of Medicine and University Hospital Motol, Charles University, Prague, Czech Republic; 14) Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, NC; 15) GeneDx, Gaithersburg, MD; 16) Division of Genetics, Department of Paediatrics, University of Western Ontario, London, ON, Canada

---

*De novo* heterozygous truncating mutations in the proline rich 12 (*PRR12*) gene have been reported in three patients with global developmental delay, intellectual disability, eye and vision abnormalities and other features. Consistently, expression pattern of *PRR12* supports its role in brain and eye development. However, the disease-association of *PRR12* and the disease-causing mechanism have not been well established. In this report, we describe 15 additional unrelated individuals with neurodevelopmental issues who carry 14 rare and damaging variants in *PRR12*. These variants were found to be *de novo* in all cases where parental samples were available (n=13). All variants were apparently loss-of-function (7 frameshift, 5 nonsense, 1 splice site), except for one *de novo* missense variant. In addition, a 3.3 Mb deletion encompassing the entire *PRR12* gene, which was a recombination product of a maternal balanced insertion, was identified in one patient. The phenotype varied among the 19 patients with defects in *PRR12*. The major clinical features were developmental delay (84%), eye and visual abnormalities (68%), cognitive impairment (63%), behavioral problems (58%), and growth retardation (53%). The congenital ophthalmic anomalies included coloboma of the iris, lens, retina and/or optic nerve, microphthalmia, anophthalmia, and cryptophthalmos. The less frequent features included abnormal outer ears (47%), congenital heart defects (37%) and tone abnormalities (37%). Notably, variations in eye anomalies were also observed in the two patients carrying the same p.Lys1092Argfs\*131 variant; one patient had a unilateral iris, chorioretinal and optic nerve coloboma with ipsilateral microphthalmia; while the other one only presented with oblong optic nerves and bilateral ptosis. The common features in patients with *PRR12* sequence variants, including iris coloboma and intellectual disability, were also observed in the patient with a large

deletion including *PRR12*, supporting that haploinsufficiency of *PRR12* is the disease-causing mechanism. In summary, our studies demonstrated that *PRR12* haploinsufficiency leads to a multi-system neurodevelopmental syndrome with a wide clinical spectrum and variable phenotypic expressivity.

## PgmNr 546: Exome sequencing enhances the diagnostic rate of perinatal autopsy: A prospective multicentre clinical utility trial with implications for prenatal diagnosis.

### Authors:

A. Vasudevan<sup>1,5</sup>; F. Chan<sup>6,8</sup>; A. Yeung<sup>2,8,9</sup>; Z. Stark<sup>2,7</sup>; S. Prystupa<sup>6</sup>; Y. Chan<sup>10</sup>; T. Leong<sup>4</sup>; K. Ireland-Jenkins<sup>4</sup>; S. Fawcett<sup>1</sup>; M. Graetz<sup>3</sup>; K. Rose<sup>9</sup>; S. Ayres<sup>2</sup>; A. Jarmolowicz<sup>2</sup>; G. Brett<sup>2</sup>; Y. Prower<sup>9</sup>; H. Chaliner<sup>4</sup>; C. Dao<sup>3</sup>; T. Davis<sup>1</sup>; L. Hui<sup>3,7</sup>; M. Teoh<sup>9</sup>; S. Rowlands<sup>1</sup>; S. Walker<sup>2</sup>; E. Lynch<sup>2</sup>; M. Martyn<sup>7</sup>; B. Chong<sup>2</sup>; C. Gaff<sup>7,8</sup>; S. Lunke<sup>2</sup>; J. Collet<sup>6</sup>; G. McGillivray<sup>1,2,3,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Genetics, The Royal Women's Hospital, Parkville, VICTORIA, Australia; 2) Victorian Clinical Genetics Services, 50 Flemington Rd, Parkville VIC 3052, Australia; 3) Mercy Perinatal, Mercy Hospital For Women, 163 Studley Rd, Heidelberg VIC 3084, Australia; 4) Austin Hospital, 145 Studley Rd, Heidelberg VIC 3084 Australia; 5) Clinical Genetics Service, Sunshine Hospital, 176 Furlong Rd, St Albans VIC 3021 Australia; 6) Department of Anatomical Pathology, The Royal Womens Hospital, Parkville, VICTORIA, Australia; 7) Murdoch Children's Research Institute, Royal Children's Hospital, Flemington Rd, Parkville VIC 3052, Australia; 8) Melbourne Genomics Health Alliance, 50 Flemington Rd, Parkville VIC 3052 Australia; 9) Monash Medical Centre, 246 Clayton Rd, Clayton VIC 3168, Australia; 10) Department of Anatomical Pathology, Monash Medical Centre, 246 Clayton Rd, Clayton VIC 3168, Australia

### Objectives:

To determine the utility of exome sequencing as an adjunct to perinatal post mortem for congenital anomalies.

To model the likely outcome of exome sequencing as a prenatal test in the same setting

### Method:

Probands with congenital anomalies were referred by perinatal pathologists. They were enrolled for sequencing if their microarray analysis was negative and their anomalies were considered to have a monogenic cause. Singleton or trio exome sequencing was performed as an adjunct to routine perinatal autopsy and the diagnostic outcomes were compared. A geneticist independently reviewed the probands' antenatal imaging findings and recommended a phenotype specific gene list to model the clinical utility of prenatal exome sequencing.

### Results:

A prospective cohort of 131 probands was referred. Forty-nine (37%) were unsuitable for inclusion. The parents of 5 (4%) declined enrolment and 10 (8%) could not be consented. Sixty-seven probands (52%) were enrolled. One proband could not be sequenced due to a degraded DNA sample. Results are available for 65 probands (32 singletons and 33 trios).

Specific diagnoses were identified at autopsy in 11 cases (17%) including 2 cases with negative sequencing. Sequencing identified specific diagnoses ('pathogenic' or 'likely pathogenic' variants) in 23 cases giving a diagnostic rate of 35%. The combined diagnostic rate of autopsy and exome was 38%. The individual autopsy and genomic diagnostic rates were highest in probands with significant skeletal findings (39% and 61% respectively, n=18). Genomic diagnoses were obtained from 34% of singleton exomes, with segregation when required, and 36% of trio exomes.

Variants of uncertain significance (VUS) were reported in 13 cases (20%). In four, a strong phenotype-genotype match together with plausible candidate variants indicated a need for additional studies. The combined rate of diagnostic or suspicious variants was 42%. The use of antenatal sequencing in this cohort using a candidate gene list approach would have identified a specific genomic diagnosis in 78% (18 of the 23 cases).

**Conclusion:**

Clinical Exome sequencing doubles the diagnostic rate of perinatal autopsy for congenital anomalies and supports the prenatal use of genomic sequencing



## PgmNr 547: *PRNP* gene deletion by CMA in a fetus with congenital abnormalities: First case report.

### Authors:

P. Onur; M. Shaver; Y. Leong; M.A. Iqbal

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, Rochester, NY.

---

Mutations in the *PRNP* gene located at chromosome 20p13 band are associated with neurodegenerative spongiform encephalopathies also known as prion disease. We report a 21-year-old pregnant woman (G4P2A1) with a history of ADHD, bipolar disorder, chronic depression, smoking and use of THC during pregnancy referred for fetal congenital anomalies including bilateral clubbed feet, rocker bottom feet and absent cavum septum pellucidum as well as intrauterine growth retardation (IUGR) identified at 29-week ultrasound. Amniocentesis revealed a normal karyotype of 46,XX by routine G-banding. A follow up chromosome microarray comparative genomic hybridization (CMA) using 4 × 180K + SNP v2.0 platform (Agilent Technologies) revealed a 759 Kb loss at 20p13 from 4053819 to 4813038 base pairs encompassing the *PRNP* morbid gene, confirmed by FISH with RP11- 960N2(20p13, SO) and Tel 20p(SG) probes. The mother had an uncomplicated C-section at 37w0d. Her APGAR scores were 9 at 1 minute and 9 at 5 minutes. Neonatal evaluation revealed that the baby was small-for-gestational-age. Her birth weight was 1.810 kg (0.3 percentile), length was 41.0 cm (0.1 percentile) and head circumference was 33.0 cm (39. 7 percentile). Her left foot revealed rocker-bottom appearance with dorsal deviation. Additionally, right foot was dorsiflexed with resistance to extension of the ankle. No cardiac or respiratory abnormalities were seen and neurologic reflexes including Moro, pupillary, fair suck were normal. This report describes the first prenatal case with a *PRNP* gene deletion and highlights the value of CMA as a highly sensitive assay in detecting critical gene deletions/disruptions which can help establish important genotype-phenotype correlation. Our case also highlights the possible role of *PRNP* in ultrasound abnormalities observed in the fetus. Additional similar cases will be helpful in elucidating role of *PRNP* in early embryogenesis.

## PgmNr 548: Vascular abnormalities in murine embryos with ubiquitous expression of *Akt1* p.(E17K).

### Authors:

M.J. Lindhurst <sup>1</sup>; W. Li <sup>2</sup>; S.M. Wincovitch <sup>3</sup>; X. Ma <sup>4</sup>; J.J. Shwetar <sup>1</sup>; Y. Mukouyama <sup>2</sup>; L.G. Biesecker <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 2) Laboratory of Stem Cell and Neuro-Vascular Biology, Cell and Developmental Biology Center, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD; 3) Genetic Disease Research Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 4) Laboratory of Molecular Cardiology, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD

---

Mosaic expression of the c.49G>A, p.(E17K) *AKT1* variant causes Proteus syndrome. To create a mouse model of the disorder, we made a CRE conditional *Akt1*p.(E17K) allele. To test the Happle hypothesis, which states that the Proteus syndrome variant only survives by mosaicism, we activated the variant allele by crossing *Akt1* conditional and ACTB-Cre mice which express Cre ubiquitously. Embryos from this cross were not viable and died between E11.5 and E17.5, confirming the Happle hypothesis. Mutant embryos were pale, had fewer visible blood vessels and increased hemorrhage in the skin. Whole-mount immunostaining of E14.5 and E15.5 mutant skin showed no mature arteries or veins in the vasculature, suggesting a vascular remodeling defect. To determine the extent of the vascular defects in these embryos, we used the iDISCO clearing method to perform whole mount immunolabeling and volume imaging on mutant and wild type E11.5 – E14.5 embryos. Examination of E12.5 embryos showed that the vertebral arteries failed to complete formation with heterogeneous, abnormal connections of the basilar artery to the aorta. In E13.5 mutant embryos, several vessels including the thoracic and external carotid arteries were missing, underdeveloped, or had abnormal patterning. In addition, hematoxylin and eosin staining of sections from E11.5 – E14.5 embryos showed heterogeneous blood-filled spaces that when visualized by whole mount immunolabeling were shown to be vascular channels surrounded by alpha-smooth muscle actin-positive cells. Taken together, we conclude that constitutive AKT signaling impairs the remodeling of the developmental vascular capillaries into a hierarchical vascular network, and leads to structural vascular anomalies. We speculate that at a cellular level, these vascular anomalies are due to abnormal endothelial cell behaviors such as sprouting, migration, and subsequent endothelial cell fusion, and that disconnection of major arterial networks may cause abnormal circulation, resulting in defective vascular remodeling in peripheral vasculature.

## PgmNr 549: An infant with ambiguous genitalia and complex mosaicism for Y-chromosome abnormalities not detected by prenatal cell-free fetal DNA testing.

### Authors:

W. Wilson <sup>1</sup>; M. Briley <sup>1</sup>; A. Kakkanatt <sup>1</sup>; L. Karns <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, UVA Health System, Charlottesville, VA; 2) Department of OB-Gyn, UVA Health System, Charlottesville, VA

---

Analysis of cell-free fetal DNA has become a commonly-used prenatal diagnostic test. We report an infant whose cell-free fetal prenatal study was “normal”, but who has ambiguous genitalia and was found to be mosaic for a complex Y-chromosome abnormality. The mother, age 30 years, presented for screening ultrasonography at 12 weeks gestation because of apparent increased nuchal translucency on an earlier ultrasound examination. Exam at 12 weeks showed a 5 mm nuchal translucency (increased). A maternal blood sample was sent for cell-free fetal DNA studies, the results of which were “consistent with a female fetus”. The fetal DNA fraction was estimated by the lab to be 4%. Follow-up prenatal ultrasound examinations at 17 and 20 weeks post LMP were unremarkable, without increase in the nuchal translucency, and were consistent with female genitalia. The baby was born at term, weighing 3.02 kg. Physical exam was unremarkable other than prominent ears and ambiguous external genitalia. She had a hooded phallus, measuring 1.7 cm, with a meatus at the tip and another small opening at the base. The labioscrotal folds were fused and slightly rugated. There were no palpable gonads. Pelvic ultrasonography demonstrated a uterus thought to be appropriate in size for age, but identified no gonads. Her state newborn screen was unremarkable, including normal 17-hydroxyprogesterone. Serum electrolytes were normal. At 8 days of age, serum DHEA SO<sub>4</sub> was 99.7 ug/dl (normal 108-6-7), 17-hydroxypregnenolone 129 ng/dl (normal 229-3104), and total testosterone 55 ng/dl (normal 7-20). Lymphocyte karyotype from a peripheral blood sample showed complex mosaicism: 45, X[20]/46,X,psu idic(Y)(q11.222)[1]/47, X ,psu idic(Y)(q11.222)X2[1], ish psu idic(Y)(q11.222)SRY++nuc ish(DXZ1X1)[84]/DXZ1x1,SRYx4[13]/(DXZ1X1,SRYx2)[1]

The infant with unanticipated ambiguous genitalia whom we present illustrates some of the limitations in using cell-free DNA for prenatal diagnostic studies. It is possible that the combination of low-level mosaicism for the abnormal Y chromosome and the relatively low amount of fetal DNA in the cell-free DNA pool resulted in a “falsely normal” result. It is important that patients and providers be aware of the potential limitations of this methodology.

## PgmNr 550: Diagnosis of fetal structural abnormalities using whole exome sequencing: A single centre study.

### Authors:

A. Haworth <sup>1</sup>; Y. Patel <sup>1</sup>; R. Dubis <sup>1</sup>; J. Rizvi <sup>1</sup>; J. Anton <sup>1</sup>; V. Ramachandran <sup>1</sup>; L. Reed <sup>1</sup>; E. Serra <sup>1</sup>; L. Ives <sup>1</sup>; H. Savage <sup>1</sup>; J. Short <sup>2</sup>; C. Crosby <sup>2</sup>; N. Lench <sup>1</sup>; B. Thilaganathan <sup>3</sup>; S. Mansour <sup>2</sup>; E. Dempsey <sup>2</sup>; T. Homfray <sup>2</sup>; S. Drury <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Congenica Limited, Biodata Innovation Centre, Wellcome Genome Campus, Hinxton, CB10 1DR, UK; 2) South West Thames Regional Genetics, St George's Hospital, London, SW17 0QT; 3) Fetal Medicine Unit, St George's Hospital, London, SW17 0QT

### Introduction

Ultrasound screening (US) in pregnancy is recommended by the WHO and is undertaken to monitor maternal and fetal health; approximately 2-5% of pregnancies display anomalies of fetal development. Current testing strategies demonstrate a genetic aetiology in around 40% of cases, leaving the majority of cases without a diagnosis.

Accurate diagnosis in the prenatal period is valuable; it provides information for prenatal, perinatal and postnatal management and treatment and can enable accurate assessment of the recurrence risk for future pregnancies. Provision of an additional comprehensive test would likely extend this benefit to a significant portion of the 60% of cases currently undiagnosed by conventional methods.

Recent published evidence has demonstrated the utility of exome sequencing for improved diagnosis in fetuses with structural anomalies; a recent review by Best et al (2018) showed reported diagnostic rates of between 6.2-80%.

### Objective

To provide a molecular diagnosis for cases of unexplained fetal US anomalies using whole exome sequencing, with phenotypes derived from ultrasound and/or post-mortem investigation.

### Methods

Fetuses with an abnormality detected on US were referred to Clinical Genetics at St George's Hospital, London. Fetal DNA was obtained from CVS, AF, fetal blood or post-mortem tissue and maternal cell contamination was ruled out where relevant. DNA was enriched using Agilent SureSelect CREv2 and sequenced using Illumina NextSeq 500. Secondary and tertiary analysis of sequencing data and variant review was performed using the Congenica clinical decision support platform.

### Results

Whole exome sequencing established a diagnosis in 24/62 cases (39%); phenotypes with causal variants included skeletal dysplasia (33.3%), fetal hydrops (20.8%), brain structural abnormalities (16.7%), cardiac structural abnormalities (12.5%) and multisystem anomalies (16.7%).

Updated figures will be presented and variants of uncertain clinical significance (VUS) of note in the *ALPL*, *KCNJ2* & *KMT2D* genes will be discussed.

### Conclusions

The diagnostic yield in this selected case series is higher than published unselected cohorts.

Molecular diagnosis was not always concordant with the primary clinical diagnosis highlighting both the challenge of reduced phenotype detail in this cohort and the importance of expanding prenatal genotype-phenotype correlations. Positive diagnostic rate was higher in cases with fetal hydrops

## **PgmNr 551: When to test fetuses for Noonan spectrum disorders? Proposition from a systematic analysis of 260 cases.**

### **Authors:**

A. Scott; M.-A. Delrue

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medical Genetics, CHU Sainte-Justine, Montréal, Quebec, Canada

---

Noonan spectrum disorders (NSD) are common genetic syndromes encountered at pre and postnatal evaluations, occurring in 1:1000-2500 live births. In recent years, studies have identified some evocative ultrasound findings for these syndromes, e.g. lymphatic dysplasia (increased nuchal translucency (NT), increased nuchal fold (NF), cystic hygroma (CH), hydrops, effusions, ...); congenital heart disease (valvular dysplasia or hypertrophic cardiomyopathy (HCM)); hydramnios and renal anomaly. The aim of this study was to establish which of these US findings are the most suggestive of NSD in an effort to better define the prenatal indications for NSD testing.

Retrospective review of 260 charts from 2012 to 2018 and review of literature. Inclusion criteria: (1) chromosomal anomalies excluded (by array-CGH), (2) prenatal cases with NSD evocative US findings (n=206) or pediatric cases with clinical suspicion of NSD (n=54), and (3) molecular genetic testing for NSD performed. Most of the patients had at least 11 NSD genes tested: BRAF, HRAS, KRAS, MAP2K1, MAP2K2, NRAS, PTPN11, RAF1, RIT1, SHOC2, SOS1. Prenatal ultrasound findings in positive and negative cases were compared. A stratified analysis of some ultrasound features was conducted.

A pathogenic or likely pathogenic variant was identified in 21% (55/260) of cases. 17/141 cases of CH and increased NT (12%) had a disease-causing variant. The rate of detecting a NSD variant was 40% for thoracic effusions, 28% for hydrops, 28% for congenital heart disease and 27% for increased NF. As for the stratified analysis, the detection rates were significantly improved in cases of persistent CH (16%), NT > 6 mm (23%), when CH was associated with another evocative US finding (21%) or when a cardiac defect was combined with HCM (100%) and in cases of persistent hydrops (36%). There was also improvement of the detection rate when US findings were found in combination with hydramnios (49%), renal anomaly (36%) or macrosomia (54%). 45/151 cases with two or more US findings (30%) had a NSD diagnosis vs 9% for cases with only one US finding.

**Conclusion:** after normal array-CGH, Noonan spectrum disorders should be considered when any prenatal finding of lymphatic dysplasia or suggestive congenital heart disease is found alone OR in association. Polyhydramnios, renal anomalies and macrosomia are frequent, significantly associated findings of NSD. Only severe prenatal ultrasound findings predict an unfavorable postnatal outcome for NSD.

## PgmNr 552: Likely pathogenic *de novo* variants in congenital diaphragmatic hernia patients are associated with clinical outcomes.

### Authors:

L. Qiao<sup>1,2</sup>; J. Wynn<sup>1</sup>; L. Yu<sup>1</sup>; R. Hernan<sup>1</sup>; X. Zhou<sup>1,2</sup>; V. Duron<sup>3</sup>; G. Aspelund<sup>3</sup>; C. Farkouh-Karoleski<sup>1</sup>; A. Zygmunt<sup>1</sup>; U. Krishnan<sup>1</sup>; S. Nees<sup>1</sup>; F. Lim<sup>4</sup>; T. Crombleholme<sup>4</sup>; R. Cusick<sup>5</sup>; K. Azarow<sup>6</sup>; M. Danko<sup>7</sup>; D. Chung<sup>7</sup>; B. Warner<sup>8</sup>; G. Mychaliska<sup>9</sup>; D. Potoka<sup>10</sup>; A. Wagner<sup>11</sup>; S. Soffer<sup>12</sup>; D. Schindel<sup>13</sup>; D. McCulley<sup>14</sup>; Y. Shen<sup>2,15,16</sup>; W. Chung<sup>1,17,18</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Columbia University Irving Medical Center, New York, NY, USA.; 2) Department of Systems Biology, Columbia University Irving Medical Center, New York, NY, USA.; 3) Department of Surgery, Columbia University Irving Medical Center, New York, NY, USA.; 4) Cincinnati Children's Hospital, Cincinnati, OH, USA.; 5) Children's Hospital & Medical Center of Omaha, University of Nebraska College of Medicine, Omaha, NE, USA.; 6) Department of Surgery, Oregon Health & Science University, Portland, OR, USA.; 7) Monroe Carell Jr. Children's Hospital, Vanderbilt University Medical Center, Nashville, TN, USA.; 8) Washington University, St. Louis Children's Hospital, St. Louis, MO, USA.; 9) Division of Pediatric Surgery, Department of Surgery, University of Michigan School of Medicine, Ann Arbor, MI, USA.; 10) Division of Pediatric Surgery, Department of Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.; 11) Medical College of Wisconsin, Milwaukee, WI, USA.; 12) Department of Surgery, Northwell Health, Great Neck, NY, USA.; 13) Children's Hospital Dallas, Dallas, TX, USA.; 14) Department of Pediatrics, Meriter-Unity Point Health, Madison, WI, USA.; 15) Department of Biomedical Informatics, Columbia University Irving Medical Center, New York, NY, USA.; 16) JP Sulzberger Columbia Genome Center, Columbia University Irving Medical Center, New York, NY, USA.; 17) Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA.; 18) Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA.

---

Congenital diaphragmatic hernia (CDH) is a major congenital abnormality that is associated with significant mortality, long term morbidity, and neurodevelopmental impairment in some but not all individuals. To identify classes of genetic variants associated with an increased risk of severe clinical or developmental outcomes of CDH patients, we prospectively enrolled neonates with CDH and followed them for up to five years. We performed whole genome or exome sequencing in 463 parent offspring trios to identify *de novo* single-nucleotide variants (SNVs), indels and copy number variants (CNVs). We classified CDH patients into two groups: (a) cases with likely pathogenic (LP) variants, including *de novo* likely gene disrupting (LGD) variants or predicted deleterious missense variants in constrained genes with high diaphragmic expression or known risk genes of CDH or common comorbid disorders, deletions in constrained genes, or CNVs implicated in known syndromes (LP-group, n=100); (b) other cases (nonLP-group, n=363). We assessed seven outcomes: pulmonary hypertension at one month (n=335) and three months of age (n=269), mortality prior to initial discharge (n=461), Bayley Scales of Infant Development (BSID-III, n=190) and Vineland Adaptive Behavior Scales (VABS-II, n=240) at two years of age, Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV, n=70) and VABS-II (n=88) at five years of age. LP-group patients have significantly lower developmental

scores than non-LP group at age of two (motor:  $p=4\times 10^{-5}$ ; cognitive:  $p=1\times 10^{-4}$ ), with a similar trend in complex CDH (with additional anomalies,  $n=69$ ) and isolated ( $n=123$ ) cases. LP-group also have a trend of higher mortality rate prior to initial discharge (29.3%,  $p=7\times 10^{-3}$ ) and severe pulmonary hypertension rate at 1 month (67.7%,  $p=1\times 10^{-2}$ ) than non-LP group. Overall, complex cases in LP-group had the worst survival rate and two-year developmental outcome. Isolated cases in nonLP-group had developmental outcome similar to population average and the best survival rate, and isolated cases in LP-group and complex cases in non-LP group had similar intermediate survival rate and developmental outcome. In summary, we found *de novo* likely pathogenic variants and CNVs to be associated with worse survival, more severe pulmonary hypertension, and worse developmental outcomes. These results have important implications for counseling about prognosis, potential intervention, and long term follow up for children with CDH.

## PgmNr 553: Homozygous mutations in *SPEF2* induce multiple morphological abnormalities of the sperm flagella and male infertility.

### Authors:

C. Liu <sup>1</sup>; X. He <sup>2</sup>; P. Ray <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Fudan university, shanghai, China; 2) Anhui Medical University, Hefei, China; 3) Université Grenoble Alpes, Grenoble, France.; 4) UM GI-DPI, Grenoble, France.

---

Infertility has become one of the major issues of human health, affecting 10% –15% of the general population worldwide. Male factors are identified in approximately half of the infertile couple, and the diagnosis and treatment of male infertility therefore elicit an increasing interest. Male infertility due to multiple morphological abnormalities of the sperm flagella (MMAF) is a genetically heterogeneous disorder. Previous studies revealed several MMAF-associated genes (*DNAH1*, *CFAP43* or *CFAP44*, *CFAP69*, et al), which account for approximately 60% of human MMAF cases. The pathogenic mechanisms of MMAF remain to be illuminated.

In our study, whole-exome sequencing was conducted for genetic analyses of a cohort of 65 Han Chinese men with MMAF. Two homozygous stop-gain variants c.910C>t (p.Arg304\*) and c.3400delA (p.Ile1134Serfs\*13) of the *SPEF2* (sperm flagellar 2) gene were identified in two unrelated consanguineous families. Consistently, an Iranian subject from another cohort also carried a homozygous *SPEF2* stop-gain variant c.3240delT (p. Phe1080Leufs\*2). All of the *SPEF2* variants found in three unrelated individuals were absent or extremely rare from control sequence databases (1000 Genomes Project, ExAC and gnomAD). Previous animal studies reported spontaneous mutations of *SPEF2* causing sperm tail defects in bulls and pigs. Consistently, our further functional studies using immunofluorescence assays showed the absence or a remarkably reduced staining of *SPEF2* and of the MMAF-associated CFAP69 protein in the spermatozoa from *SPEF2*-affected subjects. Together, our findings revealed a new human MMAF-associated gene, which will be informative in the molecular diagnosis and genetic counselling of infertile men with oligoasthenoteratozoospermia.



## PgmNr 554: Carrier screening for spinal muscular atrophy in 10170 pregnant women in China.

### Authors:

J. Tang <sup>1,2</sup>; Q.L. Li <sup>2</sup>; W. Ma <sup>1</sup>; Y.R. Chen <sup>2</sup>; H. Tang <sup>1</sup>; X. Li <sup>3</sup>; L.P. Luo <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Imaging Center, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China; 2) Medical Genetics Center, Jiangmen Maternity and Child Health Care Hospital, Jiangmen, Guangdong, China;; 3) Westlake University, Hangzhou, Zhejiang, China

---

**Background** Spinal muscular atrophy (SMA) is a genetic neuromuscular autosomal recessive disease characterized by muscle atrophy and weakness. SMA is one of the most common rare diseases. One in 6,000 to one in 10,000 children are born with the disease. One in 25 to one in 50 people are carriers of the SMA gene, so American College of Medical Genetics has recently recommended routine carrier screening for SMA. Large studies are needed in China to determine the benefits and costs of SMA carrier screening.

**Objective** To screen the carriers of spinal muscular atrophy (SMA) among pregnant women in Jiangmen area of Guangdong Province China, and to provide information for genetic counseling and further prenatal diagnosis, in order to reduce SMA newborns.

**Methods** From January 2017 to February 2019, this a prospective population-based cohort study of 10170 pregnant women from Jiang Men city in China. Exons 7 and 8 of SMN1 were amplified by real-time multiplex fluorescence quantitative PCR, and the copy number was quantitatively detected. Multiplexed-dependent probe amplification (MLPA) was used to verify the abnormal results of SMN1 gene screening. If the pregnant women are heterozygous for SMA, the husband is requested test. If both partners were SMA carriers, prenatal diagnostic testing is necessary performed.

**Results** Among 10170 pregnant women, 165 were carriers of heterozygosity deletion in exon 7 and 8 of SMN1 gene, 6 were carriers of heterozygosity deletion in exon 7 of SMN1 gene, and 55 were carriers of heterozygosity deletion in exon 8 of SMN1 gene. The carrier rate was approximately 1/59 (1.69%). 3 couples were determined to be at high risk for having offspring with SMA and prenatal diagnostic test. Prenatal diagnostic result show 3 fetuses are normal.

**Conclusion** Carrier screening and carrier rate determination are of great clinical significance for follow-up genetic counseling, disease prevention and reduction of birth defects. SMA carrier screening has reduced the burden of SMA newborns.

## PgmNr 555: A late-onset Alzheimer's disease associated gene *TM2D3* is required for neural maintenance and developmental Notch signaling in *Drosophila*.

### Authors:

J. Salazar<sup>1</sup>; D. Li-Kroeger<sup>1</sup>; P. Marcogliese<sup>1</sup>; S. Yang<sup>1</sup>; S. Yamamoto<sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 2) Dept. of Neuroscience, Baylor College of Medicine, Houston, TX.; 3) Program in Developmental Biology, Baylor College of Medicine, Houston, TX.; 4) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX.

---

Alzheimer's disease (AD) is the most prevalent neurodegenerative disorder in the aging human population. Mutations in the genes *APP*, *PSEN1*, and *PSEN2* have been found to cause familial, early-onset forms of the disease; however, about 95% of AD cases are late-onset (after age of 65) and of sporadic nature and the genetic factors that may contribute are steadily being uncovered. Large human genetics studies in the past decades have identified several loci associated with risk of late-onset Alzheimer's disease (LOAD). However, the molecular functions of many of the genes within these loci remain to be elucidated. To develop therapeutics for AD it is crucial to understand the molecular players that contribute to AD pathogenesis.

Here, we describe our efforts to functionally characterize *TM2D3*, a gene previously associated with increased risk of LOAD (7.45 OR). Using the highly tractable model organism *Drosophila melanogaster*, we have generated a null allele of the fruit fly ortholog of *TM2D3*, *almondex* (*amx*). Importantly, we have found that *amx* mutants are short lived and exhibit climbing deficits, indicative of neuronal dysfunction. Functional studies in the *Drosophila* giant fiber system and the use of electroretinogram (ERG) have uncovered defects in proper neuron function in *amx* mutants. Interestingly, introduction of wild-type human *TM2D3* in *amx* mutants suppresses lifespan and climbing defects, indicating that the human protein can functionally replace Amx and function *in vivo* in *Drosophila*. This gives us a platform to probe the molecular function of *TM2D3*, as well as the effect of the LOAD linked variant of *TM2D3* (p.P155L).

Additionally, embryos from *amx* mutant females exhibit a neurogenic phenotype, indicative of loss of proper Notch signaling during embryonic development. Past epistasis experiments have suggested that *amx* is genetically involved in the proteolytic processing of the Notch receptor. Since the metalloprotease ADAM10 and the intramembrane protease complex γ-secretase are both involved in the processing of Notch receptors and APP, we hypothesize that *TM2D3* (and Amx) acts as a regulator of ADAM10 and/or γ-secretase. Taking advantage of genetic epistasis experiments that are possible in *Drosophila*, we aim to precisely determine at what step *TM2D3* acts genetically to understand its molecular function in the context of LOAD.

## PgmNr 556: Exome sequencing in 400 individuals with malformations of cortical development reveals wide genetic heterogeneity.

### Authors:

U. Kini <sup>1</sup>; A. Pagnamenta <sup>2</sup>; M. Hashim <sup>2</sup>; R. Tripathy <sup>3</sup>; D. Keays <sup>3</sup>; J. Taylor <sup>2</sup>; DDD. Study <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oxford Centre for Genomic Medicine, Oxford University Hospitals NHS FT, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford; 3) Institute for Molecular Pathology Campus Vienna Biocenter 1 Vienna; 4) Sanger Centre, Cambridge, UK

---

Malformations of cortical development (MCD) are congenital anomalies that result from abnormal proliferation, migration and organisation of fetal neurones during brain development causing brain anomalies such as primary microcephaly, lissencephaly, periventricular nodular heterotopia (PVNH), cortical dysplasia and polymicrogyria (PMG). Over the past 2 decades a large number of novel genetic causes have been described. **Aims:** To study the spectrum of genetic pathways involved in cortical development and examine genetic heterogeneity in a cohort of 400 individuals with MCD. **Methods:** Patients with MCD who had undergone exome sequencing were identified from 2 sources: (1) DDD (Deciphering Developmental Disorders) study: Exome sequencing was carried out on 13,451 individuals with suspected rare genetic disorders. A comprehensive list of 21 HPO (human phenotype ontology) terms related to MCD were used to identify a subset of 294 patients and their exome data was accessed through a filtered list of candidate variants provided by DDD (2) Genetics of Structural Brain Abnormalities and learning difficulties (GOSBA) study: Trio exome sequencing data from 106 probands were examined for known and novel genetic variants. The phenotype was assessed in detail via a bespoke clinical questionnaire, review of MRI brain scans and neuropathological findings and studying facial dysmorphism. **Results:** 68 individuals with MCD (68/400 or 17%) were found to have a reportable genetic variant in a gene previously reported to cause a brain phenotype (n=39 genes); the most common being *DYNC1H1* (n=8 affected individuals). An additional 17 patients had variants in novel genes (n=9 genes, published) causing a specific brain phenotype - *PGAP3*, *PIGY*, *DENR*, *SOX11*, *ERMARD*, *PI4KA*, *HNRNPK*, *NEDD4L* and *CCND2*, taking the diagnostic yield up to ~27%. Putative candidate variants that need further exploration have been identified in an additional 150 patients. Cellular genetic pathways such as microtubule, GPI anchor and *AKT-PI3KA* pathways were identified as critical for cortical development. **Conclusion:** Our study demonstrates a wide genetic heterogeneity in MCD - 48 different genes accounted for the MCD in 85 patients. A large proportion of MCD cases (~70 %) therefore remain unsolved despite using newer testing technology such as exome sequencing. Due to the wide genetic and phenotypic heterogeneity reporting of novel genes can be slow and labour-intensive.

## PgmNr 557: Mutations in *RFX* family members are a common cause of autism and ADHD.

### Authors:

T. Nakayama<sup>1,2,3</sup>; H.K. Hodges<sup>3,4</sup>; N. Argyrou<sup>1,2,3</sup>; C. Gubbels<sup>1,2,3</sup>; A. Soucy<sup>1,2,3</sup>; C. Genetti<sup>1,2,3</sup>; L. Rodan<sup>1,2,3</sup>; R. Pfundt<sup>5</sup>; T. Gardeitchik<sup>5</sup>; B.B.A. de Vries<sup>5</sup>; M. Swinkels<sup>6</sup>; M.E.H. Simon<sup>6</sup>; K. van Gassen<sup>6</sup>; C.M.L. Volker-Touw<sup>6</sup>; A. Hamosh<sup>7</sup>; C.D. Applegate<sup>7</sup>; L.L. Imken<sup>8</sup>; M. Willing<sup>9</sup>; E. Fassi<sup>9</sup>; T.L. Toler<sup>9</sup>; K.W. Gripp<sup>10</sup>; L. Baker<sup>10</sup>; F. Vansenne<sup>11</sup>; P. Rump<sup>11</sup>; X. Wang<sup>12</sup>; J. Ambrus<sup>13</sup>; M. Fannemel<sup>14</sup>; G.E. Tiller<sup>15</sup>; E. Agolini<sup>16</sup>; K. McWalter<sup>17</sup>; R. Asadollahi<sup>18</sup>; A. Rauch<sup>18</sup>; P. Boonsawat<sup>18</sup>; C. Fagerberg<sup>19</sup>; G. Lesca<sup>20</sup>; A. Labalme<sup>20</sup>; L.E. Walsh<sup>21</sup>; K.K. Payne<sup>21</sup>; B. Lianoglou<sup>22</sup>; K. Aldinger<sup>23</sup>; J. Balciuniene<sup>24</sup>; J. Murrell<sup>24</sup>; C. Bupp<sup>25</sup>; G. Pascolini<sup>26</sup>; S. Bakhtiari<sup>27</sup>; M.C. Kruer<sup>27</sup>; J. Gecz<sup>28</sup>; K. Harper<sup>28</sup>; Z. Powis<sup>29</sup>; P. Agrawal<sup>1,2,3</sup>; A. Beggs<sup>1,2,3</sup>; T.W. Yu<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Developmental Medicine, Department of Medicine, Harvard Medical School, Boston Children's Hospital, Boston, MA; 2) Division of Genetics and Genomics, Department of Medicine, Boston Children's Hospital, Boston, MA; 3) Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 4) Department of Pediatrics, Harvard Medical School, Boston, MA; 5) Department of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands; 6) Division of Biomedical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 7) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 8) Department of Medical Genetics, Dell Children's Hospital, Austin, TX; 9) Division of Genetics and Genomic Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO; 10) Division of Medical Genetics, Nemours/A.I. DuPont Hospital for Children, Wilmington, DE; 11) Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands; 12) CIPHERgene, Beijing, China; 13) Division of Allergy, Immunology, and Rheumatology, SUNY at Buffalo School of Medicine, Buffalo, NY; 14) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 15) Department of Genetics, Kaiser Permanente, Los Angeles, CA; 16) Department of Medical Genetics, Bambino Gesù Children's Hospital, Rome, Italy; 17) GeneDx, Gaithersburg, MD; 18) These authors contributed equally to this work; 19) Department of Clinical Genetics, Odense University Hospital, Odense, Denmark; 20) Department of Genetics, Civil Hospices of Lyon, Bron, France; 21) Indiana University Health Neuroscience Center, Indianapolis, IN; 22) University of California at San Francisco Fetal Treatment Center, San Francisco, CA; 23) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 24) Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, PA; 25) Spectrum Health Medical Group, Grand Rapids, MI; 26) Laboratory of Medical Genetics, Department of Molecular Medicine, Sapienza University, San Camillo-Forlanini Hospital, Rome, Italy; 27) Barrow Neurological Institute, Phoenix Children's Hospital, Phoenix, AZ; 28) Robinson Research Institute, The University of Adelaide, Adelaide, Australia; 29) Ambry Genetics, Aliso Viejo, CA

---

Regulatory Factor X (RFXs) proteins are a family of developmentally important transcription factors that share highly-conserved DNA binding domains. Knockout studies of *Rfx2*, *Rfx3*, *Rfx4*, and *Rfx7* in mice and *Xenopus* demonstrate important roles for these genes in embryogenesis, central nervous system development and ciliogenesis. Expression patterns of human *RFX2*, *RFX3*, *RFX4* and *RFX7* show strong tissue specificity to human brain, and are under marked haploinsufficient constraint in human populations. We report a total of 31 patients (28 probands) with *de novo* or inherited

heterozygous mutations in *RFX* family members in association with autistic features and other neurobehavioral abnormalities: *RFX3* mutations (17 patients), *RFX2* (2), *RFX4* (3), and *RFX7* (9). These cases share in common neurobehavioral features including autism spectrum disorder (ASD), ADHD, and mild intellectual disability. Other frequent features included hypotonia, subtle dysmorphisms, and sleep disorder. Functional analysis of the spectrum of mutations encountered is consistent with a loss-of-function mechanism. Collectively, *de novo* mutations in *RFX* family members may comprise one of the more common genetically identifiable causes of ASD, and point to a network of brain-expressed targets that illuminate critical disease-associated neurobiological pathways for this disorder.

## PgmNr 558: Associated anomalies in cases with agenesis of the corpus callosum.

### Authors:

C. Stoll; Y. Alembik; B. Dott; M.P. Roth

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetique Medicale, Faculte de Medecine, Strasbourg, France

---

Agenesis of corpus callosum (ACC) is an uncommon congenital anomaly, its etiology is unclear and its pathogenesis is controversial. Cases with ACC often have other non-ACC associated congenital anomalies. The purpose of this study was to assess the prevalence and the types of these associated anomalies in a defined population. The associated anomalies in cases with ACC were collected in all livebirths, stillbirths and terminations of pregnancy during 29 years in 387,067 consecutive births in the area covered by our population-based registry of congenital malformations. Of the 99 cases with ACC, representing a prevalence of 2.56 per 10,000, 73 (73.7%) had associated anomalies. There were 16 (16.2%) cases with chromosomal abnormalities, and 13 (13.2%) nonchromosomal recognized dysmorphic conditions including syndromes 2 each: Aicardi, Dandy-Walker, and fetal alcoholism. Forty four (44.4%) of the cases had non syndromic multiple congenital anomalies (MCA). Anomalies in the musculoskeletal, the urogenital, the central nervous, the cardiovascular, and the digestive systems were the most common other anomalies in the cases with MCA. The anomalies associated with ACC could be classified into a recognizable malformation syndrome in 29 out of the 73 cases (39.7 %) with associated anomalies. This study included special strengths: it is population-based, each affected child was examined by a geneticist, all elective terminations were ascertained, and the surveillance for anomalies was continued until 2 years of age. In conclusion the overall prevalence of associated anomalies, three of four cases, emphasizes the need for a screening for other anomalies in cases with ACC.

## PgmNr 559: The nuclear receptor co-regulator RERE is required for development of the cerebellar vermis and midbrain.

### Authors:

B. Kim <sup>1</sup>; H. Zaveri <sup>1</sup>; D.A. Scott <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular & Human Genetics, Baylor College Medicine, Houston, Texas.; 2) Molecular Physiology and Biophysics, Baylor College Medicine, Houston, Texas.

---

Deletions of chromosome 1p36 are the most common terminal deletions in humans and associated with high incidence of the brain defects. The arginine-glutamic acid dipeptide repeats gene (RERE) is located in a proximal critical region on chromosome 1p36. Pathogenic mutations in *RERE* have been identified as causative variants for defects seen in individuals with proximal 1p36 deletions. Cerebellar vermis anomalies are documented in 50% of individuals with these mutations in *RERE* and *RERE*-deficient mice show incomplete foliation in the vermis. Although *RERE* play a critical role in the vermis development, morphogenic and molecular mechanisms by which tissue specific depletion of *RERE* causes in the vermis are not clearly described in early embryonic cerebellum. The development of midbrain and cerebellum is initiated by the isthmus organizer (IO) located in the midbrain-hindbrain junction at E9.5. A bilateral, wing-like shaped cerebellar primordium transforms into cylinder-shaped morphology along the medial-lateral axis between E12 and E15.5. In this study, floxed *Rere* allele is excised in the midbrain-hindbrain junction by Pax2-Cre recombinase. The bilateral wings of the cerebellum are fused in medial region in control embryos at E13.5. In contrast, midline fusion of these wings is failed in Pax2-Cre;*Rere*<sup>F/F</sup> embryos. Even though medial region of Pax2-Cre;*Rere*<sup>F/F</sup> embryo cerebellum is fused at E15.5, development of midline fusion is limited in comparison with control embryos. At E18.5, four fissures are not formed in the vermis primordium of Pax2-Cre;*Rere*<sup>F/F</sup> embryos while these fissures appear in control embryos. In addition, development of the inferior colliculus in the midbrain is attenuated in mutant embryos while the inferior colliculus is completely developed in control embryos. We also showed that expression of *Wnt1*, *Fgf8*, *En1* and *En2*-which are known to be key regulators for development of the midbrain and the cerebellum- is downregulated in the mutant embryo cerebellums in comparison with that of the control embryos at E14.5. Taken together, these results suggest that tissue-specific ablation of *Rere* is enough to lead abnormal development of the vermis and *Rere* regulates expression of genes that are important for development of the midbrain and the cerebellum during the early embryonic period.

## PgmNr 560: Genetic and clinical spectrum of neurodevelopmental disorder with microcephaly in a Chinese cohort.

### Authors:

H. Wang; L. Shu; X. Mao

[View Session](#) [Add to Schedule](#)

**Affiliation:** Hunan Provincial Maternal and Child Health Care Hospital, Changsha, China

---

**Background:** Microcephaly is a complex and multifactorial birth defect with occipital frontal circumference (OFC) of  $>2$  SDs below the population mean. There is a growing number of researchers focusing of the genes linked to the disease. To date, almost 800 genes related to microcephaly have been identified including 18 pathogenic genes of microcephaly primary hereditary (MCPH). Only 3 cohorts using exome sequencing to identify genes linked to microcephaly.

**Methods:** We included 80 patients with microcephaly. CNV-seq and whole-exome sequencing (WES) were performed on the patients. Relevant clinical profiles of each patient was collected by experienced neurologists.

**Results:** We identified 8 CNVs and 34 genes related to microcephaly. The top 6 common genes in our cohort was *DYRK1A* (5 patients), *ASPM* (4 patients), *TUBA1A* (3 patients) and *CASK*, *FOXP1*, *PNKP* (2 patients). Other genes were found in only one patient, respectively. Among the 34 genes, there were 26 reported genes for microcephaly (76.5%), 3 unreported genes for microcephaly (*ARF1*, *MPDZ*, *ZNF462*; 8.8%) and 5 candidate genes for microcephaly (*SMG6*, *BTBD9*, *FOCAD*, *TMEM147*, *THSD7A*; 14.7%). These patients presented with clinical features such as developmental delay, seizures, dysmorphic feature.

**Conclusion:** Our research included the largest cohort with developmental disorder and microcephaly and clarified the genetic and clinical spectrum of the disease to date. We also identified genes with novel clinical phenotypes as microcephaly and several novel candidate genes. Our research contributed to the understanding of genetic etiology of microcephaly and provided clues for genetic counseling.



## PgmNr 561: Investigating evolving interactions between Zika virus and host pathways using *Drosophila*.

### Authors:

J. Harnish <sup>1</sup>; N. Link <sup>1</sup>; B. Hull <sup>1,2</sup>; S. Balcazar <sup>3</sup>; P. Shah <sup>4</sup>; H. Bellen <sup>1,5,6,7,8</sup>; S. Yamamoto <sup>1,5,6,7,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Post-baccalaureate Research Education Program (PREP), Baylor College of Medicine, Houston, Texas.; 3) Summer Undergraduate Research Training Program, Baylor College of Medicine, Houston, Texas.; 4) Department of Chemical Engineering, University of California Davis; 5) Department of Neuroscience, Baylor College of Medicine, Houston, Texas.; 6) Program in Developmental Biology, Baylor College of Medicine, Houston, Texas.; 7) Howard Hughes Medical Institute; 8) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas.

---

Spread of Zika virus (ZIKV) is linked to an increase in neurological diseases including Guillan-Barré syndrome and congenital microcephaly, yet the underlying molecular mechanisms are mostly unknown. Recent work suggests increased pathogenicity of the Puerto Rican ZIKV strain is associated with missense mutations that occurred in pre-Membrane protein (prM) and NonStructural protein 1 (NS1). These changes make Puerto Rican ZIKV strains distinct from those isolated in Central Africa and Southeast Asia and may contribute to worsening outcomes of ZIKV infection. To better understand how ZIKV causes neurological sequelae in humans, and why Puerto Rican strains appear more pathogenic, we are using the fruit fly *Drosophila melanogaster* to assess the interactions between ZIKV proteins and host cellular pathways. We have also investigated if more ancient versions of ZIKV proteins elicit the same effects on these pathways as the modern Puerto Rican strain. We generated transgenic flies that allow expression of individual ZIKV proteins under the control of a UAS element to determine the effect of ZIKV protein expression in diverse tissues *in vivo*. Interestingly, using the sequences from the Puerto Rican strain, we found that most factors exhibited a scorable phenotype when expressed in the nervous system or developing tissues. Some phenotypes such as reduced brain size and sensory bristle loss suggest that multiple ZIKV proteins can modulate neurodevelopment. We have also generated transgenic flies expressing the more ancient and less pathogenic Cambodian strain ZIKV proteins. Interestingly we found the Cambodian strain NS1 protein does not display the phenotypes caused by the Puerto Rican NS1 suggesting that a single missense mutation in NS1 may contribute to the worsening outcomes of ZIKV infection in the Puerto Rican strain. This study demonstrates the value of *Drosophila* in the functional characterization of viral factors and has allowed us to begin to unravel the mechanism of ZIKV mediated neurological disease. It has also given early insight into the evolution of ZIKV pathogenicity and identified possibly causative missense mutations explaining worsening outcomes of ZIKV infection in South America.

## PgmNr 562: Preimplantation genetic testing for Schaaf-Yang syndrome using karyomapping.

### Authors:

J. Zhang<sup>1,2,3</sup>; S. Chen<sup>1,2,3</sup>; Y. Chen<sup>1,2,3</sup>; H. Fei<sup>1,2,3</sup>; S. Li<sup>1,2,3</sup>; C. Chang<sup>1,2,3</sup>; L. Zhang<sup>1,2,3</sup>; L. Wang<sup>1,2,3</sup>; L. Jin<sup>1,2,3</sup>; H. Huang<sup>1,2,3</sup>; C. Xu<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) International Peace Maternity and Child Health Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China; 2) Shanghai Key Laboratory of Embryo Original Diseases, Shanghai, China; 3) Institute of Embryo-Fetal Original Adult Disease Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

---

**Introduction:** Cerebral palsy (CP) is the most common severe motor disability in children. It's not a disease but a combination of signs and symptoms. Different patterns of inheritance and genetic heterogeneity are known to CP. Some neurologic and metabolic genetic disorders which mimic the clinical characteristics with CP are more likely to be misdiagnosed. Whole exome sequencing (WES) for a patient, initially diagnosed as CP, revealed a nonsense variant in the *MAGEL2* gene. *MAGEL2* is a maternally imprinted, paternal-only expressed, gene located in the Prader-Willi critical region 15q11-15q13. Loss-of-function variants in paternal allele of *MAGEL2* has been described in Schaaf-Yang syndrome (SHFYNG), which is a neurodevelopmental disorder shared partial phenotypes overlap with Prader-Willi Syndrome, including neonatal hypotonia, feeding problems, and developmental delay/ intellectual disability. Male individual with *MAGEL2* pathogenic variant is at 50% risk for producing an affected child. Here we report the first successful preimplantation genetic testing for monogenic defects (PGT-M) for SHFYNG using Karyomapping.

**Materials and Methods:** A Chinese family with 5 patients initially misdiagnosed as CP was recruited in our Reproductive Genetic Clinic. Four of the patients died after birth due to atelectasis. WES was used to identify pathogenic variant of the only survive 11-year-old proband. Karyomapping, a method employing genome-wide analysis of SNP, was used as PGT-M protocol. Analysis of DNA from the proband and his parent revealed informative SNPs linked with the *MAGEL2* pathogenic variant. These SNPs were subsequently used for embryo diagnosis.

**Results:** The WES for the proband, misdiagnosed and treated as CP, revealed a novel nonsense pathogenic variant of c.2895 G>A (p.Trp965\*) in the *MAGEL2* gene. The variant was inherited from his unaffected father and paternal grandmother. Five embryos underwent trophectoderm biopsy followed by vitrification in the second cycle. Through Karyomapping analysis, 2 embryos without inheritance of the *MAGEL2* pathogenic variant haplotype were diagnosed as unaffected. One unaffected embryo was transferred, with one healthy baby was born at 38 gestational weeks.

**Conclusions:** This is the first report of Karyomapping-based PGT-M for a SHFYNG family initially misdiagnosed as CP. Karyomapping could be used to identify embryos with monogenic defects.

## PgmNr 563: Hippocampal transcriptome changes associated with long-term cognitive effects of gestational air pollution toxicity in mouse.

### Authors:

A. Haghani <sup>1</sup>; R.G. Johnson <sup>1</sup>; N.C. Woodward <sup>1</sup>; J.I. Feinberg <sup>2,3</sup>; K.C. Lewis <sup>4</sup>; C.H. Ladd-Acosta <sup>3,5</sup>; N. Safi <sup>6</sup>; A.E. Jaffe <sup>7</sup>; C. Sioutas <sup>8</sup>; C.E. Finch <sup>1</sup>; H. Allayee <sup>9</sup>; D.B. Campbell <sup>4</sup>; H.E. Volk <sup>3,5,10</sup>; T.E. Morgan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Davis School of Gerontology, University of Southern California, Los Angeles, CA.; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.; 3) Wendy Klag Center for Autism and Developmental Disabilities, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.; 4) Department of Pediatrics and Human Development, Michigan State University, East Lansing, MI.; 5) Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.; 6) Department of Biomedical Sciences, Center for Bioinformatics and Genomics, Cedars-Sinai Medical Center, Los Angeles, CA.; 7) Lieber Institute of Brain Development, Johns Hopkins Medical Campus, Baltimore, MD.; 8) Department of Civil and Environmental Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, CA.; 9) Department of Preventive Medicine, University of Southern California, Los Angeles, CA.; 10) Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD.

---

Air pollution adversely alters the early phases of human development. Gestational air pollution exposure is associated with long-lasting neurodevelopmental effects including increased risk of autism, depression, and decreased intellectual capacity. The molecular mechanisms for this toxicity have not been examined at the transcriptomic level. In this study, pregnant dams were exposed to urban derived nanosized particulate matter (nPM). Three-month male and female offspring were studied for cognitive changes using forced swim and elevated plus behavior tests. At 4.5 months of age, hippocampal transcriptomic changes were analyzed by RNAseq to identify the genes associated with the observed behavioral changes. Male offspring showed increased immobility time in the forced swim test and increased open arm entry in the elevated plus test, which indicates increased depressive behavior and decreased anxiety, respectively. Weighted gene co-expression analysis of the hippocampal transcriptome changes identified a module of 43 genes with sex-specific gene responses to gestational nPM exposure. These genes were enriched for several nervous system pathways, including G-protein coupled receptor, GABA receptor, and Opioid signaling pathways. The identified genes were tested for association with the observed behavioral changes. The forced swim was associated with several nPM responsive genes related to synaptic long-term potentiation and CCR5 signaling pathways. The genes associated with elevated plus were enriched for immune-related pathways such as GP6 signaling and pattern recognition pathways. In conclusion, this study identified several transcriptomic profiles for further study on gestational air pollution toxicity on the brain. Understanding the causal relationship of the identified genes with depressive and anxiety-like behaviors may have important implications for therapeutic and intervention development research.

Funding: ES023780; ES022845; Cure Alzheimer's Fund

## **PgmNr 564: Evidence of pervasive anxiety-related symptoms in 47,XXY (Klinefelter syndrome) and the potential impact of hormonal replacement therapy (HRT).**

### **Authors:**

C. Samango-Sprouse <sup>1,2,4</sup>; A. Gropman <sup>3,2</sup>; P. Lasutschinkow <sup>4</sup>; T. Sadeghin <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) George Washington University School of Medicine and Health Sciences, Washington, DC; 2) Department of Human and Molecular Genetics, Florida International University, Miami, FL; 3) Children's National Medical Center, Washington, DC; 4) The Focus Foundation, Davidsonville, MD

47,XXY (Klinefelter syndrome) is the most common X&Y chromosomal variation (1:660 males). The phenotype includes accelerated height, language-based learning disorders, motor delays & behavioral problems. Past studies demonstrated hormonal replacement therapy (HRT) may improve neurodevelopment of boys with 47,XXY. The progression of anxiety related symptoms throughout childhood & the potential impact of HRT at each age has not been investigated.

Parents of 48 toddlers (group A), 44 boys between 6 & 10 years (group B) & 42 boys between 11-18 years (group C) with 47,XXY completed Child Behavior Checklists. 58.3% of A, 61.4% of B & 72.6% of C received HRT (T). The rest were untreated (U). HRT was based on an evaluation by a pediatric endocrinologist. Hotelling's T-squared distributions were completed to analyze differences between U&T groups.

In group A, 10% of non-HRT (A-U) showed symptoms above the 97<sup>th</sup> percentile (clinically elevated) in internalizing & externalizing problems. 3.57% of A-T had clinically elevated symptoms in externalizing & total problems. A-T had fewer symptoms in affective ( $P=0.017$ ) & total problems ( $P=0.03$ ) compared to A-U.

23.53% of B-U were clinically elevated in internalizing & total problems. 11.76% of B-U were clinically elevated in anxious/depressed, somatic complaints, thought, externalizing & affective problems. B-T had fewer symptoms but were clinically elevated in internalizing (14.81%), externalizing (11.11%) & total problems (11.11%). B-T had fewer symptoms in somatic complaints ( $P=0.01$ ), thought ( $P=0.07$ ) & somatic problems ( $P=0.01$ ).

C-U group was clinically elevated in internalizing (60%), anxiety (50%), & anxious/depressed (40%). 30% of C-U was clinically elevated in somatic complaints, total & affective problems. 21.88% of C-T was clinically elevated in internalizing & total problems. 12.5% were elevated in affective & anxiety problems. C-T had less symptoms in anxious/depressed ( $P=0.01$ ), thought ( $P=0.03$ ), externalizing ( $P=0.04$ ), total ( $P=0.03$ ), & affective problems ( $P=0.02$ ).

Our findings demonstrate anxiety related symptoms appear to be a penetrant aspect of the 47,XXY phenotype as early as 3 years, with severity possibly increasing with age. This research supports the potential positive effect of HRT on neurodevelopment in boys with 47,XXY with less reported

symptoms in HRT-boys at all ages. Further studies should explore the impact of HRT on specific symptoms of anxiety to develop more personalized & precise treatments.

## **PgmNr 565: Combinatorial overexpression of transcription factors *Atoh1*, *Gfi1* and *Pou4f3* enable the conversion of Greater Epithelial Ridge (GER) cells into ectopic hair cells in the neonatal mouse cochlea.**

### **Authors:**

A. Anand Iyer <sup>1</sup>; H. Jen <sup>3</sup>; R. Yousaf <sup>2</sup>; M.M. McGovern <sup>2</sup>; T. Cai <sup>3</sup>; S. Singh <sup>2</sup>; H. Zhang <sup>2</sup>; A.K. Groves <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 2) Department of Neuroscience, Baylor College of Medicine, Houston, Texas; 3) Program in Developmental Biology, Baylor College of Medicine, Houston, Texas

---

Hair cells in the inner ear enable the perception of sound through the process of mechano-transduction, rendering them indispensable for hearing. Regeneration of these hair cells occurs naturally in non-mammalian vertebrates like fish, amphibians and birds. In mammals, hair cells can regenerate at young ages to a minimal extent, in response to specific genetic interventions. One such intervention is the ectopic expression of transcription factor *Atoh1*. It is known that the expression of a transcription factor cocktail comprising *Atoh1*, *Gfi1* and *Pou4f3* can reprogram the transcriptomic signature of mouse embryonic stem cells towards that of inner ear hair cells. However, the effects of these transcription factor combinations on the underlying hair cell genetic network, and the relative responsiveness of 'non-hair cells' to such reprogramming signals remain unknown.

We targeted the *Rosa26* locus to allow conditional overexpression of *Atoh1* alone (*Rosa26-A*), *Gfi1*, *Atoh1* (*Rosa26-GA*) and *Gfi1*, *Atoh1*, *Pou4f3* (*Rosa26-GAP*) each. Mice harboring these conditional alleles were bred with *Sox9-CreER* mice to achieve overexpression in the greater epithelial ridge (GER) cells of the neonatal mouse cochlea after tamoxifen administration at birth. Immunostaining for hair cell markers showed the presence of a large number of ectopic hair cells in the GER of these cochleae in comparison to the wild type control animals, by one week of age. Further, these ectopic hair cells stained positive for Vglut3, an inner hair cell marker and not for Prestin, an outer hair cell marker.

Our next steps include the extensive characterisation of these ectopic hair cells by performing a single cell RNA seq followed by a bulk RNA seq. Through the single cell RNA seq, we aim to understand which cell population these ectopic hair cells most closely resemble. Once the cell population type is determined, we plan to perform a bulk RNA seq on this population to understand the difference in transcriptomic signatures of each conditional overexpression case with respect to the wildtype. Through this work, we aim to compare and understand the effect of additional transcription factors *Gfi1* and *Pou4f3* with *Atoh1*, in turning on specific gene networks to reprogram the GER cells in the neonatal mouse cochlea into hair cells. Understanding the principles governing mammalian hair cell regeneration will enable the achievement of efficient therapeutic strategies for treating human hearing loss in future.

## PgmNr 566: *MAB21L1* demonstrates a conserved role in vertebrate development.

### Authors:

S. Seese<sup>1,2</sup>; B. Deml<sup>1,2,3</sup>; S. Muheisen<sup>1</sup>; L. Reis<sup>1</sup>; R. Jamieson<sup>4</sup>; E. Semina<sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics and Children's Research Institute, Medical College of Wisconsin and Children's Hospital of Wisconsin, Milwaukee, Wisconsin 53226, USA; 2) Department of Cell Biology, Neurobiology & Anatomy, Medical College of Wisconsin, Milwaukee, WI 53226, USA; 3) Current Address: Prevention Genetics, Marshfield, WI 54449, USA; 4) The University of Sydney, NSW, Australia; 5) Department of Ophthalmology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

---

The male-abnormal 21 family has three members, *MAB21L1*, *MAB21L2* and *MAB21L3*, with their corresponding roles in development only beginning to be understood. *MAB21L1* shares 94% protein identity with *MAB21L2*, a protein previously shown to have a role in vertebrate eye development and human disease. Recent reports have also described homozygous *MAB21L1* mutations in humans associated with ocular abnormalities including corneal dystrophy, and systemic features such as facial dysmorphism, scrotal agenesis and cerebellar malformation. Null *Mab21l1* mice likewise display developmental eye defects affecting the optic cup and anterior chamber structures. Using zebrafish and human patient DNA, our study expands on the developmental role of *MAB21L1/mab21l1* along with the spectrum of phenotypes associated with its deficiency.

To study the role of *mab21l1* in development, we utilized the zebrafish model. TALENs were used to generate a genetic mutation, *mab21l1 c.107delA p.[K36Rfs\*7]*. This line was examined for ocular anomalies using gross imaging, histology, immunohistochemistry and electron microscopy. Homozygous adult fish 1-2 years of age display ocular abnormalities including cornea, lens and other anterior segment defects, microphthalmia, and retinal degeneration. Craniofacial anomalies and reduced survival were also noted. To determine age of onset, we collected specimens at 5-days post fertilization (dpf), 8-dpf, 10-dpf, 14-dpf, 1-month post fertilization (mpf), and 2-mpf. The phenotype is first observed 5-dpf (abnormally thin cornea) with additional progressive changes identified at later stages.

To evaluate *MAB21L1*'s contribution to human disorders, we screened DNA samples from patients with various developmental ocular conditions by Sanger sequencing of the *MAB21L1* coding region. This revealed two variants, *c.184C>T*, p.[Arg62Cys] and *c.658G>C*, p.[G220R], in two unrelated patients with colobomatous microphthalmia. The pathogenicity of these variants was unclear and functional tests performed to understand their effects, including protein stability, localization and complementation assays. While protein stability and localization are undisrupted, complementation assays using zebrafish mutants suggested mild functional deficiency. Additional studies are ongoing to identify new human variants. Studies are also in progress to elucidate the molecular function of the protein to aid in functional testing of newly identified human variants.



## PgmNr 567: Insights of the role of *FOXC1* gene in ocular development using a double *foxc1a/foxc1b* knockout zebrafish line.

### Authors:

J.J. Ferre Fernandez <sup>1</sup>; E. Sorokina <sup>1</sup>; S. Muheisen <sup>1</sup>; S. Thompson <sup>1</sup>; E.V. Semina <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics and Children's Research Institute, Medical College of Wisconsin and Children's Hospital of Wisconsin, Milwaukee, WI 53226, USA; 2) Department of Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, WI 53226, USA; 3) Department of Ophthalmology, Medical College of Wisconsin, Milwaukee, WI 53226, USA

The gene forkhead box C1 (*FOXC1*) encodes a forkhead/winged helix transcription factor involved in embryonic development. Mutations in *FOXC1* cause developmental eye disorders like Axenfeld-Rieger syndrome type 3 or anterior segment dysgenesis, usually following an autosomal dominant inheritance pattern. Some patients also present other non-ocular defects like cardiac and craniofacial abnormalities. There are two zebrafish orthologous genes, *foxc1a* and *foxc1b*. Both zebrafish genes demonstrate evolutionary conservation with the human ortholog, especially in the forkhead and activation domains. The lack of *foxc1a* in zebrafish produces defects in somitogenesis, heart and craniofacial region. In this study we use the zebrafish as an animal model to further study the role of *FOXC1* in the eye development.

Using CRISPRs we generated single *foxc1a* and *foxc1b* knockout zebrafish lines, carrying p.[D71\*] and p.[A86Dfs\*30] alleles correspondingly, and used them to obtain *foxc1a/foxc1b* double knockout embryos (D-KO). Both *foxc1* mutations are predicted to produce complete loss-of-function alleles because of an early truncation affecting the forkhead and C-terminal domains.

*foxc1a*<sup>-/-</sup> homozygous embryos showed anomalies in blood flow, microphthalmia and periocular edema, as well as embryonic lethality, while heterozygous *foxc1a*<sup>+/-</sup> and heterozygous or homozygous *foxc1b* embryos did not display any visible phenotype and survived to adulthood. To further evaluate *foxc1a/b* dosage requirements and to generate D-KO embryos, we raised adults carrying heterozygous *foxc1a* and homozygous *foxc1b* alleles. The *foxc1a*<sup>+/-</sup>;*foxc1b*<sup>-/-</sup> fish demonstrated reduced survival, craniofacial, skeletal and ocular defects, thus displaying a dominant phenotype for the *foxc1a* mutation in *foxc1b*-deficient background and exposing a role of *foxc1b* in the development/maintenance of the affected structures. Embryos with deficiency for a single (*foxc1a*) or both (*foxc1* D-KO) genes were examined by histology and showed similar ocular defects; also, analysis of eye vasculature was performed using Tg(*fli1a*-GFP)/*foxc1a*<sup>+/-</sup> animals and showed severe abnormalities. To identify downstream effects of *foxc1* deficiency, D-KO eye samples were collected at different developmental stages, RNAseq was performed and results will be presented. Our data show that *foxc1* is necessary for the proper eye development in zebrafish and that the developed zebrafish lines serve as a suitable model for *FOXC1*-related human ocular disease.



## PgmNr 568: cfDNA screening in triplets: A clinical laboratory experience.

### Authors:

B. Dyr; S. Caldwell; J. Rafalko

[View Session](#) [Add to Schedule](#)

**Affiliation:** Sequenom Inc., Laboratory Corporation of America Holdings, San Diego, CA

---

**Introduction:** Cell-free DNA (cfDNA) aneuploidy screening in multifetal pregnancies has been validated and clinically available since 2012. Since that time, over 900 samples from triplet gestations have been submitted for screening to one laboratory. Here, we describe the laboratory experience of the MaterniT<sup>®</sup>21 PLUS test in triplets.

**Methods:** Maternal blood samples submitted to Sequenom Laboratories<sup>®</sup> for MaterniT<sup>®</sup>21 PLUS testing were subjected to DNA extraction, library preparation, and genome-wide massively parallel sequencing as previously described by *Jensen et al.* Data analysis was performed on samples submitted for testing with a fetal number of 'three' indicated on the test requisition. Outcome data were elicited by phone or email from the ordering provider.

**Results:** From 2013 to 2019, 932 triplet samples were submitted for cfDNA screening. The majority of samples (79.2%) received a reportable result (i.e. positive or negative); the average fetal fraction (FF) in the reportable samples was 13.8%. This triplet cohort had a higher non-reportable (NR) rate due to low FF (20.2%) when compared to singleton pregnancies (<1%). In triplet samples that were NR due to low FF the average FF was 8.3%, well within the reportable range for a singleton, however the FF requirements for triplets are approximately three fold. Very few NRs were due to technical reasons (0.6%). The overall positivity rate for trisomy 21, 18, 13 in this cohort was <1% (n=7). Results of diagnostic testing were available for two of the seven positive cases, and both cases confirmed the cfDNA finding. Two additional cases had significant clinical findings consistent with increased risk for trisomy. No false negatives were reported to the laboratory. For the overall triplet cohort, the average turnaround time was 5.74 calendar days. Average turnaround times continue to decrease; in triplet samples resulted in 2018 and 2019 the average turnaround time improved to 3.8 and 3.15 calendar days, respectively.

**Conclusions:** cfDNA screening and ultrasound represent the only options for aneuploidy screening in triplets and other higher order multiples since serum screening is only available for singletons and twins. Higher NR rates are largely attributed to more stringent signal-to-noise ratio and higher FF requirements in multifetal pregnancies. Outcome information submitted from ordering providers suggest that cfDNA screening is an accurate and reliable screening method for triplet gestations.

## PgmNr 569: A novel universal PCR-based library preparation for PGT-A allowing cross-platform NGS sequencing.

### Authors:

S. Myers; K. Warren; L. Tay; S. Protopsaltis; M. Jasper

[View Session](#) [Add to Schedule](#)

**Affiliation:** PerkinElmer Health Sciences Australia, Thebarton, South Australia, Australia

Single-tube PCR-based Library Preparation (PCRLP) for Preimplantation Genetic Testing for Aneuploidy (PGT-A) by Next Generation Sequencing (NGS) is a new technology that has advantages over traditional two-step approaches, including protocol time efficiencies and reduced hands-on time. Validation of new technologies is important, and a powerful way to do this is with cross-validation. However, this is challenging with trophectoderm (TE) biopsies that typically provide limited sample template. Embryos can be re-biopsied, but biological variation in the form of mosaicism complicates data reporting. An alternative approach is possible using a novel single-tube PCRLP for PGT-A, developed based on the PerkinElmer DOPlify® WGA kit. This approach has the traditional advantages of PCRLP approaches but also allows cross-validation from a single TE biopsy.

Here, we describe the development of a novel approach for PCRLP of TE biopsies that allows cross-platform NGS using either Illumina® or Ion Torrent® sequencing technology.

Five-cell samples resembling TE biopsies were manually sorted from aneuploid cell lines (Coriell Institute). Cell lysis and WGA were performed using a modified DOPlify® kit protocol (PerkinElmer). WGA DNA was split and a second round of PCR was performed for incorporation of either Illumina® or Ion Torrent® specific adapter sequences. Amplified, indexed cell samples were purified, quantified and pooled before sequencing. Illumina® sequencing was undertaken on the MiSeq® instrument (Illumina®) while Ion Torrent® sequencing was performed on the S5 instrument (Thermo Fisher®). Sequencing data was analysed for correct aneuploidy calling using PG-Find™ software (PerkinElmer). Illumina® samples yielded an average of 580,000 reads per sample with 98.2% of reads mapping to hg19. Ion Torrent® samples yielded an average of 470,000 reads per sample with 97.9% of reads mapping to hg19. All samples passed quality control and displayed the expected karyotype when analysed by PG-Find™ software.

This novel PCRLP workflow provides rapid, scalable and economical sequencing for PGT-A and provides the capability for cross-platform sequencing validation of a single embryo biopsy for PGT-A. This flexible workflow also allows customisable throughput and tailorable resolution to detect smaller segmental aberrations. Large-scale cell-line validation of both the Illumina® and Ion Torrent® workflows are underway.

For research use only. Not for use in diagnostic procedures.

## PgmNr 570: Mathematical modeling reveals non-independence of meiosis I and meiosis II errors during human oogenesis.

### Authors:

J. Xing <sup>1</sup>; K.M. Tyc <sup>1</sup>; R.C. McCoy <sup>2</sup>; K. Schindler <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Genetics, Rutgers Univ, Piscataway, NJ; 2) Dept of Biology, Johns Hopkins Univ, Baltimore, MD

---

Approximately 10% of women world-wide experience pregnancy difficulties. Aneuploidy is a major contributor of pregnancy problems in women, resulting in pregnancy loss and *in vitro* fertilization (IVF) failure. Although most aneuploid conceptions are thought to originate from female meiotic division errors, quantitative studies that formally link the observed phenotypes to underlying error mechanisms are lacking. The goal of this project is to classify mechanisms of erroneous chromosome segregations underlying the production of poor-quality eggs.

Using mathematical modeling we investigate the rates of chromosomal mis-segregations that lead to aneuploid conceptions. In the model, the probabilities of all gamete outcomes are modelled as the combination of different types of errors. We consider the probability of non-disjunction (NDJ) errors in both Meiosis I (MI-NDJ) and Meiosis II (MII-NDJ), and premature separation of sister chromatids (PSSC) in Meiosis I (MI-PSSC). The rates were parameterized based on published preimplantation genetic testing for aneuploidy (PGT-A) results performed on trophectoderm cells during IVF procedures. Our results demonstrate that we can model all possible conception outcomes with five parameters that approximated the observed data. Using our modeling strategy, we observed MI-PSSC to be more prevalent than MI-NDJ across all the ages, a result that has long been speculated. The modeling also revealed the non-independence of Meiosis I and Meiosis II errors. Namely, abnormal secondary oocytes were more likely to suffer from additional segregating errors during Meiosis II. We show how the model can be used to identify IVF patients that are outliers for the number of aneuploid embryos based on their PGT-A results. Overall, the dynamic nature of the mathematical model makes it a potential tool for predicting the outcomes of assisted reproduction. This in turn could help to improve implantation success rates which have lingered at 30% for the past 20 years.

## PgmNr 571: Preimplantation genetic testing for aneuploidies (PGT-A) by metabolomic profiling of spent culture media using Raman spectroscopy.

### Authors:

Y.X. Zhang <sup>1</sup>; Q.S.Y. Yeung <sup>2</sup>; B. Liang <sup>4</sup>; Y.K.Y. Kwok <sup>1</sup>; W.T. Liu <sup>1</sup>; H.K. Wong <sup>1</sup>; L.M. Xuan <sup>5</sup>; L.Y. Zhang <sup>1</sup>; J.P.W. Chung <sup>2</sup>; K.W. Choy <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR, China.; 2) Assisted Reproductive Technology Unit, Department of Obstetrics and Gynaecology, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong SAR, China.; 3) Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; 4) State Key Laboratory of Microbial Metabolism, Joint International Research Laboratory of Metabolic and Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai, China.; 5) Basecare Medical Device Co., Ltd., 218 Xinghu Road, SIP., Suzhou, Jiangsu, China

### Background:

With the improvement of Preimplantation Genetic Testing for Aneuploidies (PGT-A) in live birth rate, this intervention has been widely used but controversial on its potential effect by biopsy and cost-effectiveness. The application of spent culture media (SCM) from blastocysts for aneuploidy screening by Raman spectroscopy has been reported to be success recently and we are interested to assess its performance.

### Methods:

A total of 239 SCM samples from 63 cycles were collected from 2018 to 2019. The collected media has both G1/G2 and GTL as base media from blastocyst in PGT cohort (230 SCM from 55 cycles) and were separately trained and tested. There were 111 SCM have done PGT-A on blastocyst biopsy, SCM and Raman parallelly. Nine SCM from 8 cycles have reproductive outcomes in non-PGT cohort. PGT-A results was regarded as the standard and reproductive outcome using clinical pregnancy as cutoff, Raman spectra of SCM of the corresponding embryos were classified into normal or abnormal. The algorithms were trained by a set of known PGT-A samples, and then tested the rest blindly. The spectra data were predicted by 5 machine learning algorithms: Convolutional Neural Networks (CNN), k-nearest neighbors (kNN), random forests (RF), extreme gradient boosting (XGB), and stacking.

### Results:

There were 84 abnormal and 46 normal SCM from G1/G2 group, while 34 abnormal and 28 normal from GTL group. At the first phase, all SCM in both groups were analyzed. The best prediction model was stacking for both G1/G2 and GTL groups, with an overall (GTL+G1/G2) 95.41% positive predictive value (PPV), and 53.1% negative predictive value (NPV), with 93.24% specificity and 63.03% sensitivity. The testing duration of each sample was around 5 minutes, enabling a fast turnaround time. In phase two, by further excluding segmental, subchromosomal, mosaic and sex chromosomal aneuploidies, leaving only single/multiple numerical aneuploidies and euploid to analysis on 136 samples. The best prediction model was XGB, with an overall PPV of 93.33% and NPV of 76.90%, with specificity at 95.89% and sensitivity at 66.67%.

### Conclusion:

Raman spectroscopy on spent culture media for PGT-A has a good performance on positive predictive value. Although the nature of Raman spectrum has a relatively low sensitivity, its ease of use, quick turn-around time, and non-invasiveness merits. Raman spectroscopy as a first-tier screening test for excluding single and multiple numerical aneuploidies.

## PgmNr 572: Gene discovery informatics toolkit defines candidate genes for unexplained infertility and prenatal or infantile mortality.

### Authors:

R. Dawes <sup>1,2</sup>; M. Lek <sup>3</sup>; S.T. Cooper <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Kids Neuroscience Centre, Kids Research, Children's Hospital at Westmead, Sydney, NSW, 2145, Australia; 2) Discipline of Child and Adolescent Health, Faculty of Health and Medicine, University of Sydney, Sydney, NSW, 2006, Australia; 3) Yale School of Medicine, Yale University, 333 Cedar Street, New Haven, CT, 06510, USA

---

Despite a recent surge in novel gene discovery, genetic causes of prenatal-lethal phenotypes remain poorly defined. To advance gene discovery in prenatal-lethal disorders, we created an easy-to-mine database integrating known human phenotypes with inheritance pattern, scores of genetic constraint, and murine and cellular knockout phenotypes—then critically assessed defining features of known prenatal-lethal genes, among 3187 OMIM genes, and relative to 16,009 non-disease genes.

While around one-third (39%) of protein-coding genes are essential for murine development, we curate only 3% (624) of human protein-coding genes linked currently to prenatal/infantile lethal disorders. 75% prenatal-lethal genes are linked to developmental lethality in knockout mice, compared to 54% for all OMIM genes and 34% among non-disease genes. Genetic constraint correlates with inheritance pattern (autosomal recessive << autosomal dominant < X-linked), and is greatest among prenatal-lethal genes. Importantly, over 90% of recessive genes show neither missense nor loss-of-function constraint, even for prenatal-lethal genes. Detailed ontology mapping for 624 prenatal-lethal genes shows marked enrichment among dominant genes for nuclear proteins with roles in RNA/DNA biology, with recessive genes enriched in cytoplasmic (mitochondrial) metabolic proteins.

We conclude that genes without genetic constraint should not be excluded as potential novel disease genes, and especially for recessive conditions (under 10% constrained). Prenatal lethal genes are 5.9-fold more likely to be associated with a lethal murine phenotype than non-disease genes. Cell essential genes are largely a subset of mouse-lethal genes, notably under-represented among known OMIM genes, and strong candidates for gamete/embryo non-viability. We therefore curate 3435 'candidate developmental lethal' human genes: essential for murine development or cellular viability, not yet linked to human disorders, presenting strong candidates for unexplained infertility and prenatal/infantile mortality.

## PgmNr 573: Localization of beta and gamma actin in early fetal brain development.

### Authors:

I. Niehaus <sup>1</sup>; N. Walker <sup>3</sup>; M.B. Rasonja <sup>2</sup>; N. Jovanov <sup>2</sup>; E. Schrock <sup>1</sup>; W.B. Huttner <sup>3</sup>; N. Di Donato <sup>1</sup>; M. Heide <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Faculty Carl Gustav Carus, TU Dresden, Institute for Clinical Genetics, Dresden, Germany; 2) Croatian Institute for Brain Research, School of Medicine, University of Zagreb; 3) Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

---

Actin is an ubiquitously expressed protein and plays a tremendous role in multiple cell functions. The family of Actin consists of 6 isoforms, whereby only the two cytoskeletal isoforms  $\beta$ CYA and  $\gamma$ CYA are expressed in adult as well as in fetal brains. The distribution of  $\beta$ CYA and  $\gamma$ CYA in the human cortex is not yet known. We analyzed the enrichment of these two isoforms in early fetal corticogenesis. In sections of 12 weeks post conception (wpc) neocortex,  $\gamma$ CYA is the predominant isoform in all layers except the adhesion junction belt of the ventricular zone (VZ) and marginal zone (MZ) where  $\beta$ CYA represents the major actin isoform. At 16 wpc the expression pattern changes.  $\beta$ CYA becomes the predominant isoform in the cortical plate (CP) and MZ. Additionally, we analyzed the intracellular distribution of the two isoforms in *in vitro* differentiated neural progenitor cells (NPC) and neurons. In NPC, both isoforms are concentrated to the cell periphery rather than the lumen. Neurons show an increased enrichment of  $\beta$ CYA in neurites and cell processes of single cells, whereas  $\gamma$ CYA prevails in the cell body. In migrating neurons,  $\beta$ CYA is the predominant actin isoform of the cell lumen. We will present a detailed assessment of differential enrichment of the two cytoskeletal actin isoforms during early human corticogenesis on cellular and subcellular levels.

## **PgmNr 574: Motor planning deficits in a prenatally diagnosed population of 47,XXY (Klinefelter syndrome) and the variable impact of hormonal replacement therapy (HRT).**

### **Authors:**

P. Lasutschinkow <sup>1</sup>; A. Gropman <sup>2</sup>; T. Sadeghin <sup>1</sup>; C. Samango-Sprouse <sup>3,4,1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) The Focus Foundation, Davidsonville, Maryland; 2) Children's National Medical Center, Washington DC; 3) George Washington University School of Medicine and Health Services, Washington, DC.; 4) Department of Human and Molecular Genetics, Florida International University, FL

47,XXY is the most common X&Y chromosomal variation (1:600) & the motor planning deficits have not been investigated. Specific motor planning deficits & potential impact of Hormonal Replacement Therapy (HRT) was investigated.

80 boys with 47,XXY were evaluated with Amsterdam Neuropsychological Test (ANT), Beery Buktenica Developmental Test of Visual-Motor Integration (BEERY-6), Bruininks-Oseretsky Test (BOT-2), Comprehensive Trail Making Test (CTMT) & Weschler Intelligence Scale for Children (WISC-IV &V). There were four groups: non-HRT, early hormonal therapy (EHT), hormonal booster therapy (HBT), or EHT&HBT (EB). 45 did not receive HRT; EHT-11; HBT-14; EB-10. HRT was given based on participant's pediatric endocrinologist's evaluation.

All groups had below to low average scores on fine motor precision & manual dexterity on BOT-2 & depressed scores in visuomotor & motor planning on ANT. Scores were below average on CTMT. On Beery-6, all groups had below average motor coordination (MC) discrepant to visual motor integration (VMI) & visual perception (VP). On WISC, the lowest scores were in coding.

ANOVA testing revealed significant differences for treatment groups in fine motor integration(FMI) on BOT-2 ( $p=0.031$ ), block design(BD) ( $p=0.049$ ) & matrix reasoning(MR) ( $p=0.047$ ) on WISC & VP on Beery-6 ( $p=0.013$ ). Post-hoc comparisons using Holm procedure indicated significant difference in FMI ( $p=0.022$ ) between non-HRT ( $M=12$ ) & EHT ( $M=15$ ). Significant differences for BD & MR existed between non-HRT ( $M=11.09$ ;  $10.21$ ) & EB ( $M=13.8$ ;  $13.6$ ) ( $p=0.029$ ;  $p=0.016$ ). Significant differences in VP ( $p=0.022$ ) were evident between non-HRT ( $M=100.1$ ) & EHT ( $M=114.1$ ).

This is the first study to identify motor planning deficits in boys with 47,XXY and the findings begin to delineate various aspects of motor planning dysfunction in 47,XXY. Paper & pencil tasks consistent with dysgraphia are the weakest aspect of motor planning in 47,XXY as evident from scores on the WISC, Beery-6, ANT, CTMT & BOT-2. These results demonstrate that HRT may have a positive response on perceptual capabilities, with EHT treated boys performing better than non-HRT in VP & FMI & EB performing better in BD & MR. This study supports previous research that HRT has a positive impact on specific aspects of neurodevelopment in 47,XXY & also identified a selective response to treatment within motor planning.



## PgmNr 575: Preconception carrier screening yield: Effect of variants of unknown significance in partners of carriers with clinically significant variants.

### Authors:

H. Fridman <sup>1,2,3</sup>; D.M. Behar <sup>4</sup>; S. Carmi <sup>1</sup>; E. Levy-Lahad <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Braun School of Public Health and Community Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel; 2) Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel; 3) Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel; 4) Igentify, Tirat Hacarmel, Israel

**Purpose:** Expanded preconception carrier screening (ECS) identifies couples at risk for multiple diseases. ECS reports currently include only pathogenic/likely-pathogenic (P/LP) variants. Variants of unknown significance (VOUS) are not reported, unlike genomic or chromosomal-array test results in other post- or pre-natal settings. This policy omits pairings of P/LP carriers with VOUS carriers (P/LP\*VOUS), who may be at-risk. This situation is expected to be particularly common in genes with high P/LP carrier rates. We examined the possible contribution of P/LP\*VOUS matings to the yield of ECS in an Ashkenazi Jewish (AJ) cohort, a population with well-established preconception screening.

**Methods:** We analyzed 672 AJ genome sequences (225,456 virtual matings) for variants in three different gene-sets, using automated and manual ACMG classification. Gene-sets analyzed were an AJ (55 genes), pan-ethnic (168 genes) and combined (180 genes) sets.

**Results:** Across 180 genes, we identified 4671 variants: 144 (3.1%) P/LP and 1963 (42%) VOUS. The proportion of P/LP\*P/LP and P/LP\*VOUS couples at-risk was 2.7-3.8% and 17-20% respectively (for different gene-sets). Considering only coding-VOUS (i.e. exonic non-UTR VOUS), the proportion of P/LP\*cVOUS couples at-risk was  $\approx 7\%$ . Most (65-70%) P/LP\*VOUS matings were at risk for severe/profoundly-severe diseases.

We also modeled different scenarios of VOUS reclassification, based on hypothetical fulfillment of specific ACMG criteria. Hypothetical fulfillment of the PS4 criterion (significantly greater prevalence of the variant in affected individuals vs. controls) would reclassify 89.2% of VOUS to P/LP. In contrast, hypothetical fulfillment of the PM2 criterion (absence of the variant in controls) would reclassify only 1% of VOUS to P/LP.

**Conclusions:** Disregarding VOUS in ECS may miss couples at-risk for severe diseases. Even if only 10% of couples currently classified as P/LP\*cVOUS are ultimately reclassified as P/LP\*P/LP, ECS yield would increase by  $\approx 20\%$ . This suggests that VOUS in current ECS genes are more likely to affect ECS yield than sequencing an ever greater number of rare disease genes.

These results should not be interpreted as a call for including VOUS in ECS reports, but rather as evidence for the importance of addressing variant classification in ECS-panel genes. Further data collection on variants in affected individuals is likely to have the greatest impact on such reclassification.

## PgmNr 576: Expanded carrier screening using next generation sequencing: An experience from India.

### Authors:

S. Sharda; S. Nair; P. Palan; T. Geetha; A. Arunkumar; V. Ramchander; V. Gopalan; L. Mahadevan; S. Murugan; R. Vedam

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical Genetics, MEDGENOME Labs, Bangalore, India, MD

---

Carrier screening of couples is currently limited to genetic disorders with a high frequency, a positive family history or is majorly ethnicity based. Couples may not even be aware that they could be carrying a genetic variant, which could lead to their unborn child developing a genetic disorder. In most cases, carriers of a genetic disorder are asymptomatic and may have no family history and unfortunately their carrier status is often confirmed only after an affected baby is born.

With the availability of advanced molecular technologies, like massively parallel sequencing, expanded genetic screening aims to identify carriers for a much larger group of genetic diseases, not limited to ethnicity or a positive family history. There is no carrier screening program available in India, at present except for limited screening for thalassemia. Expanded carrier testing is the need of the hour in a country with high population, high birth rate and a high rate of consanguinity.

Medgenome Labs have tested around 2000 Indian couples for their carrier status, using Next Generation Sequencing and MLPA technology. We have screened couples for Autosomal and X-linked recessive disorders and have provided clinically significant results, based on which the couples went ahead for prenatal diagnosis. The most common indications for referral were previous child affected or suspected to be affected with a genetic disorder [known/unknown], consanguinity, bad obstetric history including recurrent pregnancy and IVF loss or a family history of suspected genetic disorder. Carrier screening provides the couple an opportunity to take informed reproductive choices, including prenatal diagnosis, pre-implantation genetic diagnosis or even preparedness for the newborn.

The data would also help to know the most common genetic disorders with their carrier frequency in the Indian population and if there is any difference compared to the western literature. Most of the carrier frequencies of genes and common variants is available from the western literature and ethnicity based frequency of many genetic disorders are not available for Indian population. This information can be used to counsel couples regarding their risk as well as help them take an informed decision. This retrospective analysis will certainly help the primary care providers make guidelines for primary as well as secondary prevention of rare genetic disorders which have an emotional and financial burden on the parents and society

## PgmNr 577: Autozygosity mapping and its relation to the timing of birth in Norwegian family trios.

### Authors:

P. Sole-Navais<sup>1</sup>; J. Bacelis<sup>1</sup>; G. Zhang<sup>2,3</sup>; O. Helgeland<sup>4,5</sup>; D. Modzelewska<sup>1</sup>; M. Vaudel<sup>5,6</sup>; S. Johansson<sup>5,6</sup>; P.R. Njolstad<sup>5,7</sup>; L.J. Muglia<sup>2,3</sup>; B. Jacobsson<sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Obstetrics and Gynecology at Institutet, Gothenburg University, Göteborg, Sweden; 2) Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States; 3) Human Genetics Division, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, United States; 4) Division of Health Data and Digitalization, Department of Genetics and Bioinformatics, Norwegian Institute of Public Health, Oslo, Norway; 5) KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway; 6) Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway; 7) Department of Pediatrics, Haukeland University Hospital, Bergen, Norway

Long autozygous (homozygous IBD segments) are enriched in low frequency and rare deleterious variants. Its mapping in genotyping array studies captures the effects of non-genotyped variants with recessive effects that may lie within these segments. Birth timing has a broad-sense heritability ranging between 25-40%, and maternal and dominance components account for 15% and 11% of its variance, respectively.

Here, we leveraged the relationship between parental genetic relatedness and offspring runs-of-homozygosity (ROH) in 15000 family-trios to estimate autozygosity while avoiding the arbitrary selection of calling parameters. We estimated the maternal, paternal and fetal effects of genome-wide autozygosity and identified regions with recessive effects on spontaneous delivery risk.

DNA from 15000 triades from the Norwegian Mother and Child cohort was genotyped using different Illumina arrays. We called ROH in the offspring multiple times varying the parameters used, and selected, within each batch, those parameters that maximized the coefficient of determination between offspring ROH and parental genetic relatedness. Autozygosity was then estimated in each member of the family trios as the fraction of genome covered by autozygous segments. Genomic regions with recessive effects were identified by autozygosity mapping. Cox proportional hazard regressions were fitted in mothers, fathers and offspring to estimate the effect of genome-wide and regional autozygosity on spontaneous delivery risk.

We observed differences in  $R^2$  ranging from 0.01 to 0.75 between offspring ROH and parental genetic relatedness using different parameters, which were highly dependent on genotyping platform. Genome-wide autozygosity analysis suggested a modest effect of maternal autozygosity on spontaneous delivery risk (HR: 1.04 [95%CI: 0.99,1.09]) and gestational age (beta: -0.6 [-1.0, -0.1]). We observed no effect of fetal nor paternal autozygosity on birth timing. We identified 259 maternal and 280 fetal variants with a false discovery rate  $<0.3$  in at least two of the 3 batches studied (lowest p-values  $4.9 \times 10^{-09}$ ) in 7 protein coding genes (RELL1, TBC1D1, PGM2, RPTOR, CARD14, SLC26A11 and

GRHL2).

Genome-wide autozygosity in the mother is associated with modest effect on birth timing, but not in the offspring nor father. Autozygosity mapping highlighted loci in glucose, insulin and cell differentiation and growth pathways associated with spontaneous delivery risk.

## PgmNr 578: Expanded carrier screening research using a unique multiplex targeted sequencing panel.

### Authors:

J. Bishop; A. Harris; G. del Mistro; C. Davidson; J. Veitch; R. Putler; T. Fahland; H. Xie; V. Mozhaysky; C. Bai; L. Le; J. Burrill; T. Huan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Thermo Fisher Scientific, 5781 Van Allen Way, Carlsbad, CA 92008

---

To create a sample-to-answer workflow with flexible throughput for expanded screening research of carrier variants, we designed a multiplex next generation sequencing (NGS) panel targeting 99.96% of coding bases plus flanking regions for 420 genes implicated in Mendelian disorders. This panel includes several complex, high-value targets that are historically difficult to assess by NGS, including the *GBA*, *SMN1*, *CYP21A2*, and *HBA1/2* loci. Eight unlinked, non-exonic, high minor allele frequency SNP markers were also included in the panel to generate a unique sequencing ID for each research sample. Panel performance was tested against cell lines containing known variants, and under-performing regions were iteratively redesigned to optimize a variety of parameters, including percentage of bases covered within the targeted loci, evenness of target coverage, and ability to call SNV, indel, and CNV variants. Libraries were created from as little as 20 ng of input DNA, yielding >99% library success rate from cell line, blood, and buccal-swab samples tested. Libraries were individually barcoded and then pooled to yield on average ~3M reads per library at up to 32 libraries per each three hour semi-conductor-based sequencing run. At this read depth, analytical sensitivity and specificity for key variants were >99% for SNV detection, and >98% for CNV detection, as empirically determined using samples with known-truth variants. Analytical reproducibility of key called-variants was >99% and no-call rate for key hotspot SNVs was <0.4%. Reporting software was also created to provide functional context for, and rapid reporting of, identified variants.

## PgmNr 579: It's not all about the extra X: The potential role of copy number variants (CNVs) in the neurodevelopmental variability of 47,XXY (Klinefelter syndrome).

### Authors:

S. Tran <sup>1</sup>; C. Samango-Sprouse <sup>1,2,3</sup>; T. Sadeghin <sup>1</sup>; A. Gropman <sup>2,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Focus Foundation, Davidsonville, Maryland; 2) Department of Pediatrics, George Washington University, Washington, D.C.; 3) Department of Human and Molecular Genetics, Florida International University, Miami, Florida; 4) Department of Neurogenetics and Developmental Pediatrics, Children's National Medical Center, Washington, D.C.; 5) Department of Neurology, George Washington University, Washington, D.C.

---

**Background:** 47,XXY (Klinefelter syndrome) is the most commonly occurring sex chromosomal aneuploidy (1:650). The phenotypic profile of this syndrome ranges widely but is often associated with executive dysfunction, speech and motor delay, and language-based learning disorders. Recent research has shown that copy number variants (CNVs) may be associated with neurodevelopmental disorders. However, the prevalence of CNVs in conjunction with 47,XXY, and subsequent effects on neurodevelopmental variability, is not clearly defined. This study investigated the potential impact that CNVs may have in boys with 47,XXY.

**Method:** Six boys (47-CNV) with 47,XXY and varying CNVs (1p36.22 duplication (dup), 2p23.3 dup, 5q35.1 dup, 8q22.3 dup, 15q21.1 deletion (del), Xp22.31 dup) were compared to eleven boys (47) with 47,XXY across cognition (COG), motor (MOT), auditory comprehension (AC), expressive communication (EC), and expressive (EL) and receptive language (RL). Domains were evaluated using the Bayley Scales of Infant Development (BSID), Preschool Language Scales (PLS), Expressive One Word Picture Vocabulary (EOWPVT), and Receptive One Word Picture Vocabulary (ROWPVT). Ages for all ranged from 24-59 months ( $M=29.7 \pm 9.5$ ).

**Results:** On the BSID, no significant differences were found in COG. There was a significant difference in MOT ( $P=0.05$ ), where the 47-CNV boys had lower MOT. Scores for EC on the PLS did not reveal significant differences, but there was a difference between the 47-CNV boys and the 47 boys regarding AC on the PLS ( $P=0.48$ ), as the 47 boys exhibited higher capabilities. 47 boys had greater EL ( $P=0.043$ ) and RL ( $P=0.019$ ) than the 47-CNV boys on the EOWPVT and ROWPVT.

**Conclusion:** The 47-CNV boys presented with lower capabilities in MOT, EL, RL, and AC than the 47 boys within the first 5 years of life. Research has shown that these domains are generally areas of weakness for 47,XXY, and these results suggest that CNVs may be an underlying genetic component contributing to the neurodevelopmental variability often seen in 47,XXY. When boys with 47,XXY exhibit a greater degree of difficulty in their neurodevelopment, further genetic testing should be considered to look for the presence of CNVs. Lastly, the CNVs described in this cohort are currently reported to be of unknown clinical significance; thus, more research is warranted to determine if these CNVs are also linked to neurodevelopment, and how these CNVs manifest themselves in

conjunction with 47,XXY.

## PgmNr 580: The molecular characterization of genetic pre-disposition to fibroid growth in black women attending the Centre Hospitalier de Recherche en Chirurgie Endoscopique et Réproduction Humaine(CHRACERH)-Yaounde, Cameroon.

### Authors:

M.S.B.E Ngwafor<sup>1</sup>; J.P. Chedjou<sup>2</sup>; N. Nguediak Kaze<sup>1</sup>; H. Dongmo<sup>2</sup>; C. Tah<sup>2</sup>; A. Noubom<sup>3</sup>; S. Biboum<sup>3</sup>; C. Noa<sup>3</sup>; E. Belinga<sup>1,3</sup>; M. Toukam<sup>1,3</sup>; W.F. Mbacham<sup>1,2</sup>; J.M. Kasia<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculty of Medicine and Biomedical Sciences-University of Yaounde I, Yaounde, Center , Cameroon;  
2) Laboratory for Research and Public Health Biotechnologies-Biotechnology Center, University of Yaounde I , Cameroon; 3) Centre Hospitalier de Recherche en Chirurgie Endoscopique et Reproduction Humaine, Yaounde-Cameroon

Fibroids are the most common benign gynecological tumors of the uterus accounting for 20-50%. It's pathogenesis is poorly understood. It has been hypothesized that diet, lifestyle, immunological and inflammatory processes may influence genetic polymorphisms of growth factors of uterine leiomyoma. With paucity in data in Africa and Cameroon in particular, the *VEGF* -460 C/T is known to pre-dispose women to fibroids where as *FGF2* 754C/G mutation has been shown to pre-dispose women to adenomyosis even though its gene is implicated in the onset of fibroids. Thus we embarked on studying the presence of these mutations that could be associated to the development of fibroids in women of childbearing age attending CHRACERH in order to establish an association of these of markers likely to contribute to fibroid growth. We conducted a case-control study, with the recruitment of women aged 20 and above; 30 cases of USS diagnosed fibroma. In the control arm, we had 10 women with no fibroids. Prior to this, participants signed an informed consent and completed a questionnaire. Thereafter, a physical examination was done and blood samples collected on 3MM filter paper (dried blood spots-dbs) was taken to LAPHORB for genomic analyses. The Chelex 100 method for DNA extraction. The determination of the 754 C/G and 460 C/T polymorphisms of the *FGF2* and *VEGF* genes by PCR-RFLP using the endonucleases *HhaI* and *BshI*2361 respectively. This was carried out during a period of two months. At the end of the study we obtained 95% genotype GG (mutant) for the *FGF2* 754 gene with a preponderance of the G allele-97.5%. In the case of positive fibroids we obtained 93% GG genotype with 7% CG in the control group. We obtained 100% TT genotype (mutant) for the *VEGF* gene in both groups similar to studies done in Chinese women with T allele predisposing women to fibroids and the C allele with less exposure to fibroids. With respect to the association between fibroids and gene 754C/G mutation, we obtained an OR of 1.071 on a 95% CI of 0.974-1.179 with a p value of 0.402 which reflects no association between the mutation and occurrence of fibroid. It should be noted this mutation has not been studied as a biomarker pre-disposing women to fibroids but its gene *FGF2* has been implicated in fibroid development. Given no association was found between the mutations and onset of fibroids in this African population, a basis is set for the investigation of other putative fibroid biomarkers.



## PgmNr 581: Integration of genetic ancestry with DNA methylation identifies novel placental epigenetic signatures implicated in birthweight and adult complex diseases.

### Authors:

F. Tekola Ayele; X. Zeng; T. Workalemahu; M. Ouidir

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Intramural Population Health Research, NATIONAL INSTITUTES OF HEALTH (NIH), BETHESDA, Maryland.

---

Integrating genetically derived ancestry estimates with DNA methylation (DNAm) in fetal tissues from diverse populations may give clues about early developmental mechanisms and later life health disparities. Here, we report a study that interrogated multi-omic data on placenta collected from ethnically-diverse individuals aiming: 1) to identify ancestry-specific as well as shared DNAm signatures of genetic ancestry composition; 2) to test whether these DNAm signatures of ancestry are associated with fetal growth; and 3) to test for enrichment of ancestry-associated CpGs with mQTLs implicated in adult diseases. Placental biopsies were collected during delivery from 312 self-identified Black, White, Hispanic, and Asian women in the U.S. DNA and RNA isolated from placental samples underwent genotyping on the Illumina 2.5M array, DNAm on Illumina 450k array, and RNAseq on Illumina HiSeq2000. The percentages of African (AF), European (EU), Native American (NA) and East Asian (EA) ancestries were estimated using ADMIXTURE. Covariate-adjusted linear regression models tested associations between genetic ancestry percentage and DNAm. Analyses were performed for: EU, AF, and NA ancestry percentages in Hispanics; and AF, EU, and EA ancestry percentages in Blacks, Whites, and Asians, respectively. A total of 18 CpG sites were significantly (FDR  $P < 0.05$ ) differentially methylated with AF ancestry percentage in Blacks; 91 CpGs with EU ancestry percentage in Whites, 68 CpGs with EA ancestry percentage in East Asians; and 234 CpGs, 12 CpGs, and 6 CpGs with NA, EU, and AF ancestry percentage in Hispanics. Five CpGs (*ERI1*, *GCC2*, *BMF*, *PGAM5*, *NARFL*) showed directionally consistent associations with both NA ancestry in Hispanics and EA ancestry in East Asians. Among the top hits, 5 EU ancestry CpGs in Whites, 3 EA ancestry CpGs in Asians, and 8 NA ancestry CpGs in Hispanics were negatively correlated with the expressions of their respective genes. An EU-ancestry CpG (*SLC16A3*) and an NA-ancestry CpG (*IER3*) showed Bonferroni-corrected significant association with higher birthweight. Several ancestry-associated CpGs were enriched with mQTL SNPs in blood across the life span, including more than 50 GWAS SNPs associated with 24 adult traits. This novel study demonstrates that links between genetic ancestry and epigenetic marks in diverse populations may underlie differences in early growth, the developmental origins of health and disease, and health disparities.

## PgmNr 582: Exome sequencing expands knowledge of early phenotypes in well-known genetic conditions.

### Authors:

S. Rego <sup>1</sup>; T. Yip <sup>1</sup>; C. Fairley <sup>3</sup>; P.M. Martin <sup>1</sup>; U. Hodoglugil <sup>1</sup>; B. Anguiano <sup>1</sup>; J. Tenney <sup>2</sup>; A. Slavotinek <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Human Genetics, UCSF, San Francisco, California.; 2) Division of Medical Genetics, UCSF, San Francisco, California.; 3) UCSF Benioff Children's Hospital-Oakland, Oakland, California

---

Increased utilization of exome sequencing (ES) in infants has led to earlier diagnoses of well-known genetic conditions. Such diagnoses expand knowledge of early phenotypes and presentations which may not fulfill clinical diagnostic criteria (CDC). We identified several such cases as part of UCSF's Program in Prenatal and Pediatric Genomic Sequencing (P3EGS) study.

Patient 1 is a 16mo female of European ancestry who was evaluated at 2mo for failure to thrive, mild hypotonia, and multiple congenital anomalies including persistent left superior vena cava, hypoplastic kidneys, ventricular septal defect, bilateral ear pits, congenital sacral dimple, tethered cord, and a history of single umbilical artery. At 5mo she was also noted to have developmental delays (DD). ES identified a *de novo* variant in *KMT2D* predicting p.Met999\*fs. The patient did not meet CDC for Kabuki syndrome, which emphasize facial features that are difficult to appreciate in infancy, and had not yet been noted to have DD, which is also important for CDC, when first seen by genetics.

Patient 2 is a 11mo Latina female seen by genetics at 3mo during a hospitalization for respiratory failure. She also had a history of hypotonia and dysphagia. At 8mo, ES revealed a *de novo* variant in *MECP2* predicting p.Arg168\*. She did not meet CDC for Rett/Atypical Rett syndromes, as she was too young to manifest regression or stereotypic hand movements. While respiratory problems have been described in Rett syndrome, the severity and timing were atypical.

Patient 3 is a 15mo Latino male seen by genetics at 5mo due to cleft lip/palate and microcephaly. He also had a history of DD, feeding difficulties, hearing loss, and a patent foramen ovale. The patient's physical exam was notable for plagiocephaly/brachycephaly, upslanting palpebral fissures, and thickened eyebrows. His mother had history of cleft lip and an atrial septal defect. ES showed a maternally inherited *RAD21* variant predicting p.R478\*. CDC for Cornelia de Lange rely heavily on facial features—particularly synophrys—not present in our patient, and as such he did not meet criteria.

ES is providing infants with diagnoses of classic genetic conditions for which infantile phenotypes may not be well-described and the cardinal features emphasized in CDC may onset later. These cases suggest diagnostic utility of ES for infants with suspected genetic conditions who may not meet CDC and suggest CDC may need to be expanded to encompass infantile presentations.

## PgmNr 583: Toll-like receptor 6 (*TLR6*) SNPs P249S rs5743810 and Intron rs1039559 strengthen the association of *TLR6-TLR1-TLR10* gene cluster with preterm birth in Wisconsin cohort.

### Authors:

D.M. Pillers<sup>1</sup>; J. Eickhoff<sup>2</sup>; M. Baker<sup>3</sup>; B.R. Pattnaik<sup>4</sup>; W. Luo<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, University of Illinois-Chicago, Chicago, Illinois.; 2) Biostatistics and Medical Informatics, University of Wisconsin-Madison; 3) Newborn Screening Laboratory, Wisconsin State Library of Hygiene, Pediatrics and Genetics, University of Wisconsin-Madison; 4) Pediatrics and Ophthalmology, University of Wisconsin-Madison

**Background:** Preterm birth (PTB) occurs in 11% of US births, causing significant morbidity and mortality. A subset of PTB is complicated by infection and potential aberrations in the microbiome. The Toll-like receptors (TLR) of the innate immune system are integral to pathogen recognition and susceptibility to infection. We speculated that SNPs in fetal TLR genes may be associated with the subset of PTB with underlying infection as an etiology. We have previously reported results for SNPs in three TLRs (1, 2, and 4) in Wisconsin infants, with only *TLR1* SNPs associated with PTB. *TLR1* is in the *TLR6-TLR1-TLR10* gene cluster that functions in inflammation and infection. We tested SNPs in *TLR6* and *TLR10* and found that *TLR6* SNP P249S also associates with PTB.

**Objective:** SNP associations with PTB have not been easily replicated in prior published works. Thus, we sought to validate our results with additional SNPs at discrete positions in the *TLR6* gene. We hypothesized that *TLR6* SNP intron rs1039559 would similarly mark an increased risk for PTB.

**Design/Method:** Residual newborn screening DNA was tested with IRB approval. 3000 samples were used for *TLR1* SNP (N248S) analysis, with 866, and 1926 samples adequate for subsequent *TLR6* (P249S; rs5743810), and *TLR6* (intron rs1039559) testing. SNP genotype was performed by TaqMan SNP Genotyping Assay (Lifetechnologies, CA). Ancestral background was by maternal declaration.

**Results:** The *TLR6* SNP (P249S; rs5743810) minor allele T was associated with preterm gestational age (GA) < 34 wks, Odds Ratio (OR) 1.41 (95% CI: 1.15-1.72, P<0.00087) and with GA < 37 wks, OR 1.30 (95% CI: 1.02-1.67, P<0.032). However, in GA subset < 28 wks, the *TLR6* SNP (P249S) minor allele T findings were not consistent with those in the older GA < 37 wks and GA < 34 wks. The *TLR6* SNP (intron rs1039559) major allele A was associated with preterm GA < 28 wks, OR 1.21 (95% CI: 1.04-1.40, P<0.013) and with low birthweight (<1250g), OR 1.24 (95% CI: 1.08-1.42, P<0.0025).

**Conclusion:** We propose that in addition to *TLR1*, *TLR6* is another gene in the *TLR6-TLR1-TLR10* gene cluster that is associated with preterm birth. *TLR1* and *TLR6* are part of a tightly linked cluster of TLR genes that act as co-receptors with TLR2, a key step in the inflammation cascade. We speculate that aberrant *TLR1* and *TLR6* may lead to PTB via an altered downstream pathway causing disturbed commensal homeostasis leading to an altered microbiome and finally, to pathogen invasion.

## PgmNr 584: The bad, the ugly, and the good of direct to consumer genetic testing.

**Authors:**

P. Bitoun

[View Session](#) [Add to Schedule](#)

**Affiliation:** Groupe Medical Jarente, Medical Genetics, Paris, IDF, France

---

The author presents 3 cases of paternity search performed using Direct to Consumer Tests (DCT) from the searcher's point of view for one family, from the father's side in another family and from a relative in the third.

The first case presents to the clinic with results of his abnormal spermogram from which he is convinced that he may not be the father of his previous born son. He went on to present the DCT result of his 2 further born children, his wife, his previous born son. The discussion was about the understanding of these tests results, detailed presentation of the results data, the validity of commercial genetic tests and the psychological reasons for these tests.

The second family presents with a man whose 2 cousins who had previously performed DCTs and were later contacted by email about an unknown adopted individual claiming he had matched with them as possible cousins using the DCT databank family search tool. The man at first had not answered these emails from his cousins he thought were a joke until they called him informing him he may have been the father of this unknown individual from temporal and geographical evidence. The reaction, action and results of these will be presented in details.

The third case is a classic issue of an individual who posted his genome in the public space of 23 and me and Ancestry DNA and was contacted by a distant adopted relative searching for his father.

These cases are iconic examples of what geneticists may expect to be confronted with as the use of DCT are becoming commonplace and with their various impact on individuals and families.

# PgmNr 585: Utility of whole genome sequencing in an undiagnosed fetus with increased nuchal translucency: A case illustration.

## Authors:

M.M. Shi; Z.R. Dong; H.K. Chau; Z.J. Yang; K. Kwok; K.W. Choy

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Obstetrics & Gynaecology, The Chinese University of Hong Kong, Hong Kong, China

## Background:

Increased nuchal translucency (NT) in the first trimester is highly associated with chromosomal abnormalities. Karyotyping and chromosomal microarray (CMA) have been routinely performed for these cases, but only yield a diagnosis rate less than 20%. Monogenic syndromes such as Noonan syndrome are also related with increased NT but could not be detected by routine tests. Therefore, it warrants an integrated platform for comprehensive investigation of point mutations, structural variants (SVs) and copy number variants (CNVs). Whole genome sequencing (WGS) offers the potential of a single test detecting nearly all genomic variations. Herein, we retrospectively recruited an undiagnosed fetus with increased NT, and applied WGS for the investigation of its genetic etiology.

## Methods:

The case was recruited from Prenatal Genetic Diagnosis Centre in the Chinese University of Hong Kong. The fetus presented with increased NT (4.18 mm) at 13 gestation weeks and no other abnormalities were detected prenatally. Both karyotyping and CMA were normal. Clinical follow-up showed the fetus was live birth with multiple malformations (low set ears, increased nuchal fold, right hand single transverse crease, bilateral clinodactyly) and mild hearing loss. DNA was extracted from stored chorionic villi. WGS approach (30-fold) was performed for the fetus only to save the cost. Point mutations, CNVs and SVs was analyzed by our in-house pipeline and interpretation was performed according to the ACMG guidelines. Selected putative mutations were validated among the fetus and parents.

## Results:

No pathogenic CNVs and SVs were identified, but we detected a heterozygous frameshift mutation (NM\_0012561 82: c.2404dupC (p.L802fs)) of *ANKRD11* gene, which could cause KBG syndrome (OMIM: #148050). This mutation could result in a truncated protein and lead to nonsense mediated decay. Subsequent parental validation suggested that it arouses *de novo*. Since the child presented with clinical features associated with KBG syndrome, we classified this mutation as pathogenic. In addition, we detected a secondary finding of *GJB2* homozygous pathogenic mutation (NM\_004004 c.G109A (p.V37I)), which is consistent with hearing loss findings.

## Conclusion:

Our study demonstrated that WGS was feasible to investigate the genetic etiology of prenatal cases, and dual molecular diagnosis of KBG syndrome and *GJB2*-related hearing loss could be detected among undiagnosed fetuses with increased nuchal translucency.

## PgmNr 586: Accurate quantification of copy number of the $\alpha$ -globin genes using CarrierScan™ Assay.

### Authors:

S. Mohan<sup>1</sup>; L. Le<sup>1</sup>; R. Varma<sup>1</sup>; D. Behar<sup>2,3</sup>; C. Bormans<sup>2</sup>; J. Winters<sup>4</sup>; L. Hasadsri<sup>4</sup>; P. Lundquist<sup>4</sup>; A. Roter<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) ThermoFisher Scientific, Santa Clara, CA.; 2) . Gene by Gene, Genomics Research Center, TX; 3) Igentify, Tirat HacCarmel, Haifa, Israel; 4) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN

---

Copy number variations in the two alleles of the  $\alpha$ -globin genes *HBA1* and *HBA2* on chromosome 16p13.3 lead to varying degrees of severity of the  $\alpha$ -thalassemia trait. The most common deletion alleles of  $\alpha$ -thalassemia are found in up to 80% of the individuals in some Middle Eastern, sub-Saharan African and South East Asian populations. In the most severe form, called Hb Barts hydrops fetalis syndrome, fetuses lacking all copies of the  $\alpha$ -globin genes cause serious health complications for women during pregnancy and are usually stillborn. While treatments for managing disease symptoms in adults are available, high-throughput, reliable and cost-effective genetic carrier detection is critical to alleviating the burden of the disease during pregnancy and childbirth. Methods developed for quantifying the number of copies of *HBA1* and *HBA2* genes for the Applied Biosystems™ CarrierScan™ Assay have identified 15 previously undetected  $\alpha$ -thalassemia carriers in research samples obtained from the Coriell Institute for Medical Research Cell Repository. Among 96 samples tested during product development, a loss of copy in the *HBA2* gene was detected in 11 samples, and a gain in copy in 4 samples. Two additional samples, previously annotated as carriers, were also detected. The copy number status of the samples was verified by multiplex ligation-dependent probe amplification (MLPA) at Mayo Clinic and Gene by Gene. Additionally, the remaining 79 samples were confirmed as carrying 2 copies each of *HBA1* and *HBA2* genes, indicating that the algorithm performed with 100% analytical sensitivity and specificity in this verification study. Samples analyzed by the Genomics Laboratory at Mayo Clinic identified an additional 27  $\alpha$ -thalassemia carriers across 4 96- format CarrierScan plates; the results were confirmed with MLPA, showing 100% concordance between the assays.

The CarrierScan assay is a microarray-based technology that detects sequence and structural variants in approximately 600 genes. Deletions and duplications at the gene or exon level are identified by calculating a copy-number score across selected copy-number-responsive probesets, including those that target the  $\alpha$ -globin gene loci. This provides important information that may assist in research evaluating  $\alpha$ -thalassemia copy number variations across populations.

## PgmNr 587: Genome-wide association study of female infertility in >36,000 women.

### Authors:

G. Galarneau <sup>1</sup>; P. Fontanillas <sup>2</sup>; C. Clementi <sup>1</sup>; D.E. Parfitt <sup>1</sup>; J.Y. Tung <sup>2</sup>; P. Yurttas Beim <sup>1</sup>; the 23andMe Research Team, the Celmatix Research Team

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Celmatix, New York, NY; 2) 23andMe, Inc., Mountain View, CA

On average, women of reproductive age having timed intercourse are able to establish pregnancy within 6 months, but ~15% take longer than a year or are unable to conceive. We conducted the first genome-wide association study (GWAS) on female infertility in women of European ancestry who used 23andMe's genetic services and consented to participate in research. Cases were women who reported trying to conceive for  $\geq 13$  months in their first pregnancy attempt ( $n=9,822$ ). Controls were women with  $\geq 1$  biological child who reported conception in  $< 6$  months in their first pregnancy attempt ( $n=26,947$ ). We used LDHub to identify genetically correlated traits and MAGMA to perform a gene-set analysis.

Two loci showed genome-wide association. The first locus includes *CDC42* and *WNT4* ( $rs61768001$ ,  $p=4.6 \times 10^{-10}$ ,  $OR=1.16$ , 95% CI [1.11, 1.22]). *CDC42* controls the activation of primordial follicles and *WNT4* is a regulator of endometrial decidualization. The second locus is ~25 kb upstream of *FSHB* ( $rs11031006$ ,  $p=3.6 \times 10^{-8}$ ,  $OR=1.14$ , 95% CI [1.09, 1.19]). *FSHB* encodes the  $\beta$ -subunit of follicle stimulating hormone (FSH). The variants  $rs11031006$  and  $rs10835638$ , both in the credible SNP set, have been associated with nulliparity and lifetime parity. The variant  $rs10835638$  is in the *FSHB* promoter within an 11-bp binding site of the LHX3 transcription factor. The  $rs10835638$ -T allele has been found to reduce the binding affinity of LHX3 and reduces *FSHB* transcription in gonadotrope cells. Six additional loci showed suggestive association ( $p < 1 \times 10^{-6}$ ): *GLYR1*, *FAM168A*, 2q14.3, *HLA-A-TRIM26*, *DUSP10-HHIPL2* and *THADA-PLEKHH2*. Two of these, *GLYR1* and *FAM168A*, include genes that interact with the FSH signaling pathway, which induces ovarian follicles maturation. *GLYR1* mediates the stress activation of MAPK14, which is involved in oocyte meiotic maturation. *FAM168A* activates AKT1, which initiates oocyte growth.

The most genetically correlated traits with female infertility were number of live births ( $p=0.0002$ ,  $r_g=-0.35 \pm 0.10$ ) and lung forced vital capacity ( $p=0.0017$ ,  $r_g=-0.18 \pm 0.06$ ). The most enriched gene-set was NF- $\kappa$ B-driven pro-inflammatory genes that are negatively regulated by glucocorticoids ( $p=3.4 \times 10^{-5}$ ). Interestingly, our study implicates multiple pathways and processes known to be important for reproduction in delayed time-to-first-conception and infertility. This subclinical insight will be key to earlier detection, more efficient patient journeys, and novel therapeutics.



## PgmNr 588: Genetic evidence on the role of maternal vitamin D status during pregnancy in obstetric outcomes.

### Authors:

M.C. Borges <sup>1,2</sup>; A.L.G. Soares <sup>1,2</sup>; Q. Yang <sup>1,2</sup>; F. Kilpi <sup>1,2</sup>; M. Taylor <sup>1,2</sup>; J. Brand <sup>1,3</sup>; D.A. Lawlor <sup>1,2</sup>; MR-PREG Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, United Kingdom; 2) Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom; 3) Clinical Epidemiology and Biostatistics, School of Medical Sciences, Örebro University, Örebro, Sweden

### Background/Aims

Vitamin D deficiency or insufficiency is common in pregnancy. However, it is unclear whether vitamin D supplementation should be an integral part of routine antenatal care given the scarcity of evidence from high-quality randomized controlled trials (RCTs). Mendelian randomization is a method that uses genetic variants robustly associated with modifiable exposures to generate more reliable evidence regarding which interventions should produce health benefits. We have used Mendelian randomization to interrogate the effect of maternal vitamin D status during pregnancy on several obstetric outcomes (i.e. miscarriage, stillbirth, need of labour induction, C-section, gestational diabetes and birth weight).

### Methods

We selected six independent genetic variants strongly associated with vitamin D status, proxied by circulating 25-hydroxyvitamin D (25(OH)D), in the general population. These genetic variants were tested for replication in an independent cohort of pregnant women (N = 4,819 from ALSPAC study). Mendelian randomization analysis was conducted using inverse variance weighted method in up to 207,565 women from UK Biobank. Sensitivity analyses were performed to explore bias due to invalid genetic variants. Results are expressed as odds ratio (OR) or mean difference in outcome per unit increase in natural log 25(OH)D.

### Results

The associations of genetic variants with 25(OH)D largely replicated in pregnant women. In the Mendelian randomization analysis, there was no clear evidence that maternal 25(OH)D concentration was related to the incidence of miscarriage (OR=1.03; 95%CI: 0.88, 1.22; P=0.70), stillbirth (OR=1.31; 95%CI: 0.80, 2.13; P=0.28), need of induction (OR=0.93; 95%CI: 0.44, 1.97; P=0.85), C-section (OR=0.93; 95%CI: 0.44, 1.97; P=0.85), gestational diabetes (OR=1.43; 95%CI: 0.37, 5.53; P=0.60), or birth weight (0.03 kg, 95%CI: -0.01, 0.07, P=0.15). However, effect estimates were imprecise for several outcomes.

### Conclusion

Overall, there was no strong evidence that maternal 25(OH)D status affects any of the assessed obstetric outcomes in the target population. We have requested and will soon receive data from several additional studies to increase power and to include a broader range of outcomes in our



analyses. Mendelian randomization can be a valuable tool to probe the putative effect of pregnancy exposures on obstetric outcomes and to inform clinical practice and public health policies in the absence of high-quality RCTs.

## PgmNr 589: Placental microRNAs in the regulation of fetal growth: A meta-analysis of two independent mother-infant cohorts.

### Authors:

E.M. Kennedy <sup>1</sup>; K. Hermetz <sup>1</sup>; A. Burt <sup>1</sup>; T.M. Everson <sup>1</sup>; K. Hao <sup>2</sup>; J. Chen <sup>3</sup>; M. Karagas <sup>4</sup>; D. Pei <sup>5</sup>; D.C. Koestler <sup>5</sup>; C.J. Marsit <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA; 2) Department of Genetics and Genome Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Lebanon, NH; 5) Department of Biostatistics & Data Science, University of Kansas Medical Center, Kansas City, KS

---

The placenta is the master regulator of fetal growth and development. During implantation, the trophoblast invades the extracellular matrix (ECM) of the decidualized endometrium into the uterine wall. The trophoblast then differentiates and proliferates into a network of branching villi – the functional units of the placenta that facilitate exchange of gases, nutrients and waste for the growing embryo. Poor placental function is the most common cause of intrauterine growth restriction, which is associated with perinatal morbidity, mortality and long-term health outcomes. MicroRNAs (miRNA) are non-coding RNAs that regulate gene expression through targeted silencing of target mRNAs, based on sequence complementarity. Placental miRNAs associate with fetal growth, and are linked to the basic processes of trophoblast biology, but the pathways and molecular mechanisms by which miRNAs regulate birth weight are not fully known. To assess the role of miRNAs in placental function and fetal development, we have utilized small RNA sequencing to profile placental miRNAs from two independent mother-infant cohorts: the New Hampshire Birth Cohort Study (n=317) and Rhode Island Child Health Study (n=225). We modeled miRNA counts on infant birth weight percentile in each cohort, while accounting for race, sex, parity and technical factors, using negative binomial generalized linear models. We aggregated cohort-specific results via inverse variance weighted fixed-effects meta-analysis. We identified 62 miRNAs that were differentially expressed (DEmiRs) with birth weight percentile at false discovery rate (FDR)<0.05. Our DEmiRs were predicted to target 1130 mRNAs, among which, 90 were predicted mRNA targets of more than two DEmiRs. These 90 targets are enriched for pathways involved in ECM organization, including collagen formation (FDR=4.4e-4) and degradation (FDR=6.9e-6). Our findings support the idea that placental miRNAs control trophoblast proliferation and differentiation, apoptosis, cellular metabolism, as well as vasculo- and angio-genesis. Further, our results suggest that placental miRNAs are involved in the metabolism of the ECM protein, collagen, which may influence the invasive behavior of trophoblasts during implantation. In the largest RNA-Seq based, genome-wide miRNA profiling study conducted in human placenta to date, we implicate potential molecular mechanisms underlying the role of miRNAs in placental development, placental function and fetal growth.

## PgmNr 590: Increasing incidences of musculoskeletal abnormalities in a large population of boys with 49,XXXXY.

### Authors:

G. Porter <sup>1</sup>; C. Samango-Sprouse <sup>1,2,3</sup>; P. Lasutschinkow <sup>1</sup>; T. Sadeghin <sup>1</sup>; A. Gropman <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Focus Foundation, Davidsonville, MD.; 2) George Washington University, Department of Pediatrics, Washington, DC; 3) Florida International University, Department of Human and Molecular Genetics, Miami, FL; 4) Children's National Medical Center, Washington, DC

49,XXXXY is the rarest chromosomal variation, with 1:80,000-100,000 live male births. Case studies indicate increased incidences of musculoskeletal abnormalities in these boys, including knee subluxation, congenital talipes equinovarus, scoliosis, kyphosis, and genu valga. These issues, combined with testosterone deficiency, may lead to decreased participation in physical activity and additional complications to health and well being. The elucidation of the orthopedic profile of 49,XXXXY is necessary in order to provide health care providers with accurate information on which to base the development of care and treatment plans.

This study collected data from an international cohort of more than 100 boys with 49,XXXXY over ten years. The data was collected based on evaluations by a multidisciplinary team consisting of a pediatric orthopedist and two physical therapists. The data was segregated into three categories: under 5 years, 5-10 years, and over 10 years old. Participants with a copy number variant or mosaicism were excluded.

Boys with 49,XXXXY presented with increased incidences of torticollis (29.2%), scoliosis (20.1%), leg length discrepancy (35.9%), gait abnormalities (83.0%), genu valgum (39.1%), hamstring tightness (41.2%), ligamentous laxity (67.4%), radioulnar synostosis (66.2%), clinodactyly (66.7%), abnormal shoulder range of motion (36.8%), gunstock deformity (21.6%), foot pronation (42.4%) and eversion (22.8%), pes planus (25.8%), tight heel cords (21.3%), and feet abnormalities (87.9%).

The incidences of these anomalies appeared to increase with age, specifically scoliosis, kyphosis, number of hip surgeries, leg length discrepancy, gait abnormality, hamstring tightness, ligamentous laxity, pes cavus, and club foot.

These results indicate the high rate of musculoskeletal anomalies in boys with 49,XXXXY, and the increasing incidence of these abnormalities with age. Such findings may indicate that orthopedic issues become more complicated with time, and these boys will benefit from early targeted treatment on specific vulnerable areas. Given the rare incidence of this disorder, the investigation of associated clinical symptoms is essential to determine and establish appropriate care and treatment. This research further supports the necessity for the comprehensive treatment of boys with 49,XXXXY by a multidisciplinary team of specialists, including, among others, a pediatric orthopedist, pediatric physical therapist, and occupational therapist.

## PgmNr 591: Cell-free DNA analysis in twin pregnancies after reduction because of fetal aneuploidy.

### Authors:

K.M. Chen <sup>1</sup>; E. Bevilacqua <sup>2</sup>; K. White <sup>1</sup>; R. Stokowski <sup>1</sup>; H. Chuang <sup>1</sup>; J. Doshi <sup>1</sup>; Y. Wang <sup>1</sup>; L. Kunz <sup>1</sup>; M. Schmid <sup>1</sup>; J. Jani <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ariosa Diagnostics Inc., Roche Sequencing Solutions, San Jose, CA. 5945 Optical Court, San Jose, CA 95136; 2) Department of Obstetrics and Gynecology, University Hospital Brugmann, Université Libre de Bruxelles, Brussels, Belgium

---

**Background:** In multifetal pregnancies, cell-free DNA from a non-viable conception may lead to a non-invasive prenatal test result that is discordant with the viable fetus. However, few studies have examined changes in cell-free DNA over time in pregnancies with an aneuploid fetal demise.

**Objectives:** Using elective fetal reduction as a model for natural demise, we examine the results of targeted cell-free DNA analysis at multiple points after reduction of an aneuploid twin.

**Methods:** Two twin pregnancies, each with one aneuploid fetus (one trisomy 21 and one monosomy X), were included in the study. Maternal blood was collected at multiple time points prior to and following reduction and studied using targeted cell-free DNA analysis. Proportion values were calculated for chromosomes 21, X, and Y. The chromosome 21 and X aneuploidy values represent the relative signal from the chromosome of interest compared to that expected in disomy. The chromosome Y value represents the presence of Y signal. Fetal fraction was also determined for all samples.

**Results:** In the first pregnancy with a male fetus with trisomy 21 and one female euploid fetus, the aneuploidy values from chromosome 21 moved towards the expectation for euploidy over 13 weeks following reduction. The chromosome Y value followed a similar pattern. In the second pregnancy with a fetus with monosomy X, the aneuploidy value from chromosome X moved towards the expectation for euploidy over the 8 weeks following reduction. However, the trends were not linear: both cases showed a transient increase in the level of aneuploidy 3 to 8 weeks following reduction. Fetal fraction also increased and then decreased again over several weeks. Notably, the aneuploidy values never reached 0 (the euploid state) even months after the reduction.

**Conclusions:** These study results demonstrate the complexity of cell-free DNA analysis in multifetal pregnancies with reduction. The unpredictable patterns for both aneuploidy and fetal fraction after fetal reduction do not support cell-free DNA testing at any given interval following a fetal demise.

## PgmNr 592: Cell-free DNA testing for prenatal aneuploidy assessment: An analysis of international professional society statements.

### Authors:

R. Jones; K. White; A. Batey; E. Kostenko; M. Schmid

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medical and Scientific Affairs, Roche Sequencing Solutions Inc., San Jose, CA

---

**Introduction:** Prenatal assessment for fetal aneuploidy using cell-free DNA analysis (also called non-invasive prenatal testing or NIPT) is clinically available throughout the world. The objective of this study was to obtain a global perspective on approaches to clinical implementation by reviewing practice guidelines and recommendations published by international and national professional societies.

**Methods:** Position statements and practice recommendations were located by searching PubMed, Embase, and examining public websites available in English. Key content and themes were captured from each document.

**Results:** Twenty documents were identified from international as well as national societies from Europe, the Americas, the Middle East, Asia and Pacific regions. Statements were from 28 organizations representing medical genetics (9), obstetrics (15), prenatal ultrasound (5) and counseling (1), with some societies representing more than one area. The statements were consistent in stating: NIPT is not diagnostic (20), NIPT is an accurate screening test (20), and patient pre and post-test counseling, including test features and limitations, is recommended (19). Differences were found in three areas: which patients should be offered NIPT, what fetal conditions should be on the test panel, and what information should be on the test report. Fourteen of twenty documents focus on NIPT for patients with an increased risk for fetal aneuploidy. Eight of the statements have been issued or revised since 2017. Of 13 statements that addressed expanded panels, 9 societies expressly say that cell-free DNA screening should not be used for copy-number variants or whole genome analysis. Test report recommendations were found in 4 of 20 professional statements.

**Conclusions:** Professional societies involved in prenatal care have different recommendations regarding the incorporation of NIPT into clinical practice. Although societies agree that NIPT has high accuracy for prenatal risk assessment of common aneuploidies, confirmation of 'high-risk' results with diagnostic testing is critical and patient counseling is important, differences exist around which patients should be offered NIPT, what fetal conditions beyond common aneuploidies should be added to the testing panel and what information should be included on the test report. Less than half of the statements have been issued or revised since 2017.

## PgmNr 593: Clinical performance of noninvasive prenatal testing in high risk pregnancies.

### Authors:

Zhu <sup>1,2</sup>; H. Yan <sup>3</sup>; Y. Guo <sup>4</sup>; H. Wang <sup>5</sup>; K. Kwok <sup>1</sup>; M. Chau <sup>1,2</sup>; J. Chen <sup>3</sup>; X. Kong <sup>4</sup>; M. Chen <sup>3</sup>; K. Choy <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong SAR; 2) Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China; 3) Department of Fetal Medicine and Prenatal Diagnosis, the Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, 510150, China; 4) Genetics and Prenatal Diagnosis Center, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, 450052, China; 5) Baoan Women's and Children's Health hospital, Jinan University, Shenzhen, 518101, China

### Background

Noninvasive prenatal testing (NIPT) has greatly advanced prenatal screening for aneuploidies since its implementation for clinical use. Genome-wide NIPT allows for screening of rare aneuploidies and copy number variants (CNVs), but the clinical performance of such expansion remains unknown and invasive diagnostic testing is warranted for confirmation of the positive results. Chromosomal microarray analysis (CMA) shows superiority over conventional karyotyping in detecting a wide spectrum of chromosome abnormalities including pathogenic CNVs through the genome. In this study, we aimed to evaluate the clinical performance of genome-wide NIPT in detecting chromosomal abnormalities in high risk pregnancies by comparison with CMA, so that we can provide valuable information for genetic counselling and clinical management of high risk pregnancies.

### Methods

Pregnant women with high risk for chromosome abnormalities were retrospectively enrolled from four prenatal diagnosis centers between 2015 and 2018. Pregnancies with both NIPT and CMA performed during pregnancies were included for further analysis. The results were categorized into common trisomies, sex chromosome abnormalities, rare aneuploidies and CNVs. Positive predictive value (PPV) and false positive rate (FPR) were calculated, and the discordant results between NIPT and CMA were also analyzed.

### Results

A cohort of 675 singleton pregnancies were included in this study. Of these, the majority (74.7%, 504/675) were referred by advanced maternal age. Overall, 77.8% (525/675) of the cases were reported to be positive by NIPT. A positive CMA result was detected in 278 of 675 (41.2%) cases. Among NIPT positive cases 43.0% (226/526) were confirmed to be true positives. PPV estimates for rare aneuploidies and CNVs of NIPT were 17.5% and 33.9%, respectively. Importantly, 62 (of 675, 9.2%) additional chromosomal aberrations were detected by CMA but not reported by NIPT, including 45 CNVs smaller than 5Mb, 12 CNVs larger than 5Mb, and other 5 whole chromosomal abnormalities (Trisomy 18, 47,XXY, mosaic Monosomy 9 and 22, triploidy, uniparental disomy 4).

### Conclusion

NIPT shows high false positive rates for the detection of rare aneuploidies and CNVs in high risk pregnancies, which will increase potentially unnecessary invasive confirmatory testing and maternal anxiety. CMA is able to provide comprehensive genetic information and should be offered for high risk pregnancies undergoing invasive diagnostic procedures.

## PgmNr 594: Combined preimplantation genetic testing for *CFTR* gene mutations and aneuploidy using a single WGA PCR and amplicon panel-specific enrichment.

### Authors:

M. Jasper <sup>1</sup>; C. Robinson <sup>1</sup>; J. Nehyba <sup>2</sup>; R. Hrdlickova <sup>2</sup>; A. Kothandaraman <sup>2</sup>; S. Myers <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) PerkinElmer Health Sciences (Australia) Pty Ltd, Thebarton, Australia; 2) PerkinElmer, Inc., Austin, TX

Cystic Fibrosis is one of the most common indications reported for Preimplantation Genetic Testing (PGT) for monogenic disorders (PGT-M) (ESHRE PGD data collection). The ability to target the *Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)* gene with PCR primers during whole genome amplification (WGA), achieving concurrent PGT-M and PGT for aneuploidy (PGT-A) in a single PCR reaction represents a unique approach to maximise the screening opportunity for a single embryo biopsy.

The aim of this study was to demonstrate a novel protocol for concurrent PGT-A and PGT-M in a single PCR, using a *CFTR*-specific primer panel in combination with WGA, and analysed by Next Generation Sequencing (NGS).

Five-cell samples sorted from a 48,XXY,+21 cell line (Coriell Institute) were subjected to WGA using the DOPlify® TSE (Target Sequence Enrichment) protocol (PerkinElmer) with the inclusion of either NEXTFLEX® Cystic Fibrosis Amplicon Panel kit primer mix 1 or 2 (PerkinElmer). Further enrichment of the 31 *CFTR* target regions was achieved following reamplification of the enriched WGA using *CFTR*-specific PCR with the inclusion of gDNA (control). The *CFTR* amplicons generated were subsequently pooled with the original TSE WGA to allow single sample indexing (PG-Seq™ kit, PerkinElmer) and MiSeq® instrument (Illumina®) sequencing.

Whole chromosome aneuploidy concordant with the expected sample karyotype was achieved and 68% (23/31) of *CFTR* target regions were represented for 5-cell samples, with read depths of up to 200x. Of the 8 most common *CFTR* mutations, accounting for 75% of all mutations (CF Genetic Analysis Consortium, 1994), all were amplified from gDNA and 4/8 from 5-cell samples. An additional 13 Single Nucleotide Polymorphisms (SNPs) were identified that could potentially be used for linkage analysis.

The DOPlify® WGA kit provides a flexible technology to amplify and sequence whole genomes and relevant sequences in parallel using PerkinElmer's TSE protocol, offering a streamlined solution for combined PGT-M and PGT-A. The system accommodates the addition of multiple primer sets targeting specific sequences, allowing numerous sites to be analysed at once. Further optimisation of primer concentration and pooling will achieve improved coverage for all target regions. The workflow also provides an opportunity to generate PGT results using a variety of different NGS indexing and or NGS platform approaches.

For research use only. Not for use in diagnostic procedures.



## PgmNr 595: Discordant NIPT (non invasive prenatal test) result: Clinical review and validation.

### Authors:

D. Cha <sup>1,2</sup>; J.Y. Kim <sup>1</sup>; S.H. Shim <sup>2</sup>; S.Y. Hong <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Dept OB/GYN, Gangnam CHA Hosp, Seoul, Korea, Republic of; 2) Genetic Laboratory, Gangnam CHA Hosp. Seoul, Korea, Republic of

---

**Objective :** To confirm the reliability of NIPT and to discuss about the discordant cases.

**Methods :** This test results in large single center from 1, 2015 to 12, 2018 were analyzed. The patients who referred from other hospital due to abnormal NIPT result were included. Karyotyping was done by CVS and amniocentesis. Microdeletion test was added in case of showing abnormal sonographic soft marker.

**Results :** Total 1607 single pregnancy cases who performed NIPT were enrolled. Among the 1562 cases(97.2%) reported as low risk, 1 case revealed as false negative which showed z-score 2.49 on chromosome 18 and finally diagnoses as true trisomy 18. Forty-two cases(2.6%) reported as high risk; Trisomy 21(n=22), trisomy 18(n=2), trisomy 13(n=1), monosomy X(n=4), sex chromosome anomaly including XXX, XXY(n=7), triploidy(n=1) and 22q11.2 deletion(n=1). In massive parallel sequencing method, other uncommon trisomy such as trisomy 7(n=1), 14(n=1) and 15(n=2) was incidentally detected. Thirty-four cytogenetic confirmation tests were done in NIPT high risk group. Among the 17 karyotype tests in NIPT trisomy 21 group, all 17 cases revealed as trisomy 21(100%). One karyotype test in NIPT trisomy 18 group showed trisomy 18(100%). In NIPT monosomy X group, 3 cases showed discordancy resulting as normal karyotype and 1 case was mos45,X[14]/46,X,i(X)(q10)[6] which regarded as monosomy X(25%). The 22q11.2 deletion case had no sonographic cardiac anomaly and confirmed as false positive. Three 47,XXX and 3 46,XY were diagnosed in sex chromosome high risk group(50%). The NIPT result of triploidy was due to effect of vanished fetus in DCDA twin pregnancy. Both trisomy 15 NIPT cases showed discordancy but interestingly one of the cases was finally diagnoses as maternal uniparental disomy 15 which implies the further diverse way of interpretation in case of uncommon trisomy in NIPT. There were 3 uninformative results(0.2%); 1 case due to low fetal fraction, 1 case due to repeated CNVs on X chromosome which finally diagnosed as exon 1-29 duplication of dystrophin gene on Xq21.2 and 1 case with unknown cause.

**Conclusion :** In the present study, clinical review and validation of both discordant and general cases are introduced based on single institution's results. As the use of NIPT as a prior screening test increases, the systematic discussion of discordant cases should be conducted between the laboratories.



## PgmNr 596: Spectrum and features of clinically significant copy number variants in prenatal diagnosis: Implications on genome-wide noninvasive prenatal testing.

### Authors:

M.H.K. Chau <sup>1</sup>; Y. Cao <sup>1</sup>; Y.K. Kwok <sup>1</sup>; S. Chan <sup>2</sup>; Y.M. Chan <sup>1</sup>; Z. Yang <sup>1,4</sup>; H. Wang <sup>1,3,4</sup>; H.K. Wong <sup>1</sup>; T.Y. Leung <sup>1,5</sup>; K.W. Choy <sup>1,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, China; 2) Warwick Medical School at the University of Warwick, Coventry, UK; 3) Department of Central Laboratory, Bao'an Maternity and Child Healthcare Hospital, Jinan University School of Medicine, Key Laboratory of Birth Defects Research, Birth Defects Prevention Research and Transformation Team, Shenzhen, China; 4) Shenzhen Research Institute, Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Shenzhen, China; 5) The Chinese University of Hong Kong-Baylor College of Medicine Joint Center for Medical Genetics, Hong Kong, China

**Background:** Chromosomal microarray analysis (CMA) is widely utilized for the detection of pathogenic copy number variants (pCNVs) in prenatal diagnosis. Implementation and widespread use have expanded our understanding, curation and delineation of CNVs and the related genomic disorders. On the other hand, noninvasive prenatal testing (NIPT) is well established for the screening of common aneuploidies. More recently, efforts have been made to expand the scope of NIPT for the detection of fetal microdeletion syndromes. Expanding NIPT for microdeletions and microduplications is controversial as there are no consensual recommendations by professional societies. The spectrum and features of pCNVs including the size, genomic distribution and mode of inheritance are important considerations before genome-wide NIPT is implemented into clinical practice.

**Methods:** We retrospectively analyzed the pCNVs (<10Mb) in 1,510 pregnancies with standard indications for prenatal diagnosis from Jan 2012 to Dec 2017 at our prenatal genetic diagnosis laboratory. We have combined pCNV data from 14 published studies including of 22,355 fetuses undergoing CMA testing. Summarizing this data, we describe the spectrum and features of pCNVs including the prevalence, size, and mode of inheritance.

**Results:** Overall, 375/23,865 (1.6%) fetuses carried pCNVs (<10Mb), and 44 (11.7%) fetuses involved 2 or more pCNVs. A total of 428 pCNVs were detected in these fetuses, of which 280 are deletions and 148 are duplications. 360 (84.1%) were less than 5Mb in size and 68 (15.9%) were between 5-10Mb. The incidence of carrying a pathogenic copy number variant in the high-risk group (abnormal ultrasound finding, high risk down's screening, family history, NIPT positive) is 1 in 36, and the low-risk group (advanced maternal age ≥35, maternal anxiety, others) is 1 in 125. Parental inheritance study results were available for 311 pCNVs, 204 (65.6%) occurred *de novo*, 71 (22.8%) were maternally inherited and 36 (11.6%) were paternally inherited.

**Conclusion:** Collectively, pCNVs are common in all pregnancies. High-risk pregnancies should be offered invasive testing with CMA for a comprehensive investigation. Detection limits on size, genomic distribution, mode of inheritance should be carefully considered before implementing copy number variant screening in genome-wide NIPT.

## PgmNr 597: The application of ACMG variant classification guidelines for the variant assessment of non-invasive prenatal sequencing (NIPS).

### Authors:

L. Lu <sup>1</sup>; Y.J. Jiang <sup>1</sup>; S. Gu <sup>2</sup>; J. Dong <sup>1</sup>; H.Y. Cheng <sup>2</sup>; C. Gijavanekar <sup>1</sup>; Z. Chen <sup>1,2</sup>; Y. Wang <sup>1,2</sup>; S. Mullegama <sup>1,2</sup>; J. Xiao <sup>1</sup>; L. Wang <sup>1</sup>; J. Yang <sup>1</sup>; H.Z. Dai <sup>1,2</sup>; L.Y. Meng <sup>1,2</sup>; E. Schmitt <sup>1</sup>; S. Peacock <sup>1</sup>; G. Wang <sup>1</sup>; C. Eng <sup>1,2</sup>; J.L. Zhang <sup>1,2</sup>; F. Xia <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor Genetics 2450 Holcombe Blvd, Houston, TX 77021; 2) Baylor College of Medicine One Baylor Plaza, Houston, TX 77030

---

**Background** Classification of prenatally detected sequence variants plays a critical role in pregnancy management. Limited fetal phenotypic data and maternal admixture complicate NIPS analysis. To share experience and help improve data analysis in NIPS, genetic variants detected from a NIPS for single or multiple dominant monogenic disorders were retrospectively assessed for the performance of ACMG guidelines in this test. **Methods** A total of 73 variants with potential pathogenicity in 19 dominant monogenic disease-causing genes (*BRAF*, *CBL*, *CDKL5*, *CHD7*, *COL1A1/2*, *FGFR2/3*, *JAG1*, *KRAS*, *MAP2K2*, *MECP2*, *NIPBL*, *PTPN11*, *RAF1*, *RIT1*, *SOS1*, *TSC1/2*) were evaluated and classified by following ACMG guidelines. The strength of certain criteria was assessed and tuned for more accurate as well as efficient data interpretation. **Results** By following the ACMG guidelines, 52 variants were classified as pathogenic (P), 9 were classified as likely pathogenic (LP), and 12 were variants of uncertain significance (VUS). With certain criteria (e.g. number of patients reported in literature, applicable fetal/familial phenotypic evidence, etc.) tuned, however, 8 likely pathogenic variants could be upgraded to pathogenic, which included four that had multiple patient reports and were well-recognized as disease-causing variants. Reasonable specificity of ultrasound indications was required to fully appreciate *de novo* variants' pathogenicity. With 64.4% (n=47/73) of variants in this NIPS were *de novo*, the specificity of fetal ultrasound information impacted significantly on the accuracy of NIPS variants' classification. The importance of the phenotypic information was also reflected by the 12 VUSs that either due to early gestational ages (14.1±3.9 weeks) that precluded sonographic findings or maternal mosaicism compounded with the lack of clinical and/or scientific evidences, or evidences which were not concordant with the disease caused by the gene bearing the variant under assessment. **Conclusion** ACMG guidelines are helpful in the NIPS variant classification; however, given the nature of this test and the targeted genes, certain criteria need fine-tuning. Reasonably specific fetal ultrasound information at appropriate gestational ages, and the family medical histories significantly impact classification accuracy.

## PgmNr 598: A retrospective study of the application of chromosomal microarray analysis in prenatal diagnosis.

### Authors:

Zhu; L. Gu; W. Liu; L. He; Y. Zhu; Y. Yang; X. Wu; P. Cao; J. Li

[View Session](#) [Add to Schedule](#)

**Affiliation:** Nanjing Drum Tower Hospital, Nanjing, China

---

During the last decade, Chromosomal microarray analysis (CMA) was recommended by a few communities in different countries as a first-tier technique for genetic follow up of fetal structural anomalies identified by ultrasonography. With the accumulation of cases and experiences, it's the time to take a second look on the application and consider if we are abusing this technique. In this retrospective study, we present a comprehensive overview of 2363 singleton prenatal cases out of 4266 cases which were referred for CMA during 2014 to 2018.

The overall pathogenetic findings account for 10.41% (246/2363) of the cases, among which 49.6% (122/246) could be detected by karyotyping analysis. In those chromosomal normal cases, CMA made a further 5.53% yield. These data are perfectly in accordance with previous perspective researches. Multiple abnormalities (37%), pleural effusion (25%), cystic hygroma (23.5%) and gastral intestinal abnormalities (19.35%) were most associated with genetic abnormalities. For submicroscopic anomalies, multiple abnormalities (15.1%), pleural effusion (14.3%), gastral intestinal abnormalities (10.7%) and urinary system abnormalities (9.1%) were most relevant indications. The most striking finding is that 41.67% (6/13) cases of echogenic kidney(s) with or without size change were associated with pathogenic copy number changes (CNV). Another interesting finding is the high detection rate (50.8%) of variance of uncertain significance in fetus with growth retardation and other minor ultrasound findings. On the other hand, if there's no clinical indication for prenatal diagnosis, little benefit would be obtained from CMA, with only 1 pathogenic CNV in 110 cases.

## PgmNr 599: Disorders of sex development in a multigenerational family: Oligogenic causes, reduced penetrance and blended phenotypes with combined pituitary hormone deficiency.

### Authors:

M. Laan <sup>1</sup>; L. Kasak <sup>1</sup>; M. Grigorova <sup>1</sup>; E. Laasik <sup>1</sup>; E. Koppel <sup>1</sup>; M. Punab <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia; 2) Andrology Center, Tartu University Hospital, Tartu, Estonia

---

Disorders of sex development (DSD) represent a heterogeneous group of congenital genital anomalies. We report a family presenting DSD in three generations. The three probands were first cousins (P1, male, XY, 13 yrs; P2, male, XY, 18 yrs) diagnosed with cryptorchidism, severe penoscrotal hypospadias, small testicles, azoospermia; and their joint first cousin once removed (P3, male, XY, 30 yrs) with cryptorchidism and cryptozoospermia. Familial anamnesis also included other reproductive disorders, such as cases with ambiguous sex, prostate cancer, stillbirth, induced termination of mid-pregnancy and diagnosed premature placental ageing. Whole-exome sequencing (WES) of eight family members detected a splicing variant in *NR5A1* gene (c.G991-1C; p.Val331Ilefs) as the most likely cause of DSD in P1 and P2. The variant is not reported in human genome databases, but was shown in a Belgian patient with cryptorchidism (Robevska et al 2018, Hum Mutat). As several family members were asymptomatic carriers, the estimated penetrance rate was < 50%. In depth WES data analysis of P1-P2 and their parents uncovered oligogenic contribution as a potential modulator of the penetrance of the *NR5A1* variant. Both, P1 and P2 had inherited rare variants in genes implicated in combined pituitary hormone deficiency (CPHD) – *OTX2* p.Pro134Arg (P2), *PROP1* p.Leu102Cysfs and *GLI2* p.Arg131His (P1) from the other parent. Hormonal profiling supported the blended phenotype of DSD and CPHD in this family. P3 carried neither the *NR5A1* splicing variant nor any of the CPHD variants. The highest prioritized candidate variant responsible for his condition was p.Glu60Ter in the *DMRT2* gene located in the genomic region linked to gonadal dysgenesis. Due to complex family history, WES data were also screened for secondary findings. P2, P3 and other family members were detected as the carriers of the p.Ile157Thr variant of *CHEK2*, implicated in familial prostate cancer. The grandfather of P1-P2 and father of P3, who had been diagnosed with prostate cancer, were identified as the obligate carriers of the *CHEK2* variant. Our study demonstrates heterogeneity and oligogenic contribution to the penetrance of DSD mutations and the value of in-depth WES analysis in the pedigree to uncover the genetics of multiple segregating clinical conditions. Detailed profiling of pathogenic variants facilitates high quality counselling of index patients and their family members. Grant: Estonian Research Council IUT34-12.

## PgmNr 600: Insights from the largest genetic study of sporadic and recurrent miscarriage.

### Authors:

T. Laisk-Podar<sup>1,2</sup>; International Miscarriage Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia; 2) Department of Obstetrics and Gynecology, University of Tartu, Tartu, Estonia

---

**Introduction:** Miscarriage is a common complex trait that affects 10-25% of all clinically confirmed pregnancies. Miscarriage is associated with excessive bleeding, infection, depression, infertility and an increased lifetime risk of cardiovascular disease. A variety of predisposing factors have been associated with increased miscarriage risk, and it has a genetic component that thus far has remained poorly characterized. Our aim was to discover and map the maternal genetic susceptibility and underlying biology of sporadic and recurrent miscarriage.

**Materials and methods:** We performed a genome wide association study including up to 69,054 sporadic miscarriage cases from seven different ancestries, 750 recurrent miscarriage cases of European ancestry, and up to 359,469 female controls. We investigated the genetic architecture of miscarriage with biobank-scale Mendelian randomization, heritability, genetic correlation and functional annotation analyses.

**Results:** We identified one and four loci significantly ( $p < 5 \times 10^{-8}$ ) associated with sporadic and recurrent miscarriage, respectively. Following functional annotation linked these associations with genes related to gonadotropin regulation, placental biology and progesterone production. We found a heritability of 29% (95%CI 20%-38%) for miscarriage. Consistent with observational associations, we found significant genetic correlations between sporadic miscarriage and number of children ( $p = 7.2 \times 10^{-9}$ ).

**Conclusions:** Our results confirm miscarriage as a complex partly heritable phenotype and implicate novel biology through regulation of genes involved in gonadotropin regulation, placental biology and progesterone production.

## PgmNr 601: Maternal blood mRNA profiles associated with spontaneous preterm birth in maternity log study.

### Authors:

D. Ochi <sup>1</sup>; T. Yamauchi <sup>1</sup>; Y. Tsunemoto <sup>1</sup>; S. Kumatani <sup>1</sup>; R. Yamashita <sup>2</sup>; O. Tanabe <sup>3</sup>; S. Hiyama <sup>1</sup>; M. Nagasaki <sup>4</sup>; J. Sugawara <sup>5,6</sup>; Maternity Log Study Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NTT DOCOMO, INC., Yokosuka, Kanagawa, Japan; 2) National Cancer Center, Kashiwa, Chiba, Japan; 3) Radiation Effects Research Foundation, Hiroshima, Japan; 4) Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 5) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 6) Gynecology and Obstetrics, Tohoku University Graduate school of Medicine, Sendai, Miyagi, Japan

---

**Objective:** Despite the seriousness of pregnancy complications including preterm birth (PTB), their mechanisms of pathogenesis or progression have not been well elucidated. Pregnancy complications are multifactorial diseases caused by complex interactions of genetic and environmental factors. Messenger RNAs (mRNAs) in maternal blood can be affected by both factors. Several studies investigated maternal whole-blood gene expression profiles associated with spontaneous PTB (SPTB), *i.e.* non-induced delivery before 37 gestational weeks. However, these studies used microarray to analyze mRNA expression, and significance of their results were not clear. In this study, we used RNA-seq to investigate blood mRNA profiles associated with SPTB.

**Methods:** We used maternal peripheral blood samples collected in the Maternity Log Study, which is a prospective cohort study of pregnant Japanese women. In this study, pregnant women were recruited in the first or second trimester of pregnancy at Tohoku University Hospital. These samples were collected at the first term (between 12 and 23 gestational weeks) and the second term (between 24 and 34 gestational weeks) from 301 women using PAXgene Blood RNA Kit, and subjected to RNA-seq using the HiSeq sequencer. Multivariate logistic regression model was used to evaluate association between mRNA expression and SPTB.

**Results:** After correcting multiple testing by FDR, we identified over 900 genes which showed significantly different expression levels between the SPTB group (N=14) and the healthy control group (N=182) at both first and second terms. Especially, the expression of a gene associated with maternal placenta development was significantly higher in the SPTB group than in the control group.

**Discussions:** We identified maternal blood mRNA profiles associated with SPTB, indicating possible biomarkers for SPTB detectable in the early phase of pregnancy. Further research may lead to elucidation of molecular mechanisms and development of prevention measures for SPTB.

## **PgmNr 602: The impact of delayed evacuation after pregnancy termination on fetal tissue quality for diagnostic and research applications.**

### **Authors:**

D. O'Day <sup>1</sup>; E. Micks <sup>2</sup>; L. Benson <sup>2</sup>; M. Deng <sup>1</sup>; T. Naluai-Cecchini <sup>1</sup>; D. Doherty <sup>1</sup>; S. Prager <sup>2</sup>; I. Glass <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Pediatrics, Univ of Washington, Seattle, Washington.; 2) Department of Obstetrics and Gynecology, Univ of Washington, Seattle, Washington.

---

**Background:** In the United States, potassium chloride (KCl), digoxin and other agents are commonly used to arrest fetal circulation prior to pregnancy termination during the second trimester, largely for medico-legal and social reasons. The effects of induced demise prior to pregnancy termination on fetal tissue quality for subsequent diagnostic and research applications have not been characterized in detail. This is a critical knowledge gap, since high-quality fetal tissue is required for the accurate clinical diagnosis of fetal anomalies, and for biomedical research purposes.

**Objective:** To determine the effects of delayed evacuation after pregnancy termination on tissue quality for clinical and research purposes

**Study Design:** To compare tissue quality from second trimester fetuses terminated with and without injection of digoxin or KCl, 19-25 hours before dilation and evacuation (D&E), we obtained fetal tissue from terminations with and without prior injection of digoxin or KCl. We assessed: 1) cellular morphology by hematoxylin and eosin staining, 2) cell proliferation by Ki67 staining, 3) apoptosis by the TUNEL cell death assay, 4) tissue viability by cell culture, and 5) nucleic acid quality by integrity scores and the polymerase chain reaction (PCR).

**Results:** Tissues from fetuses after delayed evacuation demonstrated: 1) greater cellular morphology disruption, 2) decreased proliferation, 3) greater apoptosis, 4) diminished cell culture viability, 5) lower RNA quality; however, all samples yielded DNA of adequate quality for PCR.

**Conclusion:** Delayed evacuation for >18 hours is associated with fetal tissue quality unsuitable for most clinical and research applications except DNA diagnostics. Limiting the interval between arrest of fetal circulation and D&E could improve the quality of fetal tissue for clinical and research purposes.



## PgmNr 603: Transitions in transcriptome and composition of maternal peripheral blood through pregnancy.

### Authors:

L. Skotte<sup>1</sup>; B.D. Piening<sup>2</sup>; L. Liang<sup>2</sup>; J.P.S. Fadista<sup>1</sup>; C. Jiang<sup>2</sup>; F. Vallania<sup>3</sup>; M.-L.H. Rasmussen<sup>1</sup>; E. Wei<sup>4</sup>; Q. Liu<sup>2</sup>; H. Chaib<sup>4</sup>; F. Geller<sup>1</sup>; B. Feenstra<sup>1</sup>; M. Snyder<sup>2</sup>; M. Melbye<sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 3) Institute for Immunity, Stanford University School of Medicine, Stanford, CA; 4) Stanford Center for Genomics and Personalised Medicine, Stanford, CA; 5) Department of Medicine, Stanford University School of Medicine, Stanford, CA

---

**Introduction:** Throughout pregnancy the maternal body facilitates conditions allowing development and growth of the fetus. The maternal bloodstream plays an essential role not only in accommodating the increased metabolic demands and requirements for nutrients, but also in hormonal signaling and in a complex regulation of immunity. We studied the associated mechanisms in maternal blood through a detailed characterization of the transitions in the transcriptome across pregnancy.

**Methods:** We enrolled 31 pregnant women (at gestational age 4 – 25 weeks) and sampled peripheral blood on a weekly basis through pregnancy and once postpartum. We sequenced ribosome depleted total RNA in 531 discovery samples from 21 pregnant women and 271 replication samples from 10 pregnant women. The samples were sequenced at an average of 25 million reads per sample and in total the mapping (STAR v2.5.3a to GRCh38/GENCODEv28) yielded >13 billion uniquely mapped reads. This extremely dense sampling allowed analysis of changes in genetic expression at an unprecedented fine scale, carried out using a longitudinal mixed effects model.

**Results:** In a strict discovery/validation design, we identified 704 transcripts with robust pregnancy-specific expression patterns. We used expression correlation analysis to cluster these pregnancy-specific patterns in modules. These showed significant enrichment for multiple biological pathways, including Defense response to virus, T-cell activation, Defense response to bacterium/fungus, Hemoglobin metabolism, and Mast cell activation. Deconvolution into cell-types revealed consistent changes in cell populations, with 11 of 19 quantified cell-types showing robust alterations through pregnancy, including a steady increase in the fraction of hematopoietic progenitor cells up to week 35. Although these patterns demonstrate an altered regulation of the immune system during pregnancy, changes in the viral load through pregnancy were not detectable among RNA reads mapped to virus genomes for these 31 women.

**Conclusion:** We provide a comprehensive characterization of the changes in transcription and cell-type composition in maternal blood through pregnancy, which highlights its systemic role in healthy gestation. These data will be useful as a reference for future studies of transcription dynamics and changes in cell populations in pregnancies complicated by, e.g., preeclampsia, preterm delivery or gestational diabetes.



## PgmNr 604: Expanded carrier screening benefits as it relates to test enhancements and variant reclassification.

### Authors:

L. Aptekar<sup>1</sup>; D. Ilg<sup>1</sup>; E. Lagoe<sup>1</sup>; C. Sotile<sup>1</sup>; L. Elkhoury<sup>1,2</sup>; M.W. Dillon<sup>1,2</sup>; R. Zimmerman<sup>1,2</sup>; A. Birch<sup>1,2</sup>; R. Kornreich<sup>1,2</sup>; L. Edelmann<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sema4, a Mount Sinai venture, Stamford, CT; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029

---

The American College of Obstetrics and Gynecology (ACOG Committee Opinion 691) recommends discussing carrier screening options, including expanded carrier screening options (ECS) with all women who are pregnant or are considering pregnancy. With regards to repeat screening, the opinion states “the decision to rescreen a patient with previous carrier screening should be undertaken only with the guidance of a genetics professional who can best assess the incremental benefit of repeat testing for additional mutations.”

Sema4 began offering its comprehensive expanded carrier screening panel in 2015. The original panel screened for 281 genetic conditions (260 autosomal recessive disorders and 21 X-linked). In 2018, a newer version was launched that included multiple enhancements: the addition of 21-hydroxylase related congenital adrenal hyperplasia (*CYP21A2*), full gene sequencing for *HBA1* and *HBA2*, full gene sequencing for *GBA* (genotyping was performed previously) and the addition of copy number variant (CNV) detection via a bioinformatics algorithm of the NGS data. For patients with a history of testing on the previous platform, the option of an “upgrade”, which included repeat analysis of the NGS data (which was rerun both for the purposes of CNV detection and variant reclassification), as well as the other previously described enhancements, was offered.

Since June 2018, our lab has completed 847 upgrades for patients previously tested on the 281 platform. Of those, there were 143 new variants reported due to a new feature of the test (133 *CYP21A2* and 9 large deletions or duplications of other genes). Another 48 new variants were reported due to classification changes and 7 variants were removed due to downgrading.

From our current data, new information regarding carrier status was reported in 22% of individuals pursuing the upgrade, 6% being solely due to variant reclassification. Importantly, utilizing ancillary assays and technology to increase detection rates, 17 new carrier couples were identified. Post test telegenetic counseling is offered for positive ECS results through our laboratory, and as such, we include a discussion on how evolving genetic knowledge and testing impact family planning and pre and postnatal care as well as disease and carrier-state identity.

## PgmNr 605: Non-invasive prenatal testing using a novel, clinic-friendly digital PCR platform.

### Authors:

M. Dueck; R. Lin; L. Jiang; S. Gallagher; A. Zayac; P. Hung

[View Session](#) [Add to Schedule](#)

**Affiliation:** COMBiNATi, Burlingame, California.

---

Noninvasive prenatal genetic testing (NIPT) describes the detection of fetal chromosomal aneuploidies using cell-free fetal DNA found in the blood of a pregnant woman. The fetal fraction of the total maternal cell-free DNA (cfDNA) varies among pregnant women but ranges from 0.5% to 19%. In 2007 a group first showed the general feasibility of using digital PCR (dPCR) for NIPT, but their assay required an unreasonably high 25% fetal fraction in the cfDNA tested. Improvements in modern dPCR make it the new ideal technology for NIPT. To address the needs of using dPCR as a routine tool for NIPT, a novel dPCR platform was developed. The dPCR platform consists of a cost-effective, micro-molded plastic consumable and a fully-integrated instrument that combines consumable sample loading, thermal cycling and 5-color fluorescence detection. The consumable partitions a 10 microliter sample into 20,000 micromolded wells without requiring valves, films or any moving parts. The instrument is 100% dry, contamination-free and requires no routine maintenance. The platform achieves sample-to-answer in less than 90 minutes using a simple, qPCR-equivalent workflow. Of the 10 microliters loaded into each well on the consumable, over 95% of the sample is analyzed. Here we show the dPCR platform's ability to accurately conduct NIPT fetal genotyping of trisomies 13 (Patau syndrome), 18 (Edwards syndrome) and 21 (Down syndrome). A commercially available assay was used that exploits primers and probes against conserved regions of the chromosomes of interest (13, 18 and 21). In a non-trisomy DNA sample, the ratio of any of these 3 chromosome markers against the other should be 1:1. If cfDNA from both the mother and fetus are non-trisomy, all possible copy number ratios will be 1:1. If any of these ratios are elevated significantly above 1, this signals that the fetal DNA may have a trisomy and the patient would be advised to undergo more rigorous testing. This assay was run on the novel dPCR platform for 8 previously characterized maternal cfDNA samples that had a range of fetal fractions. The dPCR platform was able to accurately determine the fetal trisomy genotype in all cases. The platform presented here offers qPCR equivalent workflow, affordability and 90 minute time-to-answer while maintaining high precision and unbiased, absolute chromosome tag quantification for precision NIPT.

## **PgmNr 606: A comprehensive evaluation of the importance of follow-up prenatal diagnostic testing in the era of increased utilization of non-invasive prenatal screening.**

### **Authors:**

J. Guiltinan; T. Sahoo; M. Strecker; S.B. Hay; M.K. Travis; E. O'Toole; J. Proffitt; J. Oldzej; N. Dzidic; S. Aradhya; K. Hovanes

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical Genomics, Invitae, Irvine, CA.

---

As the use of non-invasive prenatal screening (NIPS) has become widespread, multiple studies have substantiated its superior performance in relation to traditional screening approaches, including demonstrating its significant predictive value. However, studies estimating the false positive (FP) and false negative (FN) rates and predicting pregnancy outcomes based on confirmatory results from prenatal diagnostic testing (PDT) have been limited.

**Method:** We examined 1766 consecutive singleton pregnancies referred to our laboratory with documented NIPS results, including 73 cases with one or more NIPS failures or inconclusive NIPS results. Karyotype and/or microarray analyses were performed on amniotic fluid and/or chorionic villi samples.

**Results:** Among cases for which both NIPS and PDT results were available (1693), the overall concordance between those results was 81% (1374/1693). Specifically, of the 894 cases reported as abnormal by NIPS, 63% (559/894) had concordant PDT results, representing true positives (TP). An additional 4% of these cases (33/894) showed partial concordance. The discordance rate for this subgroup of cases was therefore 33% (299/894) (FP). Conversely, NIPS was reported normal in 47% of cases (799/1693) and follow-up PDT demonstrated 98% (779/799) concordance, representing true negative (TN) results, while 2% of cases (15/799) with normal NIPS results yielded abnormal PDT results (FN). An additional 5 cases with normal NIPS results showed a sex chromosome discrepancy.

Considering specific abnormalities, the NIPS FP rate was highest for triploidy (78%; 25/32), followed by other autosomal aneuploidies (71%; 27/38), monosomy X (70%; 65/93), trisomy 13 (56%; 48/86), segmental deletions and duplications (55%; 36/66), trisomy 18 (26%; 40/151), other sex chromosome abnormalities (17%; 9/53), and trisomy 21 (13%; 49/375). The NIPS TP rate was highest for trisomy 21 (87%) and lowest for triploidy (22%). These TP rates include cases with additional findings identified by PDT and cases with partial concordance (PC); PC rates ranged from 0%-11%.

**Conclusion:** Our data demonstrating low false negative rates supports professional society recommendations for the use of NIPS as the primary screening test for detecting fetal aneuploidies. However, the considerably high false positive rates for abnormal NIPS results, other than for trisomy 21, strongly support the use of follow-up prenatal diagnostic testing, irrespective of risk estimated solely from NIPS.

## PgmNr 607: Evaluation of non-linear comorbidity structure and genetic risk for classification of preterm birth from electronic health records.

### Authors:

A. Abraham; C. Bejan; D.R. Velez-Edwards; J. Capra

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt University, Nashville, Tennessee.

---

**Background:** Preterm birth (PTB) is the leading cause of infant mortality worldwide and increase risk for immediate and long-term morbidities. Although there is a substantial heritable component to PTB, its genetic etiology remains poorly understood. Many different factors influence PTB risk. The relationships between these comorbidities, their genetic drivers, and their effects on PTB outcome remain largely unknown. Electronic health records (EHR) are a rich resource to characterize this phenotypic complexity. We hypothesize that accounting for the phenotypic complexity will reveal clinically relevant sub-phenotypes and yield insights into the genetic architecture of PTB.

**Methods:** In our delivery cohort, we define preterm from non-preterm deliveries using billing codes and estimated gestational age. First, we test ICD9 codes for association with PTB to characterize the clinical phenome. We then explored how well billing codes (other than those related to delivery status), demographic, and clinical features could classify deliveries as preterm or non-preterm using boosted decision trees. To evaluate the classification performance of genetic risk and detect any interaction benefits, we compared models with and without PTB genetic risk scores (GRS). Performance was evaluated using ROC/PR curves and top features were extracted by mean absolute SHAP value. We then used unsupervised clustering on the feature importance values to discover sub-phenotypes of deliveries and quantify their risk for PTB.

**Results & Conclusions:** Within our delivery cohort (n=11,270 PTB; 41,793 non-PTB), 298 codes were enriched for PTB and identified known reproductive and systemic comorbidities. Decision trees accurately identified PTB using billing codes (AUC=0.92) and the addition of GRS did not improve performance. To evaluate the potential for predicting PTB, we trained decision trees on features available 90 days before delivery. This algorithm accurately predicted PTB (AUC=0.82). Clustering on feature importance values yielded 12 sub-phenotypes, three of which had significantly increased risk for PTB, with distinct enrichment of comorbidities (diabetes + psychiatric traits vs. hypertensive + PTSD vs. thyroid + infectious agents). These sub-phenotypes provide interpretability and reveals non-linear combinations of comorbidities influencing risk for PTB.

## PgmNr 608: Parents' satisfaction and emotional responses to newborn genomic sequencing: The role of uncertainty.

### Authors:

M.K. Uveges<sup>1,2</sup>; L.A. Frankel<sup>3</sup>; I.A. Holm<sup>2,4</sup>; A.L. McGuire<sup>5</sup>; S. Pereira<sup>5</sup>; A.H. Beggs<sup>2,4</sup>; R.C. Green<sup>1,2,6</sup>; K.D. Christensen<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Brigham and Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) University of Houston, Houston, TX; 4) Boston Children's Hospital, Boston, MA; 5) Baylor College of Medicine, Houston, TX; 6) Broad Institute of MIT and Harvard, Cambridge, MA and Partners Personalized Medicine, Boston, MA

---

### Purpose

Genomic sequencing results can generate findings with variable actionability, and individuals who dislike uncertainty may be less enthusiastic about learning this information. We sought to understand the association between parents' tolerance of uncertainty and their emotional responses and satisfaction to receiving genomic information about their newborns.

### Methods

The BabySeq Project was a randomized clinical trial of newborn genomic sequencing (nGS) in sick and well babies. At baseline, parents completed the 12-item Intolerance to Uncertainty Scale (IUS). At three months after results disclosure sessions, parents completed the 10-item Patient Health Questionnaire (PHQ) and the 7-item General Anxiety Disorder scale (GAD), scales that assessed symptoms of depression and anxiety, respectively. At 3 months post-disclosure, parents also completed a 9-item scale assessing satisfaction with disclosed results.

### Results

337 parents completed the 3-month follow-up survey. In regression analyses that controlled for baseline scores and cohort (sick or well newborns), a one-point increase in IUS scores was associated with a 0.4 point increase in PHQ scores and a 0.3 point increase in GAD scores (both  $p < 0.001$ ). No interactions were observed between IUS scores and randomization status on either outcome (all  $p > 0.10$ ). No association was observed between IUS scores and satisfaction with results among fathers. However, among mothers, a one-point increase in IUS scores was associated with a 0.3 point decrease in satisfaction scores if newborns did not receive sequencing ( $p = 0.015$  for interaction between IUS and randomization status), but was unassociated with satisfaction scores if newborns received sequencing.

### Conclusion

Parents with more intolerance of uncertainty had more symptoms of anxiety and depression regardless of whether or not their newborn received genomic sequencing. When results were broken down by gender, mothers who reported more intolerance of uncertainty were also less satisfied with disclosure, but only when randomly assigned to the non-genomic sequencing condition. Findings highlight the importance of considering dispositional characteristics when discussing health information about newborns with parents, and suggest that newborn genomic sequencing may satisfy parents' questions about the health of their newborns rather than raise more concerns.

## PgmNr 609: Genetic burden and gestational age accurately predict complications in preterm infants.

### Authors:

K.M. Gnona<sup>1,2,4</sup>; WCL. Stewart<sup>2,5,6</sup>; P. White<sup>3,5</sup>; M. Klebanoff<sup>1,5</sup>; L.D. Nelin<sup>1,5</sup>; I.A. Buhimschi<sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Perinatal Research, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio; 2) Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio; 3) Institute of Genomics Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, Ohio; 4) Department of Biophysics, The Ohio State University, Columbus Ohio.; 5) Pediatrics, The Ohio State University, Columbus Ohio.; 6) Department of Statistics, The Ohio State University, Columbus Ohio.

---

**OBJECTIVE:** Premature infants are at increased risk for neonatal complications (NC) including: bronchopulmonary dysplasia (BPD), necrotizing enterocolitis (NEC), intraventricular hemorrhage (IVH), and retinopathy of prematurity (ROP). Moreover, many of these conditions can lead to long-term disability. However, even among preterm infants, there is still remarkable variation in the presence and severity of NC. We hypothesize that the accumulation of evolutionarily conserved variants (i.e. genetic burden) in Non-Hispanic White preterm infants influences the risk for NC.

**METHODS:** We analyzed the whole-exome sequences (WES) of 131 Non-Hispanic White preterm infants born between 28-31 completed weeks. Fifty-six infants denoted resilient (RES) did not develop severe complications (BPD  $gr \geq 3$ , IVH  $gr \geq 3$ , NEC  $gr \geq 2$ , and ROP  $gr \geq 2$ ) while 75 infants denoted susceptible (SUS) developed at least one severe complication. All preterm infants were extensively phenotyped, and cared for in the same time period, and hospital environment. We performed logistic regression across 24,254 genes to test burden for association to NC; gestational age was included as a covariate in the analysis. In a *post hoc* analysis, we also examined the predictive power of the polygenic risk score (PRS), which we computed from the genetic burden of 10 genes showing the largest evidence for association.

**RESULTS:** Our genetic burden analysis (which included gestational age as a covariate) showed that genetic burden is associated with NC at the exome-wide level ( $p = 0.05$ ). After using a 10-fold cross-validation procedure to protect against over-fitting, we also showed that PRS predicts NC (average AUC=67%,  $p = 0.01$ ). Note that, no single gene was statistically significant at the exome-wide level. Furthermore, our results confirm that gestational age is associated with NC ( $p = 7.31E-12$ ), and that prediction accuracy improves when gestational age and genetic burden are combined into a single composite predictor (average AUC=87%).

**CONCLUSIONS:** Genetic burden in evolutionarily conserved regions of the genome is contributing to the heritability of NC in Non-Hispanic Whites, with excess variation leading to increased risk. Therefore—similar to our composite predictor of NC—it should be possible to construct improved biomarkers that integrate information from the genetic burden of associated genes.

## PgmNr 610: The fate of paternal mitochondria during early embryonic development.

### Authors:

W. Zou <sup>1,2</sup>; J. Slone <sup>1</sup>; T. Huan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229, USA; 2) Reproductive Medicine Center, Department of Obstetrics and Gynecology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022, China

---

Paternal mitochondria are eliminated through different mechanisms in different animals in order to make sure that strict mitochondrial maternal inheritance is maintained. However, the elimination mechanism of paternal mitochondria in mice has been somewhat controversial, and the mechanism in humans has not yet been fully defined. Therefore, to better understand this process, we have conducted a close examination of the changes of paternal mitochondrial DNA (mtDNA) during early embryonic development. Mice and human embryos were obtained by intracytoplasmic sperm injection (ICSI), and single molecular, long-read sequencing was used to detect and analyze the percentage of sperm mitochondria in unfertilized oocyte and different development stages of embryos. In this way, we obtained the abundance of paternal mtDNA during the development of mice and human early embryos. The results provide a further clue to elucidate the molecular mechanisms of mice and human paternal mitochondria during fertilization and early embryonic development, as well as the elimination during oocyte meiosis.



## PgmNr 611: The genetic landscape of sperm mosaicism and its inter- and intra-subject variability.

### Authors:

M.W. Breuss<sup>1,2</sup>; X. Yang<sup>1,2</sup>; X. Xu<sup>1,2</sup>; D. Antaki<sup>1,2</sup>; R.D. George<sup>1,2</sup>; L.L. Ball<sup>1,2</sup>; J. McEvoy-Venneri<sup>1,2</sup>; V. Stanley<sup>1,2</sup>; J. Sebat<sup>3,4,5,6</sup>; J.G. Gleeson<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurosciences, Howard Hughes Medical Institute, University of California, San Diego, La Jolla, CA 92093, USA; 2) Rady Children's Institute for Genomic Medicine, San Diego, CA 92025, USA; 3) Beyster Center for Genomics of Psychiatric Diseases, University of California, San Diego, La Jolla, CA 92093, USA; 4) Department of Psychiatry, University of California, San Diego, La Jolla, CA 92093, USA; 5) Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA 92093, USA; 6) Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA

Somatic mutations are defined by their presence in only a fraction of the cells that comprise a tissue – a phenomenon that is referred to as mosaicism. Within the individual, in which mosaic variants arise, they have been connected to a range of diseases – including cancer or dementia. However, if these mutations occur prior to the specification of or within the germline, they have the potential to be transmitted to offspring. In a cohort of 8 families with 14 children, we previously described that 4% of 501 apparent *de novo* single nucleotide variants (dSNVs) on the paternal haplotype were in fact mosaic in father's sperm. Moreover, the majority of these variants was not detectable in blood or only at significantly lower levels. These data and additional analysis of untransmitted SNVs suggested that sperm and by extension the spermatogonial stem cells harbor a rich collection of somatic mutations that allow insights into the early embryonic development and the establishment of the germline; moreover, understanding the extent and patterns of sperm mosaicism allows for the refinement of clinical risk assessment of recurrence and *de novo* transmission.

Here, we extend our analysis to a cohort of 12 individuals aged 18-22 years (young), and 4 between the ages of 48-62 years (old). We performed at least 300x whole genome sequencing on the sperm and blood of each individual to reliably detect mosaicism down to 2% with a low false-positive rate (<10%). In addition, for the majority of the young cohort we collected additional time points for a total of 16. This comprehensive data set allows us to determine the inter- and intra-subject variability of sperm mosaicism, within and across age groups. We find that each individual harbors around 20 sperm-mosaic variants with around two-third being only detectable in sperm. Across our data sets, allelic fractions of mosaic variants suggest that a restricted founder population of 3-6 cells dominates the germline. Furthermore, the mutational signatures reveal mutagenic processes that are distinct between early embryonic development and post germline establishment and differ from those mutations that are accumulated in the germline with age. In addition, we predict that our framework has the ability to detect clinically relevant mosaicism that conveys significant risk to inherit a potential or observed pathogenic mutation.



## PgmNr 612: Pedigree analysis of rare variants identified in monozygotic concordant transgender twin males through whole genome sequencing.

### Authors:

J.G. Theisen; L. Chorich; H.G. Kim; L. Layman

[View Session](#) [Add to Schedule](#)

**Affiliation:** The Medical College of Georgia at Augusta University, Augusta, GA.

---

**Introduction:** Transgender identity (TGI) occurs in 0.3% of the population and is associated with both significant societal discrimination and even frequent discrimination within the healthcare system. One root cause of this discrimination is a lack of understanding regarding its etiology. While prior studies have suggested a genetic contribution, no candidate genes have yet been verified through pedigree or functional analysis. We hypothesized that rare variants in genes involved in sexually dimorphic brain development contribute to TGI, and that these variants would follow known Mendelian patterns of inheritance upon segregation analysis.

**Methods:** Whole genome sequencing was performed on the genomic DNA of a family with monozygotic concordant transgender male (natal female) twins as well as their cisgender parents and sister. Variants were considered if they met the following criteria: 1) not present in 88 cisgender controls, 2) frequency  $\leq 0.01$  in the ExAC, Yale, and 1000 Genomes databases, 3) predicted to have a functional effect compared to the wild-type gene. Segregation analysis was then performed on the remaining variants, and those consistent with either paternal-dominant, recessive, or *de novo* inheritance pattern of TGI were considered for further evaluation.

**Results:** Whole genome sequencing, with an average coverage of 99.18% and average read depth of 33 demonstrated 50,715 genetic variants. After filtering and pedigree analysis, there were 119 remaining variants showing patterns consistent with either paternal-dominant ( $n=100$ ), recessive ( $n=13$ ), or *de novo* ( $n=6$ ) inheritance. A compelling heterozygous variant in *MAP4K3* was noted in both twins as well as their father, while their mother and sister were both identified as wild type. Our prior research has identified *MAP4K3* as a candidate gene that may play a role in the development of gender identity in humans, making it interesting for further investigation.

**Conclusion:** In this family, whole genome sequencing identified 119 rare variants with inheritance patterns that would be consistent with either paternal-dominant, recessive or *de novo* inheritance of TGI. While we believe that TGI likely has multifactorial genetic and environmental contributors, these findings add to the growing body of knowledge regarding the understanding of its genetic etiology. This is especially true in the case of *MAP4K3*, which we previously identified as a candidate gene, potentially affecting gender identity in the brain.

## PgmNr 613: Accurate fetal sex determination from maternal blood at 8 weeks gestation.

### Authors:

J. Casanova; S. Cacia; H. Milot; C. Jacob

[View Session](#) [Add to Schedule](#)

**Affiliation:** Gateway Genomics, 7590 Fay Ave. La Jolla, CA. 92037

---

**Background:** Prior studies have shown that SneakPeek® Early Gender Test can accurately determine fetal sex as early as 9 weeks gestation using a qPCR-based assay. The purpose of this study was to evaluate the clinical performance of SneakPeek® for noninvasive prenatal testing (NIPT) to determine fetal sex at an earlier time point: 8 weeks gestation.

**Methods:** A multicenter blinded study was conducted at five ultrasound clinics. Maternal blood samples were collected from 108 pregnant women between 7.00 and 10.57 weeks of gestation. Plasma was separated from whole blood by centrifugation. A commercial DNA extraction kit (MagMAX® Cell-Free DNA Isolation Kit) was used to isolate circulating cell-free DNA from maternal plasma. Real-time quantitative PCR was used to detect fetal DNA using a multi-copy sequence on the Y-chromosome. Total cell-free DNA (maternal and fetal circulating free DNA) was measured using an autosomal control gene.

**Results:** Male Y-chromosome DNA was detected in all samples from women carrying a male fetus. SneakPeek® correctly identified fetal sex in 107 of 108 pregnancies. Four of the 108 samples initially yielded an inconclusive result. A second sample was collected from the four participants and a result was obtained on the second round of testing. Fetal sex for all samples were unknown prior to genetic testing and confirmed through a sonogram. In this study, SneakPeek® accuracy, sensitivity, and specificity were 99.1%, 100%, and 98.2% for fetal sex identification, respectively.

**Conclusion:** This study showed that SneakPeek® Early Gender Test is highly accurate for fetal sex determination in pregnancy as early as 8 weeks gestation.

**Keywords:** SneakPeek, NIPT, Early Gender Test, Pregnancy, Fetal Sex, Maternal Blood, Maternal Plasma, Cell-free Fetal DNA

# PgmNr 614: Fast and accurate SNP discovery and allele frequency estimation from large-scale low-pass NIPT genome sequencing.

## Authors:

S. Huang <sup>1</sup>; Z. Li <sup>2</sup>; R. Liu <sup>1</sup>; X. Jin <sup>1,3</sup>; S. Liu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) BGI-shenzhen, Shenzhen, Guangdong, China; 2) BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, Guangdong, China; 3) School of Medicine, South China University of Technology, Guangzhou 51006, Guangdong, China

---

## Background

Non-invasive prenatal test (NIPT) for fetal trisomy by sequencing of maternal plasma cell-free DNA to 0.06 – 0.3 fold has become the fastest adopted molecular test in history. Nowadays, over ten millions of NIPT tests have been carried out globally, representing a data set with a sample size capacity similar to world-wide array data. The NIPT sequencing data, when combined with clinical records from the hospitals, are valuable for genome-wide association studies of many important maternal and child's traits. However, SNP discovery from the NIPT sequencing remains difficult since traditional multi-sample variation calling algorithms are incapable of fast and accurate variation calling from large-scale low-pass genomes although it is the first step for downstream genetic analysis.

## Findings

In our study, we present a novel and highly efficient approach that integrates maximum likelihood estimation and multiple likelihood ratio tests for SNP discovery and allele frequency estimation for massive amount of NIPT genomes. The algorithm can not only detect and estimate allele frequency for bi-allelic SNPs but also multi-allelic SNPs. Analysis on simulation data suggests a close-to-zero false positive rate for SNP discovery and negligible difference on allele frequency estimation. Comparison on the speed and accuracy for variation discovery and allele frequency estimation on hundreds of thousands and up to million of individuals between our method and the traditional bayesian approaches implemented in GATK and Samtools suggests that our method is more than 20 times faster while maintains similar accuracy given the same computational resource. Genome-wide association analysis using variations called from our approach indicates greater power than directly using the known polymorphic SNPs from the reference panel. The method so far has been applied on the study of more millions of NIPT individuals and can be downloaded from <https://github.com/ShujiaHuang/BaseVar>.

## Conclusions

We present an important algorithm that can be used for accurate SNP discovery and allele frequency estimation from up to millions of NIPT sequencing data. Evaluation on simulation and massive amount of real data prove its largely improved efficiency and accuracy compared to traditional variation calling approaches. Our approach supports the great potentiality of the continuously accumulating and improving NIPT data for genome-wide association mapping of important complex traits.

## PgmNr 615: Fetal fraction cutoff established for NIPS accuracy may not predict higher risk of trisomy.

### Authors:

J. Lopes <sup>1</sup>; G. Lopes <sup>2</sup>; E. Enninga <sup>3</sup>; R. Rowsey <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota.; 2) Department of Biomedical Statistics and Bioinformatics, Mayo Clinic, Rochester, Minnesota.; 3) Department of Obstetrics and Gynecology, Mayo Clinic, Rochester, Minnesota.

---

**Introduction:** The ability to isolate cell free DNA (cfDNA) from the maternal bloodstream is a powerful tool for noninvasive prenatal screening (NIPS) of trisomies 13, 18, and 21. Accurate prenatal assessment depends largely on the fraction of fetal cfDNA present in the maternal serum sample. Recently published findings revealed that the presence of trisomies is associated with decreased fetal fraction, prompting guidelines for women with failed NIPS testing due to inadequate fetal cfDNA to undergo invasive testing. In the current study, we aimed to determine whether decreased fetal fraction was associated with trisomy in our cohort, and if NIPS failure due to a fetal fraction below the established clinical cut off of 4% is an indicator of risk.

**Methods:** Information including maternal BMI, maternal age, fetal sex, gestational age, fetal cfDNA fraction, and NIPS results was collected on a cohort of 2,374 pregnancies referred to the Mayo Clinic Genomics Laboratory between January 2016 and September 2018. Additional clinical information was available for 1,181 patients who delivered at Mayo Clinic, including fetal outcome, fetal and maternal complications, and maternal smoking status. We determined the success rate of retesting after initial failure due to low fetal fraction and conducted linear and logistic regressions to investigate associations between fetal fraction and the aforementioned variables.

**Results:** Consistent with previous reports, fetal fraction was positively associated with gestational age ( $p < 0.001$ ) and negatively associated with BMI ( $p < 0.001$ ). Fetal fraction was negatively associated with the presence of trisomy ( $p = 0.031$ ) and intrauterine fetal demise (IUFD;  $p = 0.046$ ), controlling for BMI and gestational age. However, the proportions of trisomies or IUFD were similar in women who had fetal fraction below and above 4%. Fetal fraction was not associated with maternal age, fetal sex, pre-eclampsia, gestational hypertension, or maternal smoking status. 66% of redraws were successful.

**Conclusion:** Although decreased fetal fraction is associated with fetal trisomy and IUFD, failure to meet the clinical cutoff of 4% fetal fraction established for NIPS accuracy does not suggest increased risk for these characteristics. Because most redraws were successful and the majority of failed NIPS were explained by high BMI and low gestational age, a redraw may be a more appropriate next step before invasive screening due to concerns of trisomy risk.

## PgmNr 616: Single circulating trophoblast testing provides a new window for detecting placental mosaicism.

### Authors:

A. Breman<sup>1</sup>; L. Vossaert<sup>2</sup>; Q. Wang<sup>2</sup>; R. Salman<sup>2</sup>; W. Mahony<sup>2</sup>; A. McCombs<sup>2</sup>; V. Patel<sup>3</sup>; C. Qu<sup>3</sup>; P. Liu<sup>2,3</sup>; C. Shaw<sup>2,3</sup>; B. Levy<sup>4</sup>; R. Wapner<sup>5</sup>; W. Bi<sup>2,3</sup>; I. Van den Veyver<sup>2,6</sup>; A. Beaudet<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical & Molecular Genetics, IU School of Medicine, Indianapolis, IN, USA; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 3) Baylor Genetics, Houston, TX, USA; 4) Pathology and Cell Biology, Columbia University Medical Center, New York City, NY, USA; 5) Obstetrics and Gynecology, Columbia University Medical Center, New York City, NY, USA; 6) Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX, USA

---

Confined placental mosaicism (CPM) and true fetal mosaicism are known to occur in human pregnancy. Mosaicism, mostly CPM, is detected with cell-free DNA-based noninvasive prenatal testing (NIPT), but there is limited information with cell-based NIPT. We have been developing Single Circulating Trophoblast (SCT) testing as a novel form of noninvasive prenatal diagnosis. Trophoblasts obtained after immunomagnetic enrichment from 30-40 mL of maternal blood are analyzed for copy number variants by single cell whole genome next-generation sequencing. Since June 2017, we have analyzed 247 samples for which at least one isolated trophoblastic cell was scorable for fetal aneuploidies, with a median of 3 scorable cells/case (sequenced up to 5 cells/singleton case). Evidence for CPM was found in 5/247 cases (2.0%).

CPM in CVS samples can be divided into three categories: abnormal cells confined to the cytotrophoblast layer (CPM I, direct preparation), confined to the mesenchymal core (CPM II, long-term culture) or found in both lineages (CPM III); only CPM I & III would be detected by our SCT testing assay. The frequency observed is 2-fold higher than the reported combined frequency for CPM I & III (0.78% and 0.22%, respectively, in CVS samples from first-trimester pregnancies). The chromosome imbalances in the analyzed SCT cells from each case (shown as abnormal cells/total cells analyzed) were monosomy X (3/7), trisomy 13 (2/2), trisomy 15 (1/1), trisomy 4 (3/3) and gain and loss of 13q (4/4). In four of the five SCT CPM cases, CVS and/or amniocentesis testing revealed normal cytogenetic results, consistent with CPM I. For one case, CVS analysis indicated very low-level mosaicism, but follow-up amniocentesis showed a normal profile.

These results suggest that CPM I may be a more common phenomenon than currently recognized, which is perhaps not surprising given the high-frequency mosaicism seen in day 3 embryos and day 5 blastocysts. It is possible that the detection of CPM I by direct CVS biopsy is an underestimate given the tiny fraction of the placenta sampled. It is also of note that uniparental disomy (UPD) 15, causing Prader-Willi or Angelman syndrome, is assumed to occur through trisomy and monosomy rescue, respectively; yet this hypothesized mosaicism has rarely been seen in CVS samples leading to UPD15. These data have important implications for normal human biology and for disease states that might be caused by mosaicism that has been undetected to date.

## PgmNr 617: A pilot study of exome sequencing of cell free fetal DNA for detection of pathogenic variants.

### Authors:

D.L. Filer <sup>1,2</sup>; K.L. Gilmore <sup>3</sup>; B.C. Powell <sup>1,2</sup>; P.A. Mieczkowski <sup>2</sup>; K.C. Wilhelmsen <sup>1,2,4</sup>; N.L. Vora <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 2) Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 3) Department of Obstetrics and Gynecology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 4) Department of Neurology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

---

Prenatal genetic diagnosis provides earlier results and expanded testing options over traditional newborn screening, promoting more effective treatment in the immediate newborn period. Expanding the scope of prenatal genetic testing will enable additional treatment options, especially with the ongoing development of gene therapies. The most comprehensive genetic testing requires obtaining a direct fetal sample, either through amniocentesis or chorionic villus sampling, with risk for pregnancy loss and other complications. While the overall risk of these procedures is low, we aim to eliminate invasive procedures by interrogating fetal-derived cell-free DNA fragments in maternal blood. As a pilot study, we performed whole-exome sequencing (WES) on cell-free DNA in three pregnant women -- two with structurally anomalous fetuses, and one with previous pregnancies affected by Menkes syndrome. We focused our analyses on high-quality single-nucleotide polymorphisms (SNPs), cross-referencing SNPs identified in maternal cell-free DNA with likely- and known-pathogenic SNPs listed in the NIH ClinVar database. In case one, we identified a recessive causal variant c.1120 G>T (p. Glu374Ter) for type VIII osteogenesis imperfecta in the *P3H1* gene. The variant fit with the sonographic findings of shortened long bones, bowed femurs, and hypoechogenic bones. The homozygous variant in *P3H1* was confirmed by amniocentesis and Sanger sequencing. The minor allele frequency for the *P3H1* variant was 53%, which, despite sequencing coverage >200x, would have been insufficient to suggest fetal homozygosity without the sonographic findings. Case two occurred in a mother with five prior pregnancies affected with X-linked recessive Menkes syndrome; we did not find any SNPs within the *ATP7A* gene. Case two did reveal the mother, but not the fetus, carries the c.1187G>A (p.Gly396Asp) variant in the *MUTYH* gene for MYH-associated polyposis. In case three, the fetus had arthrogryposis with clubbed feet and elbow contractures on sonography. We did not find any likely- or known-pathogenic SNPs via cell free DNA analysis or via whole exome sequencing from the direct fetal sample. Taken together, these results demonstrate some ability to identify causal variants from WES on cell-free DNA, but illustrate the need for greater efforts.

## PgmNr 618: Single circulating trophoblast testing for noninvasive prenatal diagnosis.

### Authors:

L. Vossaert<sup>1</sup>; Q. Wang<sup>1</sup>; R. Salman<sup>1</sup>; W. Mahony<sup>1</sup>; A. McCombs<sup>1</sup>; B. Lang<sup>2</sup>; V. Patel<sup>3</sup>; C. Qu<sup>3</sup>; P. Liu<sup>1,3</sup>; C. Shaw<sup>1,3</sup>; B. Levy<sup>4</sup>; R. Wapner<sup>5</sup>; W. Bi<sup>1,3</sup>; A. Breman<sup>6</sup>; I. Van den Veyver<sup>1,2</sup>; A. Beaudet<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, TX, USA; 2) Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX, USA; 3) Baylor Genetics, Houston, TX, USA; 4) Department of Pathology and Cell Biology, Columbia University Medical Center, New York City, NY, USA; 5) Department of Obstetrics and Gynecology, Columbia University Medical Center, New York City, NY, USA; 6) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA

Cell-based noninvasive prenatal testing (NIPT) as a novel form of prenatal diagnosis has been explored for decades. We have developed a Single Circulating Trophoblast (SCT) testing method for the analysis of trophoblasts from maternal blood during the late 1st/early 2nd trimester to reliably detect  $\geq 1$  Mb fetal copy number variants (CNV) (PMID 27616633, 27761919). We here report on 1) the results of two pilot studies in preparation of a more rigorous CAP/CLIA validation study, 2) data from additional samples collected at early gestational age (GA), and 3) our experience with multiple pregnancies.

1) A total of 95 samples (30-40 ml) were collected at 8w2d-20w6d GA in the two pilot studies. After enrichment, putative trophoblasts were picked to undergo CNV analysis by low coverage single-cell whole genome next generation sequencing (NGS). Overall, an average of 0.20 cells/ml was identified. About 55% of all sequenced trophoblasts were of high quality (scorable for both aneuploidy and CNV), and with up to 5 sequenced cells/case, there were on average 1.93 high-quality cells/sample. This corresponded to 53% of cases having  $\geq 2$  high-quality cells.

2) An additional 44 samples were collected specifically at 6w4d-10w6d GA. An average of 0.12 cells/ml were identified by microscopy. For 8 samples no trophoblasts were available for analysis, but NGS data are available for 29 cases and still pending for 7 cases. Of all sequenced cells, 47% were of high quality, and for 12/29 cases  $\geq 2$  high-quality cells were available.

3) When considering multiples, 26 twins and 1 triplet case were recruited since June 2017 and yielded higher cell numbers compared to singletons. An average of 0.42 trophoblasts/ml was identified, yielding on average 3.6 high-quality cells per case with NGS analysis and 75% of cases having  $\geq 2$  high-quality cells. For the confirmed opposite-sex twins, cells from both fetuses were recovered in 6/8 cases, and also for the triplet (2 male and 1 female), both normal male and female cells were recovered.

In conclusion, SCT testing is a rapidly evolving technique with strong potential as a clinical test in the near future. We show that SCT testing is also feasible at early GA, although the yield is less than at 10-15w GA. Increasing the trophoblast yield would allow earlier prenatal testing and possibly earlier pregnancy management than conventional testing methods. Additionally, the ability to recover cells from separate fetuses is an advantage over cell-free NIPT.



## PgmNr 619: Sex differences in gene expression found in term uncomplicated human placentas.

### Authors:

K. Olney<sup>1,2</sup>; T. Phung<sup>1,2</sup>; M.A. Wilson<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Life Sciences, Arizona State University , Tempe, AZ; 2) Center for Evolution and Medicine, Arizona State University , Tempe, AZ

---

Pregnancy complications vary based on the genetic sex of the fetus. Pregnancies with a male fetus are more likely to be delivered by cesarean section and delivered preterm. Subchorionic hemorrhage is the most common pregnancy complication occurring in ~25% of pregnancies and is 6.8 times more likely with a male fetus than pregnancies with a female fetus. Pregnancies with a female fetus, on the other hand, are more likely to experience intrauterine growth restriction in which the fetus does not grow to normal healthy weight. The placenta, which is the genotype of the fetus, is critical for sustaining and maintaining a healthy pregnancy by acting as an immune modulator to protect the developing fetus from the mother's immune system. Improper placenta function can result in pregnancy complications and may be driven by changes in gene expression. Indeed, several studies have shown that placenta gene expression differs between healthy term placentas and placentas from pregnancies characterized by preeclampsia and intrauterine growth restriction. While these studies enhanced our understanding of these phenomena there was limited sample size, and sex differences were not analyzed. Here we fill this gap by characterizing genome-wide expression differences and similarities between 30 male XY and 30 female XX placentas from term uncomplicated pregnancies to create an overview of typical sex differences in one of the earliest developed tissue, the placenta. Our preliminary results show that genes that are more highly expressed in the female samples are involved in the stimulation of the ovaries to synthesize steroids for sustaining the pregnancy whereas higher expressing genes in male placental samples are involved in T cell differentiation in immune response. Further characterizing sex differences in these full-term placentas will provide the groundwork for future studies to investigate sex-specific disease-related gene expression.



## PgmNr 620: X chromosome inactivation is heterogeneous in the human placenta.

**Authors:**

T. Phung<sup>1,2</sup>; K. Olney<sup>1,2</sup>; M.A. Wilson<sup>1,2</sup>

[View Session](#)[Add to Schedule](#)**Affiliations:**

1) School of Life Sciences, Arizona State University, Tempe, AZ; 2) Center for Evolution and Medicine, The Biodesign Institute, Arizona State University, Tempe, AZ

---

In mammals, genetic females (XX) have two X chromosomes while genetic males (XY) have one X chromosome. To equalize dosage between males and females, one of the X chromosomes in XX females is silenced through a process called X chromosome inactivation (XCI).

Previous studies of patterns of XCI in the placenta were either in non-human placentas, or only examined a few SNPs and genes in humans. In addition to the patchiness of X-inactivation across regions of the placenta, XCI is not complete in humans; as many as one-third of the genes on the X chromosome escape XCI. It is not known whether the same genes escape XCI in the same tissue across multiple individuals, and gene-specific escape in the placenta has not been investigated.

To better understand the patchiness, or potential maternal or paternal bias in XCI in the placenta, and whether genes escape XCI vary between individuals, we sequence 30 placenta tissues from 30 full-term, healthy pregnancies with female offspring. This is important because the placenta is the genotype of the fetus. Importantly, for each placenta, we collect extract RNA from two different sites. We choose the placenta in this study because heterogeneity in patterns of XCI can give us insights into when during development the X chromosome is silenced, and because understanding heterogeneity in this early-formed tissue is important for its downstream developmental effects. We perform whole-exome DNA sequence of each individual and RNA-seq for each extraction site. We quantify XCI for each extraction site from the same placenta to understand how heterogeneous XCI is in the placenta. For each gene on the X chromosome, we determine if it escapes XCI across all 30 placentas or the pattern of escape genes is heterogeneous. We discuss our findings on the heterogeneous patterns of XCI within and between individuals and their clinical implications.

## PgmNr 621: A *BRIP1* likely pathogenic variant and a *CDH1* variant of uncertain significance in a patient with lobular breast cancer.

### Authors:

T. DeMarco <sup>1</sup>; A. Schott <sup>1</sup>; K. Matthijssen <sup>1</sup>; D. Alaeddin <sup>1</sup>; M. Turner <sup>1</sup>; N. Hauser <sup>2</sup>; J. Deeken <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

- 1) Cancer Genetics Department, Inova Schar Cancer Institute, Inova Fairfax Hospital, Fairfax, Virginia.;
- 2) Inova Translational Medicine Institute, Inova Fairfax Hospital, Fairfax, Virginia

---

We present the case of a 48-year old female with lobular breast cancer and a family history of breast and gastric cancer. In 2016, the patient underwent genetic testing utilizing a multi-gene panel. This testing revealed a likely pathogenic variant (LPV) in the *BRIP1* gene, *c.206-2A>G* and a variant of uncertain significance (VUS) in the *CDH1* gene, *p.T285I*. The patient's father passed away from gastric cancer with signet ring cell features (pathology confirmed). Following the patient's testing, a paternal aunt developed breast cancer and was also found to carry the *CDH1* VUS, confirming paternal inheritance. According to the International Gastric Cancer Linkage Consortium Guidelines, identification of a *CDH1* VUS in the setting of a family history consistent with hereditary diffuse gastric cancer (HDGC) should prompt consideration of endoscopy with multiple random biopsies, although this has not been proven to reliably detect this cancer at an early stage. *CDH1* mutations are also associated with about a 56% lifetime risk for breast cancer, often lobular. *BRIP1* mutations are associated with about a 6% lifetime risk for ovarian cancer and possibly a modestly increased risk for breast cancer. The patient recently opted for a total hysterectomy. A sister with LCIS was also found to carry the *BRIP1* LPV, but not the *CDH1* VUS, which was a finding of particular interest, given the association between lobular breast pathology and *CDH1* pathogenic mutations. A second sister was also found to carry the *BRIP1* LPV and *CDH1* VUS and developed breast cancer a year after testing positive. The identification of the *CDH1* VUS and confirmation of paternal inheritance has proven to be more distressing for this family than the identification of the *BRIP1* LPV, a result with an established NCCN recommendation for prophylactic oophorectomy. The uncertainty of screening for HDGC and the possibility that this variant may be reclassified as pathogenic has significantly raised the cancer related anxiety in this family. However, the patient and her relatives have been counseled that only if the *CDH1* VUS is upgraded, would this prompt a consideration of prophylactic gastrectomy. This case illustrates the complexity of multi-gene panel results, the importance of dynamic family history as a contributor to result interpretation, and the intricacy of genetic counseling for variants of uncertain significance, in particular in families consistent with a high-risk cancer susceptibility syndrome.

## **PgmNr 622: Towards equity in precision medicine: Population differences in the use of genetic testing for cancer in a regional hospital system.**

### **Authors:**

L. Landry<sup>1,2,3,4</sup>, N. Lindeman<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Harvard Medical School, Boston, MA; 2) Dana Farber Cancer Center, Boston, MA; 3) Brigham and Women's Hospital, Boston, MA; 4) Harvard School of Public Health, Boston, MA

---

Precision Medicine centers around the use of biomarkers as diagnostic, prognostic, treatment predictive, and risk-based tests. These genetic tests are the *defacto* gateway to precision medicine. However, the lack of diversity in genetic research, the demonstrated differences in clinical utility, and limited access to clinical genetic tests in some disease areas suggests that precision medicine may exacerbate existing disparities. There has been a major push to increase the diversity in genetic research through programs like the *All of Us Program* and *Million Veteran's Program*. There is growing concern that even if the differences in utility are resolved, the genomic tests necessary for precision health care will not be available to all. Complicating the issue is the lack of data on use of genetic testing across diverse racial groups. Using a hospital system-wide centralized clinical data registry and warehouse with more than 1 million patient notes from more than seven hospitals, we examined the demographic characteristics of select cancer focused genetic tests from (1988-2019). Specifically, we conducted a cross-sectional study of genetic testing for select hereditary cancers, select heme malignancies, as well as testing for select genes involved in targeted cancer therapy in the Partners' Research Patient Data Registry. Our study shows large significant population differences in cancer testing for pharmacogenetics [84.9% White, 3.8% Asian, 3.5% Black, and 2.0% Hispanic]; heme malignancies [82.5% White, 2.4% Asian, 4.9% Black and 2.1% Hispanic], as well as hereditary cancer [88.6% White, 2.7% Asian, 2.5% Black and 1.7 % Hispanic]. In the more than 20,000 patient notes with genetic testing more than 80% were for White patients. Most of these patients received testing at large hospitals (> 750 beds), with fewer patients receiving genetic testing at smaller community hospitals (< 200 beds). Although these findings are limited to a single hospital system and are not nationally representative of all genetic testing, these data show the problem may be pervasive in all types of genetic testing for cancer. A national study on use of genetic testing for cancer diagnosis, prognosis, treatment, and disease management is needed. Furthermore, investigation into the potential cause of these differences in use, as well as solutions including public advocacy, patient engagement and provider education is required.

## **PgmNr 623: Disclosure of familial implications of pathogenic *BRCA1* to probands: Opportunity for prompting family communication.**

### **Authors:**

S. Makhnoon; B. Arun; R.J. Volk; S.K. Peterson

[View Session](#) [Add to Schedule](#)

**Affiliation:** MD Anderson Cancer Center, Houston, TEXAS.

---

Familial communication of pathogenic *BRCA1* results is needed to maximize the clinical utility of sequencing and its public health benefits. Cascade *BRCA1* genetic testing is suboptimal and practical strategies for promoting familial testing are needed. Disclosure of *BRCA1* results present opportunities to counsel probands about familial communication. Little is known about what familial communication recommendations are offered to probands by genetic counselors (GCs). We examined differences in how GCs discussed familial implications of pathogenic *BRCA1* results based on: specificity of familial risk-messaging; offer of logistical help for relatives; and, provision of family letters. We completed structured content analysis of 159 *BRCA1* result disclosure notes, abstracted from electronic medical records (EMRs), for probands seen at MD Anderson Cancer Center from 2014-2019. Probands were 94% female, median age of 48.6 years, 52.3% non-White, and were counseled by 20 unique GCs with mean post-graduate experience of 4.2 years. 53.3% of notes described specific at-risk relatives who would benefit from genetic counseling; the remainder described familial implications more generically. The practice of including specific information differed widely among individual GCs [range: 6-88%]. 27% provided information on logistical genetic counseling assistance for relatives (e.g., targeted testing cost, assistance locating GCs or testing services for underinsured). More experienced GCs were more likely to provide detailed familial implication information ( $p=0.001$ ) and offer logistical help ( $p=0.04$ ). Presence of an accompanying relative to an in-person appointment ( $n=34$ ) did not change the specificity of familial implications offered, compared to unaccompanied ( $n=48$ ) or telephone disclosures ( $n=73$ ) ( $p=0.91$ ). The provision of family letters was mentioned in 48.4% of notes, but the intended recipients or purpose were frequently not specified. Some notes indicated familial communication primarily as a means to determine paternal or maternal inheritance of mutation. Although there may be discordance between EMR disclosure notes and actual counseling, findings suggest that offering specific risk-messaging, logistical help and explaining intended recipients of family letters are opportunities for improving familial communication. Counseling to determine inheritance patterns of *BRCA1* can be framed as opportunities to promote familial communication and cascade genetic testing.

## PgmNr 624: The Gene Machine: Exploring the road less traveled.

**Authors:**

N. Merner; E. Stallworth; M. Bishop; A. Huskey; S. Omeler; I. McNeely

[View Session](#) [Add to Schedule](#)

**Affiliation:** Auburn Univ, Auburn, Alabama.

---

Robert Frost inspired the world by taking the road less traveled. Interpret as you may, he claimed it made all the difference. New perspectives are gained from breaking away from the norm, yet today, as scientists, we find ourselves continuously following the path of least resistance regarding the recruitment of human subjects into research. We are all aware of health disparities in the U.S. and the recruitment bias towards urban-living Caucasians. Individuals living in rural America, as well as those of ethnic minorities, are the most susceptible to health disparities but the most underrepresented in research. As geneticists, particularly in the era of precision medicine, we hold the power to influence change. We need to step outside our comfort zones and make a concerted effort to include vulnerable populations in our research.

The back roads of Alabama are secluded, winding through hardwood and pine forests and open arable land. One can also see pockets of housing, some beautifully kept and others that truly do not seem habitable. The Gene Machine — a mobile recruitment and enrollment bus — witnesses this first-hand as its travels through the state to secure study participants for hereditary breast cancer genetics research. This eye-catching pink bus is a part of an IRB-approved community-based recruitment (CBR) protocol centered on overcoming barriers to research participation. We previously published our rudimentary strategies to recruit in the rural and medically underserved state of Alabama, which also has nearly double the national percentage of Black/African American individuals. Including underrepresented individuals in research is an arduous process, but it is also gratifying, and we are now implementing modifications to enhance enrollment based on lessons learned.

Overall, the Gene Machine brings research opportunities to the underrepresented, along with the option to receive genetic research reports. To date, 99% of the CBR study probands have chosen to be informed of clinically relevant findings; only 41% previously had clinical genetic testing. Notably, we established a collaborative telegenetic counseling project to provide local services to mutation carriers; interestingly, upon receiving a report, to date, only Caucasians have accepted the service, indicating adjustments are necessary. Overall, we all can learn from exploring the road less travelled, and we hope the Gene Machine can inspire a movement towards true inclusion.

## PgmNr 625: A modern approach to cascade genetic screening: Use of a chatbot to enable family sharing.

### Authors:

S. Nazareth<sup>1</sup>; E. Simmons<sup>1</sup>; M. Snir<sup>1</sup>; M. Shohat<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clear Genetics, San Francisco, CA; 2) Tel Aviv University, Tel Aviv, Israel

---

**Introduction:** Cascade screening is an ideal way to target healthy at-risk individuals, but suboptimal family communication, reliance on paper documents, and downstream geographic barriers all make the execution challenging. There is a need to develop novel approaches to family sharing that leverage modern societal communication modalities.

**Methods:** In collaboration with board-certified genetic counselors, Clear Genetics, Inc. developed a Family Sharing Tool that patients can use to share genetic test results with at-risk relatives using a “cascade chatbot” via email, text, or Facebook private message. We assessed the ability of a BRCA1 Cascade Chatbot to communicate genetic information to hypothetical at-risk relatives. The chatbot—named GIA for Genetic Information Assistant— was programmed to describe the familial mutation, relay associated disease risk, and facilitate next steps, including triage to telephone genetic counseling. The tool was tested on 100 users, ages 45-55 with at least a high school degree, to assess ease of use, satisfaction and comprehension.

**Results:** When asked about ease of use, 96% said the tool was “easy/very easy” to use, and 92% felt it was an effective way to get information. Only 3% would have preferred a call with a genetic counselor, and 4% would prefer a letter in the mail. When given the option to describe GIA’s personality, the top 3 descriptors were: “Informative” (84%), “Friendly” (63%), and “Professional” (52% ). To understand satisfaction levels, users were asked to describe how the interaction with GIA made them feel: 91% of reported “Informed”, 58% felt “Engaged”, and 34% “Empowered.” Other feelings were “Grateful” (23%), “Scared” (9%), and “Uncomfortable” (7%). Finally, 99% understood that testing was necessary to determine personal risk, and 76% understood that a positive result would stratify them as high-risk.

**Conclusion:** Preliminary data suggests that offloading the burden of family sharing to a chatbot can facilitate effective communication to at-risk relatives. This mode of communication alleviates many of the logistical barriers of cascade screening, while ensuring the delivery of authoritative genetic information to targeted relatives. Testing of this tool in a clinical setting is the next step to determining whether it can optimize cascade screening and change patient behaviors.

## PgmNr 626: Telegenetics: A new method for genetic counseling.

**Authors:**

A. Radfar <sup>1</sup>; I. Filip <sup>2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Encino, California.; 2) Kaiser , California

---

**Objective:** Telegenetics allows physicians to provide medical genetic care and counseling services remotely through audiovisual technology. Telegenetics may address many challenges facing our society in Middle -Eastern population living in U.S. Due to high level of genetic diseases in this population, telegenetics can increase detection , diagnosis and treatment of disorders in the young and aging population and decrease healthcare cost. With this work, we attempt to evaluate how telegenetics can effect a change in these challenges, and evaluate what obstacles prevent some providers from using it.

**Methods:** In this work, the cost-effectiveness, success of telegenetic care, usefulness in reaching developing and underdeveloped areas, dif?culties preventing the use of telegenetics, and proposals to overcome these challenges were reviewed and analyzed.

**Results:** Cost of telegenetics was reported 19% less expensive than traditional face-to-face care. In several studies, telegenetics was documented to have had equal or better outcomes than face-to-face encounter. Dif?culties in using telegenetics include affordability of equipment, lack of technical support in developing or underdeveloped areas, legality of licensure and patient privacy and satisfaction.

**Conclusions:** Although cost savings and convenience are major advantages of this technology, concerns with delivery barriers and challenges require cautious embracement of telemedicine. A great deal of research is needed to show that telegenetics improves patient centered outcomes.

## PgmNr 627: In the age of Twitter, email still works to engage clinicians in genomics education.

### Authors:

K. Reed <sup>1</sup>; P. Tanner <sup>1</sup>; L. Steinmark <sup>1</sup>; E. Edelman <sup>1</sup>; J. Lillpopp <sup>2</sup>; T. Ingram <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical and Continuing Education Program, The Jackson Laboratory, Bar Harbor, Maine.; 2) Genomics Office, Connecticut Department of Public Health

---

For genomic medicine to have an impact on public health, clinicians without formal training in genomics need to incorporate genetics/genomics into their practice. While dissemination of genomic education is recognized as an important aspect of reducing practice gaps, it has received limited study. The Jackson Laboratory Clinical Education Program (JAX) with support from the Connecticut Department of Public Health Genomics Office (DPH) developed and implemented a marketing strategy for *Cancer Genetic Clinical Education*, a well-tested set of 11 online, interactive modules on cancer genetics. This program seeks to train primary care providers (PCPs) to use evidence-based guidelines about Tier I cancer genetic applications. The goal was to enroll 100 new CT clinicians in the program as part of the larger DPH effort to help clinicians identify and manage patients at risk for hereditary cancer. **Methods:** Based on initial consultation with external and institutional marketing experts, JAX implemented a strategy including in-person presentation, email campaign, social media, and content marketing. **Results:** From Feb-Sept 2018, JAX sent 25 emails to CT licensed PCPs, posted 135 tweets on JAX channels, provided 58 posts for DPH's Facebook page, posted 7 paid social media ads, and gave 5 presentations to 133 clinicians. 133 new clinicians enrolled in 170 modules with 12,522 pageviews, including 70% PCPs and 7.5% oncology clinicians. Email was most effective in increasing enrollment in a module, paid ads in exposing a large number of individuals to the program, and social media in raising awareness of the program, defined by page views. Few social media users enrolled in modules and presentation attendance was not associated with module enrollment, suggesting presentations and online modules reached different audiences. **Conclusion:** Genomic educators often have limited resources and expertise to support dissemination, but it is critically important. To our knowledge, this is the first study reporting on dissemination of genomic education to a general population of clinicians. Our results suggest that different channels may be useful to achieve outcomes of awareness and enrollment, although the impact of awareness needs further study. Frequent email communication and local presentations were the most effective methods to reach clinicians with genetic education. These preliminary findings should be confirmed through study in other populations.



## PgmNr 628: Identification of *RECQL* pathogenic variants in women undergoing multi-gene panel testing.

### Authors:

K. Theobald <sup>1</sup>; S. Hiraki <sup>1</sup>; E. Sutcliffe <sup>1</sup>; P. Murphy <sup>1</sup>; K. Hruska <sup>1</sup>; R.T. Klein <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) GeneDx Inherited Cancer, GeneDx, Gaithersburg, Maryland.; 2) My Gene Team, Miami, Florida

---

**Background:** The *RECQL* gene, a member of the DNA helicase family, has been reported as a candidate gene for hereditary breast cancer. While initial publications support a significant association with breast cancer risk, subsequent authors demonstrate conflicting evidence and the penetrance of the gene has yet to be delineated. Phenotypic variability, including possible associations with ovarian and colorectal cancer, is not well-defined. We aim to describe the spectrum of *RECQL* variants identified in our testing cohort, and to characterize the clinical phenotype of women with *RECQL* pathogenic and likely pathogenic variants (collectively denoted PV).

**Methods:** We performed a retrospective review of 11,449 women undergoing multi-gene hereditary cancer panel testing for up to 64 genes, including *RECQL*, between December 2017 and February 2019. Personal and family histories of cancer were evaluated. Women with a PV in another gene were excluded from analysis.

**Results:** Among all women tested for *RECQL*, the positive yield was 0.3% (35/11,449). Fourteen different PV were identified among the 35 women. Over half had a recurrent variant: sixteen with the Polish founder variant (c.1667\_1667+3delAGTA), and three with the French Canadian founder variant (c.643C>T). Of the 35 positive women, 37.1% (13/35) had breast cancer with an average age at first diagnosis of 57.7 years. There were 3 patients with ovarian cancer (8.6%), with an average age of diagnosis of 50.3 years, and none with a personal history of colorectal cancer. Family histories of breast, ovarian or colorectal cancer were reported by 26/35 (74.3%), 7/35 (20.0%), and 8/35 (22.9%) women, respectively. Almost half of all women (n=5,452) had a personal history of breast cancer, in whom the positive yield for *RECQL* was 0.24% (13/5,452).

**Conclusion:** PV in *RECQL* are identified in a small proportion of patients undergoing panel testing and with breast cancer. The inclusion of *RECQL* in multi-gene panels may enhance cancer risk screening opportunities. While this case series may be supportive of a potential association with breast cancer, larger, diverse case-control investigations will help to clarify the risk estimations from prior studies and delineate the role of *RECQL* in female breast and other cancers.

## PgmNr 629: Population-based cancer predisposition testing as a component of newborn screening: A cost-effectiveness analysis.

### Authors:

J.M. Yeh <sup>1,2</sup>; N.K. Stout <sup>1,3</sup>; A.B. Chaudhry <sup>2</sup>; M. Gooch <sup>1,3</sup>; P. McMahon <sup>1,3</sup>; K.D. Christensen <sup>4</sup>; L.R. Diller <sup>5</sup>; A.C. Wu <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Harvard Medical School, Boston, MA; 2) Boston Children's Hospital, Boston, MA; 3) Harvard Pilgrim Health Care Institute, Boston, MA; 4) Brigham and Women's Hospital, Boston, MA; 5) Dana-Farber Cancer Institute, Boston MA

---

**Background:** The role of population-based newborn genetic testing for pediatric cancer predisposition has not been studied, despite the availability of cancer surveillance guidelines for early detection in high-risk infants and children.

**Methods:** We developed the Precision Medicine Policy and Treatment (PreEMPT) Model to estimate the value of population-based newborn genomic sequencing (NBS) for a select panel of genes associated with early onset pediatric malignancy. Cohorts of US newborns were simulated under NBS screening vs. usual care, from birth to death. Eleven pediatric cancer predisposition syndromes were included, with mutations in *RET*, *RB1*, *TP53*, *DICER1*, *SUFU*, *PTCH1*, *SMARCB1*, *WT1*, *APC*, *ALK*, *PHO2XB* assigned at birth, using pathogenic/likely pathogenic (P/LP) variant prevalence and disease risks drawn from clinical studies and SEER, ClinVar and gnomAD databases. Newborns with mutations underwent cancer surveillance based on established guidelines. Survival benefit was modeled as a reduction in proportion of advanced disease, cancer deaths, and treatment-related late mortality risks. Costs were based on published literature and national databases. To capture uncertainty, we conducted 100 simulations in which each parameter was sampled from its underlying distribution, and report the mean and 95% uncertainty interval (UI) for all outcomes.

**Results:** In a typical US birth cohort of 4 million newborns, we estimated 3208 (95% UI, 2988-3426) P/LP variant carriers and 2167 (95% UI, 2123-2212) cancer cases in the malignancies associated with this gene panel before age 20 under Usual Care, resulting in 507 (95% UI, 494-519) cancer deaths and 842 (95% UI, 821-867) living with radiation exposure risks. NBS would prevent 8 (95% UI, 6-10) cancers (in *RET* carriers), avert 41 (95% UI, 28-60) deaths via surveillance, and result in a 7% (95% UI, 4-10%) reduction in adult survivors at risk for radiation-associated late mortality. Given a sequencing cost of \$55 (e.g., \$5/gene), the incremental cost-effectiveness ratio (ICER) for NBS was \$218,500 per life year (LY) gained; if sequencing cost declined to \$10 for the gene panel, the ICER fell to \$50,800 per LY.

**Conclusions:** Population-based genetic testing of newborns may reduce mortality associated with pediatric cancers and could potentially be cost-effective as sequencing costs decline. Decision modeling provides a novel approach for understanding the potential benefits and costs of newborn genomic screening.

## PgmNr 630: Implementation of genetic cancer risk assessment (GCRA) in Monterrey, Mexico: Increasing reach, cascade testing and access to risk reduction surgery.

### Authors:

D. Aguilar <sup>1,2</sup>; J.N. Weitzel <sup>3</sup>; K.R. Blazer <sup>3</sup>; D. Castillo <sup>3</sup>; J. Herzog <sup>3</sup>; R. Mejia <sup>3</sup>; J.C. Garcia Marrufo <sup>2</sup>; S. Cardona <sup>1</sup>; C. Villarreal Garza <sup>1</sup>; Clinical Cancer Genomics Community Research Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centro de Cancer de Mama, Tec Salud, Monterrey, Nuevo León, Mexico; 2) Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud; Ave. Morones Prieto 3000, Monterrey, NL, México, 64710; 3) City of Hope Cancer Center, Beckman Research Institute, Division of Clinical Cancer Genomics, 1500 E. Duarte Rd , Duarte, CA, USA, 91010

**Background:** In limited-resource countries like Mexico, GCRA is not routinely provided due to lack of public insurance coverage for counseling, testing and prophylactic surgeries. There is limited infrastructure of clinicians with GCRA expertise, and little awareness among providers and patients. Thus, GCRA services in Mexico are 20 years behind higher income countries, resulting in suboptimal care and missed opportunity for prevention.

The Breast Cancer Center TecSalud established formal GCRA services for underserved patients by recruiting a cancer geneticist who provides *pro bono* care for patients, building academic collaborations for training and research with City of Hope, and enabling free genetic testing through the Clinical Cancer Genomics Community Research Network, supported in part by the Breast Cancer Research Foundation.

**Aim:** To report the uptake of the GCRA program by patients and relatives and of risk reduction surgery at TecSalud from January 2016 - May 2019.

**Methods:** Eligible probands are Mexican women who meet hereditary breast cancer testing criteria and at risk relatives of mutation carriers. Risk reduction procedures (Risk reduction mastectomy: RRM; Risk reduction salpingo-oophorectomy: RRSO) are covered by Seguro Popular and/or the NGO Fundación Santos y de la Garza Evia.

**Results:** GCRA was offered to 317 patients and 96 at risk relatives, with an increasing reach (proportion of those eligible who access the program). There is a significant growth of cascade testing and receipt of risk reduction surgery over time

	2016	2017	2018	2019
<b>Patients tested</b>	71	86	98	62
<b>Relatives tested</b>	0	6	68	22
<b>Prophylactic surgeries</b>	0	4 2 RRM; 2 RRSO	17 5 RRM; 12 RRSO	4 3 RRM; 1 RRSO

**Conclusion:** Increasing GCRA visits represent improving reach, due in part to integration of a skilled multidisciplinary team, outreach at the hospital and increased patient awareness. The expansion of cascade testing and performance of risk-appropriate surgeries reflects a maturing prevention

program. Key determinants of our progress are collaborations that provided us with crucial mentorship, access to genetic assays, as well as the increasing support of local leaders, administrators and benefactors. Sustainment of progress requires increased multilevel efforts and establishment of GCRA as a supported standard of care, so the opportunity for prevention is not missed for vulnerable and underserved high-risk patients and families.

## PgmNr 631: Universal tumor screening for Lynch syndrome: Qualitative perspectives from newly diagnosed colorectal cancer patients.

### Authors:

S. Gille <sup>1</sup>; J. Schneider <sup>1</sup>; A. Firemark <sup>1</sup>; J. Davis <sup>1</sup>; A. Burnett-Hartman <sup>2</sup>; E. Blum-barnett <sup>2</sup>; M. Epstein <sup>3</sup>; J. Lowery <sup>4</sup>; C. Lu <sup>5</sup>; P.A. Pawloski <sup>6</sup>; R.N. Sharaf <sup>7</sup>; S. Liang <sup>8</sup>; J. Hunter <sup>1</sup>; A. Rahm <sup>9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Kaiser Permanente Center for Health Research, Portland, OR; 2) Kaiser Permanente Colorado Institute for Health Research, Aurora, CO; 3) The Meyers Primary Care Institute and Department of Medicine, University of Massachusetts Medical School, Worcester, MA; 4) University of Colorado, Denver, CO; 5) Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Cambridge, MA; 6) HealthPartners Institute, Bloomington, MN; 7) Weill Cornell Medicine, New York City, NY; 8) Palo Alto Medical Foundation Research Institute, Palo Alto, CA; 9) Geisinger Genomic Medicine Institute, Danville, PA

---

Lynch syndrome (LS) is an inherited cancer syndrome, caused by pathogenic variants in the four mismatch repair (MMR) genes (MLH1, MSH2, MSH6 and PMS2) and EPCAM. Individuals with LS are at increased risk of colorectal cancer (CRC), endometrial cancer and other cancers, and their first-degree relatives are at 50% risk to carry the identified variant. Increasingly, health care systems are implementing universal programs for LS which screen all patients newly diagnosed with CRC for molecular signs of LS, followed by genetic testing to confirm the diagnosis. Little is known about patient perspectives on tumor screening for LS, such as importance of knowing and communication about results; this understanding will guide successful implementation of screening programs. We performed semi-structured phone interviews among individuals newly diagnosed with CRC at two integrated health care systems—one with a screening program and one without: Kaiser Permanente Northwest (KPNW) and Kaiser Permanente Colorado (KPCO). Twenty-one patients (10 KPCO/11 KPNW, 52% female, 90% White) were interviewed up to 7 months from date of CRC diagnosis (mean age 63 years, median cancer Stage 2). While most patients (20/21) had limited prior knowledge of LS or tumor screening, almost all (19/21) felt learning LS status was important to them and their family, as it could inform their future medical and lifestyle choices. One third of patients felt that tumor screening should be standard care protocol, and most (17/21) had no concerns with documentation of the results in the medical record. A minority of participants expressed concerns about LS tumor screening which included: how documentation of results could affect future insurance coverage (6/21); privacy and data security (3/21); lack of consent for screening if automatic (2/21); and potential anxiety while awaiting results (2/21). Many (14/21) felt that screening results should be returned at an in-person visit, preferably by a familiar provider (13/21). Overall, most individuals interviewed felt positively about tumor screening for LS and had few concerns about the screening or documentation of results in their medical record. Patients felt they and their families would benefit from tumor screening and knowledge of their LS status. Importantly, 62% (13/21) of participants also had a blood relative seen by the same health care system, highlighting a possible opportunity for LS testing of blood relatives by health care systems.

## PgmNr 632: No woman left behind: Is universal *BRCA1/2* testing possible for women with high grade serous ovarian cancer?

### Authors:

L. Lynch <sup>1</sup>; K. Bell <sup>1,2</sup>; S. Lopiccolo <sup>3</sup>; B. Loiseau <sup>4</sup>; K. Zbuk <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Juravinski Cancer Centre- Hamilton Health Sciences, Hamilton, Ontario, Canada; 2) Department of Oncology, McMaster University, Hamilton, Ontario, Canada; 3) Michael G DeGroote School of Medicine, McMaster University, Hamilton, Ontario, Canada; 4) Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada

---

**Background:** The Society of Gynecologic Oncology of Canada has created the “No Woman Left Behind” campaign, seeking to ensure that every woman with high grade serous ovarian cancer (HGSOC) is informed of her *BRCA* mutation status. Women with *BRCA1/2* mutations may benefit from targeted therapies and their blood relatives may benefit from predictive genetic testing and risk reducing measures. We previously published that 62% of patients with HGSOC at our centre were referred for genetic counselling/testing.

**Objective:** To document the care pathway for women referred to the Juravinski Cancer Center (JCC) in Hamilton, Ontario, Canada with: (i) a diagnosis of ovarian cancer (OC), fallopian tube cancer (FTC), or primary peritoneal cancer (PPC) (ii) a suspicious pelvic mass or (iii) clinical symptoms suggestive of OC to determine if there is a common touch point that would allow for 100% of this patient population to access *BRCA1/2* testing.

**Methods:** A retrospective chart review was conducted on 284 new patient referrals to the gynecologic oncology service at the JCC from January 2016-December 2017. Data related to the patient’s diagnostic assessment, surgery and/or treatment was extracted to construct patient care pathways. These pathways were analyzed to identify the area of greatest convergence.

**Results:** Outpatient blood work was obtained for 94.7% (269/284) of patients by their first appointment at the JCC. Eleven patients had bloodwork at their second or subsequent appointment. Beyond the first appointment, there was significant divergence in care pathways. There were four individuals who did not have blood work due to poor health/functional status.

**Conclusions:** Blood samples for germline *BRCA1/2* testing could be obtained at the first appointment as bloodwork is a standard component of the order set for all women with (i) a confirmed diagnosis of OC, FTC, or PPC (ii) a suspicious pelvic mass or (iii) clinical symptoms suggestive of OC. This would allow the majority of these patients to have the opportunity to provide a DNA sample which can be used for *BRCA1/2* testing upon patient/family consent and/or confirmation of the diagnosis. While our study shows it was not possible to obtain a blood sample from 100% of women in our patient population, our proposed suggestion of obtaining blood at the patient’s first appointment would be a 32.7% improvement on the previously reported low referral rates for *BRCA1/2* testing.

## PgmNr 633: Exploring healthcare providers perspective on the clinical utility of cfDNA testing.

### Authors:

M. Clausen<sup>1</sup>; L. Oldfield<sup>3</sup>; C. Mighton<sup>1,2</sup>; A. Sebastian<sup>1,2</sup>; S. Shickh<sup>1,2</sup>; S. Manjrekar<sup>1</sup>; J. Ho<sup>1</sup>; A. Pollett<sup>2,5,6</sup>; R. Kim<sup>3,4,5</sup>; T. Pugh<sup>2,3,7</sup>; Y. Bombard<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) St. Michael's Hospital, Li Ka Shing Knowledge Institute, Toronto, Ontario, Canada; 2) University of Toronto, Toronto, ON, Canada; 3) University Health Network, Toronto, ON, Canada; 4) The Hospital for Sick Children, Toronto, ON, Canada; 5) Mount Sinai Hospital, Sinai Health System, Toronto, ON, Canada; 6) Cancer Care Ontario, Toronto, ON, Canada; 7) Ontario Institute for Cancer Research, Toronto, ON, Canada

---

### Background

Individuals with an inherited cancer syndrome undergo costly and invasive serial screenings that are ineffective at detecting all associated malignancies. Detecting cancer early is critical to improving patient outcomes and research has shown that cell-free DNA (cfDNA) analysis can be utilized as a minimally invasive and effective cancer screening modality. While implementation of cfDNA testing into the clinic for early cancer detection is promising, no studies to date have consulted with healthcare providers and researchers on the realities surrounding the utility and implementation of cfDNA screening.

### Aim

Understand healthcare providers and researchers perceptions of the clinical utility of cfDNA testing for early cancer detection and identify the factors impacting clinical uptake.

### Methods

We are conducting 25-30 semi-structured interviews with researchers and healthcare practitioners, including oncologists, medical geneticists, genetics counsellors, and primary care providers. Interviews will be transcribed and analyzed thematically. Data collection and analysis will be completed by August of 2019.

### Results

Interviews with 10 practitioners and researchers have been completed. Preliminary analyses reveal that the utility of cfDNA screening depends on whether there exist effective screening modalities for particular cancer types. For example, for patients with BRCA1/2 mutations cfDNA could serve as a screening modality for ovarian cancer since no effective screening strategy exists, and as an adjunct test for breast cancer to help decipher benign breast lumps from malignancies. However, participant's perceived limited utility for screening colorectal cancers in Lynch syndrome carriers, given the effectiveness of colonoscopies at decreasing cancer risk. When discussing barriers and facilitators to uptake, participants expressed that a cfDNA screening test would need to have a low false positive rate, strong scientific evidence base and be accompanied by end-user education and practice guidelines. Cost and the inability of non-specialists to interpret results were seen as potential barriers.

## Discussion

Our preliminary analyses provide insights into how practitioners may use cfDNA in care and how the value of the test is largely dependent on whether effective screening strategies are available for a hereditary cancer syndrome. This study provides valuable information that will enable the optimal adoption of cfDNA into clinical care.



## PgmNr 634: Informed consent for genomic sequencing in diverse populations: Does diversity impact consent time?

### Authors:

L. Desrosiers<sup>1,2</sup>; S. Scollon<sup>1,2</sup>; A. Recinos<sup>1,2</sup>; R. Raesz-Martinez<sup>1,2</sup>; T. Wang<sup>3</sup>; J. Robinson<sup>4</sup>; M. Jibaja-Weiss<sup>3</sup>; D.W. Parsons<sup>1,2</sup>; A. McGuire<sup>4</sup>; S. Plon<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Baylor College of Medicine, Houston, TX.; 2) Texas Children's Cancer Center, Texas Children's Hospital, Houston, TX; 3) Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, TX; 4) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX

**Background:** Most studies of the informed consent (IC) process for genomic sequencing (in both research and clinical settings) have been conducted in institutions with limited racial/ethnic patient diversity. Findings demonstrated large variability in the estimated length of time to provide IC that differed by factors including experience, provider and population. The Texas KidsCanSeq (KCS) Study, an NIH funded Clinical Sequencing Evidence-Generating Research consortium project is studying implementation of genomic sequencing for pediatric cancer patients (germline for all subjects and tumor included for high risk patients) across the diverse population of Texas. Analysis of the time and effort required to obtain consent provides an opportunity to explore the IC process among ethnically diverse participants.

**Methods:** Study staff trained in IC offered KCS enrollment to patients treated in the Texas Children's Hospital (TCH) Cancer Center. The IC process was performed in the family's preferred language, with a native Spanish-speaker consenting in Spanish. All study materials (including consent forms and videos) were available in English and Spanish. Data analyzed of the IC process includes language, number of times parents discussed the study with staff (IC attempt), and total time spent in direct patient contact until enrollment. Participant-reported ethnicity was used to compare consent characteristics of enrolled patients in 3 groups (Non-Hispanic consented in English, Hispanic consented in English and Hispanic consented in Spanish) via Kruskal-Wallis Test.

**Results:** Enrollment opened in June 2018. To date, 129 families have consented and are included in the analysis. Overall, the median was 1 attempt and the maximum was 7 with 12% enrolled with more than 3 attempts. Time required to enroll participants ranged from 25 to 165 minutes with a median of 60 minutes. Among this cohort, 57 (44%) are Hispanic including 23 (18%) consented in Spanish. Among the 3 groups, the majority of participants were enrolled with 1 attempt (52.9%-56.5%) and average consent times were 62-73 minutes. There was no difference in IC attempts or time between the 3 groups ( $p=0.99$  and  $p=0.16$ , respectively).

**Conclusion:** In a setting with provision of study materials in English and Spanish and native speakers describing the study, there was no significant difference in the staffing requirements and parental time commitment to complete enrollment into a complex cancer genomic study.

## PgmNr 635: Sharing exome sequencing results with family members: Experiences of participants in a clinical sequencing study of hereditary colorectal cancer and polyposis.

### Authors:

S.M. Fullerton <sup>1</sup>; S. Makhnoon <sup>2</sup>; R. Hart <sup>2</sup>; L.M. Amendola <sup>3</sup>; X. Niu <sup>4</sup>; D.L. Veenstra <sup>5</sup>; G.P. Jarvik <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Bioethics and Humanities, University of Washington, Seattle, WA; 2) Institute of Public Health Genetics, University of Washington, Seattle, WA; 3) Departments of Medicine (Division of Medical Genetics) and Genome Sciences, University of Washington, Seattle, WA; 4) Department of Epidemiology, University of Washington, Seattle, WA; 5) Department of Pharmacy, University of Washington, Seattle, WA

---

**Background:** Family sharing of genetic test results improves the clinical and economic value of genetic testing for cancer risk. However, relatively little is understood about the impact of result type on sharing motivations and behavior. To learn more about these features of familial communication we interviewed colorectal cancer and polyposis (CRCP) patients with positive, negative, or uncertain results who reported either no sharing, or extensive sharing, with relatives. **Methods:** Patients referred for genetic testing were enrolled in the New EXome Technology in Medicine (NEXT Medicine) Study, a Clinical Sequencing Exploratory Research (CSER) consortium Randomized Controlled Trial (RCT) that compared exome sequencing versus usual care for genetic evaluation for CRCP. We interviewed 15 participants one month post-disclosure: 8 who reported no immediate sharing (2 with negative results, 2 positive, 4 VUS) and 7 who reported sharing with 4 or more relatives (2 negative, 2 positive, 3 VUS). Interviews were designed to elicit attitudes toward results received, familial sharing behavior, and relatives' responses to result sharing. Interviews were audio-recorded, transcribed, and subjected to a directed content analysis. **Results:** Interviewees described a variety of family communication practices in response to receiving CRCP results. Those who regarded the information received as useful, either because it explained personal or family history of disease or because it suggested a lower than anticipated risk, appeared more willing and able to share results with their relatives than those who remembered their results as inconclusive or unimportant. Close relatives were informed more often than distant ones due to barriers such as infrequent contact, estrangement, or ambiguity regarding responsibility for sharing results. Interviewees described relatives responding differently to the genetic risk information received: some family members expressed worry and became proactive in passing on the information to other relatives, while other family members responded with skepticism or dismissed the information. **Conclusions:** Familial communication of CRCP-related exome sequencing results was complex and dependent on individual factors and relationships within families. In this study, the perceived utility of the test result was an important determinant of how widely and in how much detail the result was shared with family members.

## PgmNr 636: Development and testing of a pre-test genetic counseling educational tool for inherited cancer.

### Authors:

D. Cragun <sup>1</sup>; A. Tezak <sup>2</sup>; A. Weidner <sup>2</sup>; G. Wiesner <sup>2</sup>; B. Zuniga <sup>2</sup>; T. Pal <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Public Health, Univ South Florida, Tampa, Florida.; 2) Ingram Cancer Center, Vanderbilt, Nashville, TN

**Introduction:** National organizations recommend genetic counseling (GC) prior to testing for hereditary cancer. Yet lack of trained genetic specialists limits the ability to meet the increasing demand due to expanded test indications. GC could be streamlined with automated and scalable tools to meet this demand. We sought to develop and test a web-based educational tool to cover pre-test GC elements.

**Methods:** A 12-minute web-based educational tool ("GeneCounsel") was iteratively developed to cover the standard pre-test GC elements and tested among patients without prior counseling scheduled for a new patient appointment in the Vanderbilt Hereditary Cancer Clinic. Participants consented and completed pre- and post-tool surveys online before and after viewing the tool. The surveys assessed knowledge, attitudinal values about genetic testing, informed/empowered decision-making, and health literacy. Participants were randomly assigned to two groups, of which only Group A completed knowledge questions on the pre-tool survey to evaluate for a priming effect. Pre- and post-tool scores were compared, and group comparisons were made based on randomization and literacy level.

**Results:** 122 patients were approached for the study to achieve a sample size of 100 participants (50 per group). Participants included 77% female, 56% with a cancer diagnosis, and a mean age of 49. Group A completed the 14 knowledge questions on both the pre- and post-tool surveys, with a median number correct of 7 at baseline which significantly increased to 12 after viewing the tool ( $p < 0.001$ ). The median correct for Group B was 11 on the post-tool survey, which was not significantly different from Group A ( $p = 0.675$ ). Pre- and post-tool surveys indicated: 1) Proportions who felt informed and empowered were 32% and 77% respectively ( $p < 0.001$ ); and 2) Attitudinal values did not change significantly ( $p = 0.607$ ). Median pre-tool knowledge scores among low and high literacy groups were 4 and 8, respectively ( $p = 0.016$ ); this gap narrowed to 11 and 12 after viewing GeneCounsel ( $p = 0.214$ ).

**Conclusion:** Viewing GeneCounsel resulted in significant gains in knowledge across all literacy levels and the majority of individuals felt informed and empowered to make decisions about testing, based on accurate knowledge and would be congruent with patients' values. These findings support the potential to automate components of pre-test GC to streamline the delivery of GC and testing services for inherited cancers.

## PgmNr 637: Large scale group genetic counselling: A novel model for genetic counselling and genetic testing for hereditary cancer.

### Authors:

Z. Lohn <sup>1</sup>; A. Fok <sup>2</sup>; H. Derocher <sup>1</sup>; M. Richardson <sup>3</sup>; S. Mung <sup>1</sup>; J. Nuk <sup>1</sup>; J. Yuson <sup>1</sup>; M. Jevon <sup>1</sup>; K. Schrader <sup>1,4,5</sup>; S. Sun <sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Hereditary Cancer, BC Cancer, Vancouver, BC, Canada; 2) Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada; 3) Department of Interdisciplinary Oncology, The University of British Columbia, Vancouver, BC, Canada; 4) Department of Molecular Oncology, The University of British Columbia, Vancouver, BC, Canada; 5) Department of Medical Genetics, The University of British Columbia, Vancouver, BC, Canada; 6) Division of Medical Oncology, The University of British Columbia, Vancouver, BC, Canada

**Introduction:** Meeting the demand for hereditary cancer genetic services is increasingly challenging which has led to the development of novel service delivery models. To shorten the wait time for non-urgent patients, we piloted a large scale group genetic counselling (GC) model that offered genetic testing to unaffected individuals within a publicly funded population-based healthcare system.

**Materials and Methods:** All patients were recruited from the Hereditary Cancer Program in British Columbia, Canada and were eligible to participate if they met provincial criteria for hereditary cancer risk assessment, had no personal history of cancer, had no prior genetic testing performed in the family, and had no living testable family member residing in BC or the Yukon. Pre-test group GC (up to 50 patients per group) was compared to traditional one-on-one GC. Patients who pursued testing underwent a post-test one-on-one GC results session. Patient reported outcome measures included the Genetic Counselling Outcome Scale (T1: immediately prior to pre-test GC and T2: 4 weeks after post-test GC), a satisfaction survey (immediately after pre-test GC) and the Multidimensional Impact of Cancer Risk Assessment (4 weeks after post-test GC) for those undergoing testing.

**Results:** To date, 392 patients have been seen (190 in the group arm and 202 in the traditional one-on-one arm). Eight group sessions have been held (median group size: 26 patients). Patients in both arms showed high satisfaction with the majority of patients reporting that the appointment was helpful (98% group arm, 99% traditional arm), that they understood the information presented (99% group arm, 99% traditional arm), and that their questions were answered (82% group arm, 99% traditional arm). Of the 299 patients who have completed testing, 5.5% (N=16) had a pathogenic or likely pathogenic variant identified including: *BRCA1* (2), *BRCA2* (2), monoallelic *MUTYH* (2), *MLH1* (1), *MSH2* (1), *MSH6* (1), *PALB2* (1), *MITF* (1), *FH* (2), *BRIP1* (1), *NBN* (1), and *TSC1* (1). A small portion of patients (7%) declined participation in the group session because they preferred one-on-one GC.

**Conclusions:** Large scale group genetic counselling is feasible and acceptable to patients, and represents a new streamlined model for cancer genetic counselling.

## PgmNr 638: Is there a lack of personal and family history information in genetics referrals for evaluation of hereditary cancer risk?

### Authors:

K. Muessig<sup>1</sup>; J. Zepp<sup>1</sup>; M. Gilmore<sup>1</sup>; J. Hunter<sup>1</sup>; S. Knerr<sup>2</sup>; E. Shuster<sup>1</sup>; H. Feigelson<sup>3</sup>; K. Goddard<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Health Research, Kaiser Permanente, Portland, OR; 2) Department of Health Services, University of Washington, Seattle, WA; 3) Institute for Health Research Kaiser Permanente Colorado, Denver, CO

---

Genetics providers are facing increasing numbers of referrals for evaluation of hereditary cancer risk. Providing detailed information about the indicators of hereditary risk, such as personal and family cancer history, can facilitate timely patient triaging for appointments. We evaluated referrals to the Kaiser Permanente Northwest (KPNW) genetics department containing diagnosis codes associated with hereditary cancer risk made between June-December 2017. Referrals were excluded if the patient was pregnant at the time of the referral or age  $\leq 17$ , or the referral did not contain personal or family history diagnosis codes for cancers associated with hereditary risk. A genetic counselor evaluated content in the medical chart, including age, notes, and diagnosis codes, to characterize referrals based on two criteria. First, the genetic counselor coded the referrals on whether they contained enough information to determine the patient's eligibility for hereditary cancer evaluation. If there was enough information, the ACMG/NSGC practice guidelines were used to determine if patients met criteria for genetic testing. Demographic and clinical characteristics were also abstracted. A total of 1047 eligible referrals were included in the analysis for 927 unique patients. The majority of patients who received a referral were female (93%), White (88%), and less than 50 years of age (50%) at the time of the referral. Thirty-nine percent had a personal history of cancer. Of those, 26% had hereditary breast or ovarian related cancer, 7% were associated with Lynch syndrome, and 6% reported other cancer types. Referred patients had been KPNW members for an average of 14 years (variance 13 years). Of eligible referrals, 68% contained enough information to determine whether or not to offer genetic evaluation; of those, 63% referrals also met ACMG/NSGC criteria for genetic counseling. Referring clinic location (metropolitan area versus outlying areas) did not affect the amount of information provided. Referring provider specialty was associated with the amount of information provided, with Oncology (68%) having the highest proportion of referrals with enough information to assess whether to offer genetic counseling. Most patient referrals to KPNW genetics for evaluation of hereditary cancer risk include enough information to determine eligibility for further evaluation, however more than a third of referring providers would benefit from additional training or support.

## PgmNr 639: A comprehensive cancer center's experiences utilizing alternative service delivery models for cancer genetic testing.

### Authors:

A. Murad <sup>1</sup>; E. Koeppe <sup>2</sup>; J. Osborne <sup>1</sup>; D. Samad <sup>1</sup>; M. Jacobs <sup>1</sup>; S. Austin <sup>1</sup>; M. Marvin <sup>1,3</sup>; E. Stoffel <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

- 1) Michigan Medicine, Department of Internal Medicine, Division of Genetic Medicine, Ann Arbor, MI.;
- 2) Michigan Medicine, Department of Internal Medicine, Division of Gastroenterology, Ann Arbor, MI.;
- 3) Michigan Medicine, Department of Human Genetics, Ann Arbor, MI.

---

Genetic medicine is an integral part of cancer care. The number of individuals who qualify for genetic testing is increasing and the traditional model of pre- and post-test counseling provided by a genetic counselor (GC) or geneticist in a dedicated cancer genetics clinic is struggling to meet growing demands. Here we describe one institution's experiences implementing non-traditional models of cancer genetic service delivery.

Between 3/1/18 – 5/31/19, genetics services were integrated into 3 specialty clinics within the University of Michigan Rogel Cancer Center as follows: (1) a nurse coordinator ordered genetic testing on select patients in a multidisciplinary pancreatic (MDP) cancer clinic; (2) a GC staffed a multidisciplinary colorectal (MDCRC) cancer clinic; and (3) a GC was embedded as part of the care team in an endocrine oncology (EndoOnc) clinic. We queried the number of patients who received a genetic evaluation in each setting and evaluated factors contributing to successful genetic service delivery.

269 patients underwent genetic evaluation, including 40 in the MDP, 81 in MDCRC, and 126 in EndoOnc. The EndoOnc model was the most successful in identifying patients for genetic evaluation and in acceptance of a genetic counselor into the clinical team, in large part due to strong physician support within the clinic. Barriers to service delivery were encountered when non-genetic providers struggled to appropriately identify high risk individuals, particularly in cases where an individual's history might not meet traditional insurance guidelines but an evaluation was still indicated. This occurred most commonly in the MDP clinic, where genetic evaluation criteria are complex and multifaceted, and differences exist between guidelines for a genetic evaluation vs. insurance guidelines for genetic testing.

In conclusion, integrating genetic services into multidisciplinary cancer care clinics can help meet high demands for genetic services. Appointments can be more streamlined and eliminate the need for extra clinic space, allowing a greater number of individuals to undergo a genetic evaluation than would be possible utilizing only the traditional model of referring patients to cancer genetics. Non-traditional service delivery models are more likely to be successful when there is physician-support, the criteria for what constitutes high risk is clear, and when a genetic counselor is incorporated as an established part of the multidisciplinary care team.

## PgmNr 640: Genetic counseling experience of a large scale web-based return of *BRCA2* research results in Iceland.

### Authors:

E. Thorolfsson<sup>1</sup>; V. Stefansdottir<sup>1</sup>; H.B. Hognason<sup>1</sup>; B.B. Gunnarsdottir<sup>1</sup>; T. Jonsdottir<sup>1</sup>; A. Ulfarsdottir<sup>1</sup>; J.J. Jonsson<sup>1,2</sup>; H.T. Bjornsson<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Landspítali National University Hospital, Reykjavik, Iceland; 2) Faculty of Medicine, University of Iceland

---

**Background.** Increasingly, genetic results from research are being returned to participants. deCODE Genetics, a genetic research company based in Iceland, has genotypic data from almost half of the Icelandic population (160.000 Icelanders). Almost 1% of Icelanders carry a known pathogenic *BRCA2* founder variant, NM\_000059.3:c.767\_771delCAAAT. In an effort to report an important actionable result back to the population, deCODE genetics recently (May 2018) opened the website arfgerd.is to enable Icelanders to find out their *BRCA2* founder mutation status.

**Methods.** Positive individuals were advised to contact the genetic counseling unit (GCU) at Landspítali-National University Hospital for confirmatory testing and counseling. Here we have quantified some observations from this experience. We report a one year clinical experience of the GCU from the opening of the webpage up until the end of May 2019. At the time, 46.000 individuals had signed onto the website.

**Results and discussion.** A total of 352 (average age 44 years), had received a positive result by the end of the period. Based on information from local support groups, we estimate that at least 40 of the 352 had received their results and genetic counseling prior to signing up. A total of 195 (63%) of the remaining 312 had contacted the GCU by the end of May 2019. A total of 117 (37%) had not, without there being a clear reason for their absence. Some may yet come, and others may seek information and/or surveillance in other venues (through family, online or outside Iceland). In addition, 129 relatives contacted the GCU directly after receiving news of the pathogenic variant within the family. Of the *BRCA2* positive (n=195), 74 (38%) knew about the pathogenic variant in the family prior to signing up to the website, while 121 (62%) did not. The majority were grateful for being able to access this information. A subset reported being surprised to learn the results (46, 24%), while others were expecting them (61, 31%). Some expressed their concerns with the positive results (45, 23%), while others (24, 12%) communicated that they were not concerned. The increased need for counseling and surveillance was significant, especially given the small size of the units involved at Landspítali-National University Hospital. However, given the frequency of this founder variant in Iceland, it offers a unique opportunity to study how to best deliver actionable genetic results to a large subset of a single nation.



## PgmNr 641: Patient empowerment following genetic counselling: A Patient Reported Outcome Measure (PROM) study.

### Authors:

J. Yuen <sup>1</sup>; S.Y. Lee <sup>1</sup>; E. Courtney <sup>1</sup>; J. Lim <sup>2</sup>; S.T. Li <sup>1</sup>; H. Soh <sup>1</sup>; Y. Chen <sup>1</sup>; M. McAllister <sup>3</sup>; E.K. Fenwick <sup>4,5</sup>; J. Ngeow <sup>1,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cancer Genetics Service, Division of Medical Oncology, National Cancer Centre Singapore, Singapore; 2) Biostatistics Unit, Department of Clinical Trials and Epidemiological Sciences, National Cancer Centre Singapore, Singapore; 3) Centre for Medical Education, School of Medicine, Cardiff University, Cardiff, UK; 4) Duke-NUS Medical School, Singapore, Singapore; 5) Singapore Eye Research Institute, Singapore, Singapore; 6) Oncology Academic Clinical Program, Duke-NUS Medical School, Singapore, Singapore; 7) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

---

**Background:** Measurement of patient-derived benefits from cancer genetics services (CGS) is lacking but is essential to guide service evaluation and quality improvements. We measured improvement in patient empowerment, if any, following genetic counselling in Singapore using a culturally-adapted version of the Genetic Counselling Outcome Scale (GCOS-24); with the secondary intention to identify and understand the factors associated with change in empowerment.

**Methods:** The GCOS-24 was administered to 155 patients, who attended the CGS at the National Cancer Centre Singapore (NCCS) (May 2016 - May 2017), at pre- and post-counselling or testing stages. There were 106 patients who underwent genetic testing. Individual pre- and post-counselling or testing responses were subjected to Rasch analysis; the scale was subsequently split into Cognitive Control (CC) and Emotional Control (EC) domains. Associations of baseline characteristics with changes in CC and EC scores were assessed using multiple regression analysis.

**Results:** Both CC and EC scores showed significant improvement following genetic counselling and testing. While all items in the CC domain showed increases at follow-up, aspects of EC related to alleviating feelings of being upset ( $p = 0.88$ ) and hopelessness ( $p = 0.2$ ) did not demonstrate significant improvement. Having a family history of cancer was associated with less improvement in CC ( $p = 0.035$ ) while genetic testing uptake was associated with greater improvement in CC ( $p = 0.001$ ).

**Conclusions:** We found significant improvements in empowerment following genetic counselling and testing, demonstrating the value of cancer genetic services. Interventions addressing hope and negative emotions associated with genetic conditions and testing may be warranted.



## PgmNr 642: Uptake and predictors of genetic testing in unselected pancreatic ductal adenocarcinoma.

### Authors:

C. Cremin<sup>1,3</sup>; M.K-C. Lee<sup>2,3</sup>; Q. Hong<sup>1</sup>; S. Kalloger<sup>3,4</sup>; J. Karasinska<sup>3</sup>; A. Mackenzie<sup>1</sup>; M. McCullum<sup>1</sup>; J. Nuk<sup>1</sup>; D. Schaeffer<sup>3,4</sup>; S. Sun<sup>1,2</sup>; D. Renouf<sup>2,3</sup>; K.A. Schrader<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Hereditary Cancer Program, BC Cancer, Vancouver, BC, Canada; 2) BC Cancer, Division of Medical Oncology, Vancouver, BC, Canada; 3) Pancreas Centre BC, Vancouver, BC, Canada; 4) Department of Pathology & Laboratory Medicine Vancouver General Hospital, Vancouver, BC, Canada

**Objectives:** Recent guidelines recommend consideration of genetic testing (GT) for all patients with pancreatic ductal adenocarcinoma (PDAC). We implemented a familial pancreatic cancer program to offer expedited appointments (within 90 days of referral if non-urgent), flexible service models (in-person, telephone and group counselling) and research-funded genetic testing using the 30 gene Color panel for any referred PDAC patient. The aim of this study was to assess uptake of GT, identify potential barriers and suggest areas for more efficient test coordination.

**Methodology:** Retrospective chart review was approved by the BC Cancer Research Ethics Board and undertaken for all PDAC patients referred to the FPCP from June 1, 2016 to January 31, 2019. Demographics, medical/family history, disease-specific data, mode of service delivery and testing uptake were compared across the cohort to determine predictors of GT uptake. Main motivations for attending the appointment were solicited by survey. Reasons for decline of testing were reviewed.

**Results:** A total of 311 patients were referred over the 2.5 year study period (55% female, 52% European ancestry, mean age at diagnosis was 64 years (SD=11). Referrals came from oncologists (71.7%), surgeons (7.7%), GP (7.1%), other (6.8%) and patient self-referral (6.4%). Disease stage was metastatic in 42.8%. Median time from cancer diagnosis to genetics referral was 47 days (2-5809) and from referral to appointment was 55.4 days (0-244). 79% (n=247) attended a genetic counselling appointment and 63% (n=196) completed GT. Patients with earlier stage disease stage were more likely to attend an appointment (p-value <0.05). Common reasons for not having an appointment booked were deteriorating health status or death (n=31; DNA banked in 40%); patient declined (n=27) and patient could not be reached (n=6). For patients who did not complete GT, 22 passively declined by not providing a sample; 12 patients died within one month of their appointment and did not complete GT; and 17 patients actively declined testing. Factors significantly associated with GT uptake (p-value <0.05) were age at diagnosis, smoking history, geographic location and mode of service delivery.

**Conclusions:** Given the sub-optimal uptake of genetic counselling/testing in PDAC, up front DNA banking for PDAC patients at initial oncologic visit may be the most effective approach to facilitate future hereditary cancer assessment for at-risk relatives.

## PgmNr 643: *TP53* variant discordance between and within families enrolled in the NCI's Li Fraumeni syndrome study.

**Authors:**

M.N. Frone; D.R. Stewart; J.T. Loud; S.A. Savage; P.P. Khincha

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical Genetics Branch, Division of Cancer Genetics and Epidemiology, National Cancer Institute, National Institutes of Health, Bethesda, MD.

---

**Introduction:** The use of ACMG-AMP guidelines and semiquantitative approach to clinical variant curation improved concordance and accuracy of variant classification. However, discordant variant classification persists and poses a significant challenge for medical management. We evaluated discordant variant interpretations resulting from clinical genetic testing for germline *TP53* variants in families enrolled in the NCI's Li Fraumeni Syndrome (LFS) Study.

**Methods:** *TP53* test results were obtained from >300 individuals enrolled in the NCI IRB-approved LFS study between Sept 2011 and Feb 2019. A discordant test result was defined as a report of pathogenicity that differed between two clinical testing laboratories, between a testing laboratory and the ClinVar database, or between the lab or ClinVar database and the final variant classification provided by our study after review.

**Results:** 86 unique variants were identified in 143 LFS families. 54/143 (38%) were discordant by at least one interpretation (i.e., Pathogenic/Likely Pathogenic (P/LP), or LP/Variant of Uncertain Significance (VUS), or P/VUS). 14 variants in 16 families (11%) had discordant interpretations that would lead to a difference in medical management (P or LP vs. VUS). Interfamilial discordance - i.e., unrelated families sharing the same variant but given discrepant variant classifications, was observed in 4 families (2 different variants). Intra-familial discordance was observed in 6 families, with family members carrying the same variant but receiving discordant interpretations.

**Conclusions:** Discordant variant interpretations are frequent and may lead to differences in medical management. Inter- and intra-familial *TP53* variant discordance results in a significant ethical medical challenge for clinicians of how to follow individuals with an identical variant receiving discordant variant interpretations and thus different recommendations for medical management. In the case of a *TP53* P/LP variant vs. VUS; there is a substantial difference between getting comprehensive cancer screening and potential prophylactic surgeries vs. no clinical follow-up. The scope of this problem may not be fully appreciated by providers who use only one or two genetic testing laboratories or who are unaware of discordant calls for a given variant. Centralized data sharing of genetic variants and creation of variant curation consensus guidelines are crucial for unified variant interpretation and optimal patient care.

## PgmNr 644: Cancer risk management among women with mutations in *BRCA* genes compared to other inherited breast cancer genes.

### Authors:

T. Pal; A. Weidner; A. Tezak; K. Clouse; D. Cragun

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medicine, Hereditary Cancer Clinic, Nashville, Tennessee.

---

Authors: Tuya Pal, Anne Weidner, Ann Tezak, Kate Clouse, Deborah Cragun

**Introduction:** Identification of inherited breast cancer risk through genetic testing may guide medical care. Given the expanding numbers of women identified with a pathogenic/likely pathogenic (P/LP) variant in genes other than *BRCA1* and *BRCA2* (*BRCA*), we sought to compare cancer risk management(CRM) practices between these carriers.

**Methods:** Female participants with a P/LP variant in *BRCA*, *PALB2*, *CHEK2*, or *ATM* were surveyed about CRM practices. Frequencies of CRM based on screening (i.e., mammograms and/or breast MRI, among those with at-risk breast tissue) and risk-reducing surgery (i.e., risk-reducing mastectomy (RRM) and risk-reducing salpingo-oophorectomy (RRSO)) were compiled and compared. Descriptive statistics, bivariate analyses, and non-parametric tests were performed to assess for associations.

**Results:** The 238 participants with P/LP variants were as follows: 186 *BRCA*, 28 *PALB2*, 15 *CHEK2*, and 9 *ATM*. Median age of participants was 54 and most were non-Hispanic white (89%) with a prior cancer diagnosis (59%). Rates of RRM (defined as at least one breast removed prophylactically regardless of a prior breast cancer diagnosis) was 65.1% among *BRCA* carriers, with lower rates among *PALB2*, *CHEK2*, and *ATM* carriers (32.1%, 33.3%, and 33.3%, respectively). Rates of RRSO were appropriately higher among *BRCA* carriers (89.8%), but were also relatively common among *PALB2*, *CHEK2*, and *ATM* carriers (32.1%, 33.3%, and 66.7%). Among those with at-risk breast tissue, most had screening for breast cancer through mammogram and/or MRI. No significant differences were found between *BRCA* carriers vs. others in terms of insurance, employment, income, house size, education, race/ethnicity, cancer status, and/or involvement of a genetic healthcare provider (all  $p>0.05$ ).

**Conclusion:** Our results suggest that adherence to CRM for *BRCA* carriers is appropriately high. However, RRM among those with *PALB2*, *CHEK2*, and *ATM* P/LP variants are also frequently observed, despite the lack of practice guidelines to recommend RRM. Even more concerning are the appreciable RRSO rates in the absence of practice guidelines or data to confirm high ovarian cancer risk. Our data suggests that many patients with P/LP variants in genes other than *BRCA* receive surgeries for which guidelines are lacking. These efforts highlight the tremendous need to promote risk appropriate care and reduce over-treatment among these individuals.

## PgmNr 645: Prepared to live with the unknown: Adult cancer patients' experiences with uncertainty when deciding to learn incidental genomic sequencing results.

### Authors:

C. Mighton<sup>1,2</sup>; M. Clausen<sup>2</sup>; R. Kodida<sup>2</sup>; S. Shickh<sup>1,2</sup>; E. Reble<sup>2</sup>; A. Sebastian<sup>1,2</sup>; J. Lerner-Ellis<sup>3,4</sup>; R.H. Kim<sup>5,6,7</sup>; A. Eisen<sup>8</sup>; C. Elser<sup>6,9</sup>; M. Aronson<sup>10,11</sup>; W.J. Ungar<sup>1,12</sup>; R.Z. Hayeems<sup>1,12</sup>; R. Upshur<sup>1,13</sup>; Y. Bombard<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Health Policy, Management & Evaluation, University of Toronto, Toronto, ON; 2) Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health System, Toronto, ON; 3) Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Sinai Health System, Toronto, ON; 4) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON; 5) Princess Margaret Cancer Centre, University Health Network, Toronto, ON; 6) Department of Medicine, University of Toronto, Toronto, ON; 7) Clinical Genetics and Metabolics, The Hospital for Sick Children, Toronto, ON; 8) Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, ON; 9) Marvelle Koffler Breast Centre, Mount Sinai Hospital, Sinai Health System, Toronto, ON; 10) Department of Molecular Genetics, University of Toronto, Toronto, ON; 11) Zane Cohen Centre for Digestive Diseases, Mount Sinai Hospital, Sinai Health System, Toronto, ON; 12) Program of Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON; 13) Bridgepoint Health, Sinai Health System, Toronto, ON

---

**Background:** Genomic sequencing (GS) can inform diagnosis and management of hereditary conditions, such as hereditary cancer syndromes, and can also identify genetic predisposition for additional diseases (secondary findings or incidental results, IR). There is substantial uncertainty about when or if diseases associated with IR will occur, as well as if and how IR should be sought, reported, and managed. Patients may therefore experience uncertainty when making decisions about learning IR.

**Aim:** To explore how adults with cancer perceive, appraise and manage uncertainty when deciding whether or not to learn IR.

**Methods:** This was a qualitative, interpretive description study among adults receiving GS through an ongoing RCT. Participants selected IR from five categories: 1) medically actionable diseases, 2) common disease SNPs, 3) rare Mendelian diseases, 4) early-onset neurological diseases, and 5) carrier status. Semi-structured interviews were conducted with participants after selecting but before receiving IR. Transcripts were analyzed thematically using constant comparison.

**Results:** Fifteen participants were interviewed (12/15 women, on average 56 years old). All participants chose multiple categories of IR; 9/15 chose all IR. Participants' uncertainty related to their cancer (e.g. uncertainty about why they had cancer, their prognosis, family members' risks) was pervasive, "a grey cloud hanging over my head." Experiences with cancer uncertainty helped participants accept uncertainty related to IR (e.g. that results are "not 100%" or "nebulous"), despite generally disliking uncertainty. Participants viewed IR as valuable information, even though the information could be uncertain. Cancer experiences heightened the perceived value of learning IR, as IR may allow them to prepare for or prevent other diseases. Participants were fairly confident in their

ability to “deal with” IR-related uncertainty, in part because of their experiences managing cancer-related uncertainty. Participants anticipated managing IR-related uncertainty in a variety of ways, such as through information seeking, support from healthcare providers, or trying not to focus on uncertainty.

**Conclusions:** Cancer experiences fostered participants’ tolerance of uncertainty related to IR, and heightened the perceived value of IR. Genetic counseling for IR could involve exploration of patients’ attitudes toward uncertainty, and consideration of how illness experience may shape expectations for IR.

## PgmNr 646: The policy implications of the cancer genomics in Japan.

### Authors:

K. Muto <sup>1</sup>; Z. Yamagata <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Public Policy, Inst of Medical Sciences, Univ Tokyo, Tokyo, Japan; 2) Dpt of Health Sciences, School of Medicine, Univ Yamanashi, Yamanashi, Japan

---

The purpose of this paper is to review the impact of the current policy change on cancer genomics in Japan. In 2017, the Ministry of Health, Labour and Welfare (MHLW) convened a Roundtable Consortium on the Promotion of Cancer Genomic Medicine. In addition to providing Japanese citizens with early access to the world's leading-edge cancer genomic medicines under the national health insurance system, the consortium clarified ownership of the constructed framework as an asset of Japanese citizens and set directions for the functions and roles of medical institutions providing cancer genomic medicine, and institutions that aggregate, manage and promote the use of cancer genomic medicine (Report, June 27, 2017). In February 2018, the Japanese government designated 11 hospitals throughout Japan to serve as core hospitals for cancer genomic medicine. The government also appointed 156 facilities as liaison hospitals to work with the core hospitals. The MHLW approved the marketing of two cancer genome profiling systems in December 2018 - OncoGuide™ developed by the National Cancer Center of Japan and health instrument-maker Sysmex Corp., and FoundationOne® CDx sold by Chugai Pharmaceutical Co. The MHLW decided that these tests would be covered by the public health insurance system after June 2019. While the fee for these products would be ¥560,000 (USD 5,100), patients will have to pay 10 to 30 percent of the total fees. The MHLW will also ask the hospitals to submit the anonymized data of the genomic profiling to the Cancer Genome Information Repository, after obtaining patients' consent. The accumulated information is expected to be utilized for the development of new treatments. Insurance coverage will be applied to patients with solid tumor who have not responded to conventional therapy, which is expected that up to 26,000 people will use annually. The secondary findings of hereditary cancers will be disclosed to patients who gave consent to know. However, we have no law to regulate genetic discrimination. The Life Insurance Association has been considering the guidelines about notification when making an insurance contract. Furthermore, we have no roundtable to discuss employment issues. It is an urgent issue to respond toward concerns about genetic discrimination.

## PgmNr 647: Pre-emptive hereditary cancer genetic testing in primary care: Leveraging early program data for process improvement.

### Authors:

A.A. Lemke; H.M. Dunnenberger; J. Thompson; K. Dilzell-Yu; A. Newlin; K. Hughes; S. Purcell; E. Kaplan; A. Matsil; C. Johnson; P.J. Hulick

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Personalized Medicine, NorthShore Univ Health System, Evanston, IL

---

Primary care physicians (PCPs) are in a unique position to present pre-emptive genetic testing to patients, yet little is known about the service delivery and outcomes in this setting. To assess interest, system feasibility, and patient outcomes, as part of a learning health care system initiative, complimentary hereditary cancer genetic testing (30-gene next generation sequencing-based panel) was offered through 4 primary care sites during a patient's annual exam. Of the patients who read the test offer message in their healthcare portal prior to their appointment, 39% agreed to undergo testing through their PCP. A total of 1006 patients had testing in November and December of 2018. Systematic chart review and descriptive data analysis revealed that patients who had testing were primarily non-Hispanic (95.5%), Caucasian (64.5%), and female (67.2%), with ages ranging from 18-78 (median age 49). Ninety-two pathogenic variants were identified in 90 (9.1%) patients, in 16 genes: *MUTYH*, *CHEK2*, *APC*, *BRCA2*, *BRCA1*, *ATM*, *BARD1*, *MITF*, *BRIP1*, *NBN*, *PALB2*, *PMS2*, *RAD51C*, *MSH6*, *CDH1*, and *CDKN2A*. Of the 90 positive patients, 38 (42%) received 41 referrals to specialty clinics; 21 (51%) completed their referrals as of May 2019. Eighteen (20%) had 1 or more cancer screening tests recommended (32 tests), and 14 (78%) completed at least one screening test. One case of occult disease was detected on mammogram in a *BRCA1* positive patient. To elicit patient feedback, a focus group was conducted. This subset of patients reported satisfaction with the cancer genetic testing through their PCP, benefits from learning both positive and negative results, practice workflow issues, and a need for more pre- and -post test education. The study data has informed service delivery improvements in a subsequent larger pre-emptive genetic testing program in primary care: a coordinated care center has been initiated to assist in patient follow through with medical referrals and recommendations; additional multi-modal educational resources are provided to patients and PCPs at various points throughout the testing process; a PCP advisory group was developed to provide programmatic input; and mechanisms were created to provide continual clinical and operational feedback to providers. Our findings highlight the importance of early systematic appraisal and opportunity for changes throughout the pre-emptive genetic testing process to improve service delivery and patient care.



## PgmNr 648: What factors influence completing cancer genetic testing in a population screening study using a medical model? Preliminary data from the BRCA Founder Outreach (BFOR) study.

### Authors:

J. Hamilton <sup>1</sup>; K. Morgan <sup>1</sup>; J. Levin <sup>1</sup>; H. Symecko <sup>5</sup>; C. Gabriel <sup>2</sup>; J. Lester <sup>4</sup>; C. Jenkins <sup>2</sup>; D. Kamara <sup>4</sup>; K. Spielman <sup>5</sup>; M. Batson <sup>5</sup>; Y. Fesko <sup>6</sup>; J. Heitler <sup>7</sup>; J. Block <sup>1</sup>; L. Pace <sup>2</sup>; M. Robson <sup>1</sup>; K. Nathanson <sup>5</sup>; N. Tung <sup>3</sup>; S. Domcheck <sup>5</sup>; B. Karlan <sup>4</sup>; J. Garber <sup>2</sup>; K. Offit <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry & Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, New York.; 2) Dana-Farber Cancer Institute, Boston, MA; 3) Beth Israel Deaconess Medical Center, Boston, MA; 4) David Geffen School of Medicine at UCLA, Los Angeles, CA; 5) University of Pennsylvania, Philadelphia, PA; 6) Quest Diagnostics, Secaucus, NJ; 7) Lifelink, Brooklyn, NY

**Background:** In order to develop and evaluate a feasible approach to population-based genetic screening, the BRCA Founder Outreach (BFOR) study offers online education, study enrollment, and consent to *BRCA1/2* founder mutation testing to individuals of Ashkenazi Jewish descent in 4 US cities. Participants then visit a local lab facility to have blood drawn for no-cost *BRCA1/2* testing, with results review and management by their primary care providers or genetics specialists. To better understand testing barriers, we sought to determine what characteristics distinguish those individuals enrolling in the study who ultimately complete genetic testing from those who do not.

**Method:** During online study enrollment (Bforstudy.com), participants answered survey items about demographic (gender, age, location, parental status), medical (has a primary care provider, personal cancer history, family cancer history, known familial *BRCA1/2* mutation, personal genetic testing experience), and psychological (perceived cancer risk, cancer-specific distress, general anxiety, knowledge about *BRCA1/2* testing) characteristics. Data were collected on whether each participant did or did not obtain *BRCA1/2* testing by up to 6 months after enrollment. We conducted univariate analyses followed by multivariable logistic regression to evaluate which characteristics were associated with *BRCA1/2* test completion.

**Results:** From March 2018 through November 2018, 2363 participants enrolled in the BFOR study (79% female; 6.7% <30 years old, 40.6% 30-50 years, 52.6% >50 years; 12.2% personal cancer history). Of these, 2085 (88.2%) had blood drawn and completed *BRCA1/2* testing, while 278 (11.8%) did not obtain testing within 6 months of enrollment. In multivariable regression, factors associated with greater odds of test completion included having a personal cancer history (OR=2.73, 95%CI=1.57-4.74) and having children (OR=1.42, 95%CI=1.09-1.87). No other demographic, medical, or psychological factors were significantly associated with test completion.

**Conclusion:** A minority of those enrolling in a population screening study ultimately decide against *BRCA1/2* genetic testing. Limited demographic and medical variables are associated with completing genetic testing in this context. Novel strategies may be needed to better engage individuals without



the motivators of a personal cancer history or potentially at-risk children. Ongoing data collection and a second study phase will address these issues.

## PgmNr 649: Achieving equity in genomic services for hereditary cancer syndromes.

### Authors:

T. Kauffman<sup>1</sup>; A. Reyes<sup>1</sup>; B. Arnold<sup>1</sup>; S. Kraft<sup>2,3</sup>; D. Duenas<sup>2</sup>; K. Porter<sup>2</sup>; G. Jarvik<sup>8</sup>; L. Amendola<sup>8</sup>; K. Anderson<sup>7</sup>; G. Joseph<sup>4</sup>; C. Guerra<sup>5</sup>; L. Riddle<sup>6</sup>; K. Goddard<sup>1</sup>; on behalf of the CHARM team

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Health Research, Kaiser Permanente, Portland, Oregon.; 2) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Hospital and Research Institute, Seattle, WA; 3) Department of Pediatrics, University of Washington School of Medicine, Seattle, WA; 4) Department of Anthropology, History & Social Medicine, University of California, San Francisco, United States; Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco; 5) Department of General Internal Medicine, University of California, San Francisco; 6) Department of Bioethics, University of California, San Francisco; 7) Denver Health Ambulatory Care Services, Denver, CO; 8) Division of Medical Genetics, University of Washington

---

**Background:** Lynch syndrome (LS) and hereditary breast and ovarian cancer (HBOC) are the two most common hereditary cancer syndromes. While more patients are accessing genetic testing than ever before, rates of testing for and diagnosis of LS and HBOC remain low, and patients who are racial/ethnic minorities and/or low-SES (hereafter referred to as underserved) are even less likely to have access to this testing. The Cancer Health Assessments Reaching Many (CHARM) study, part of the Clinical Sequencing Evidence-generating Research (CSER) consortium funded by NHGRI, is studying methods to increase access to genetic testing for hereditary cancer syndromes in primary care for underserved patients.

**Methods:** English and Spanish speaking patients at Kaiser Permanente Northwest, an integrated delivery system in Portland, OR, and Denver Health, an integrated federally qualified health system in Denver, CO, are contacted via email, text message, postcard, and/or phone calls to complete an online cancer family history risk assessment. Those that screen at higher risk or have limited family history knowledge are offered clinical exome sequencing for 39 genes related to hereditary cancer risk and 92 genes related to medically actionable or carrier findings. Participants receive post-test genetic counseling.

**Results:** To date, we have invited ~6500 patients (56% underserved) to complete the risk assessment; 12% of underserved and 9% of non-underserved completed it. Among patients who completed the risk assessment, 82% underserved and 74% non-underserved were eligible. A similar percentage of underserved and non-underserved joined the study (80% vs. 79%). In total, 67% of 402 participants are underserved - 31% racial or ethnic minority, 40% low-SES, and 29% meet both of those criteria. Of the first 220 results reported, 16% had a positive cancer finding, 1% had a positive non-cancer medically actionable secondary finding and 16% had a carrier status finding.

**Conclusion:** At every step in the process, we have been successful at recruiting and enrolling underserved patients. The implementation of an online family history tool in the primary care setting

may be an effective and scalable approach to increase access to genetic testing for underserved populations.

## PgmNr 650: Development and early implementation of a hereditary cancer risk assessment program in a diverse patient population.

### Authors:

K.F. Mittendorf <sup>1</sup>; C. Ukaegbu <sup>2</sup>; T.L. Kauffman <sup>1</sup>; D.J. Eubanks <sup>1</sup>; E. Shuster <sup>1</sup>; J. Allen <sup>1</sup>; C. McMullen <sup>1</sup>; N.M. Lindberg <sup>1</sup>; K. Anderson <sup>3</sup>; B.S. Wilfond <sup>4</sup>; C.A. Bellcross <sup>5</sup>; S. Syngal <sup>2</sup>; K.A.B. Goddard <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Kaiser Permanente Northwest, Center for Health Research, Portland, OR.; 2) Dana-Farber Cancer Institute, Boston, MA; 3) Denver Health Ambulatory Care, Denver, CO; 4) Division of Bioethics and Palliative Care, Department of Pediatrics, University of Washington School of Medicine, Seattle, WA; 5) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA

---

**Introduction:** Lynch syndrome and hereditary breast and ovarian cancer are the most common hereditary cancer syndromes (HCS); each confers increased risk of multiple types of cancer. Selective screening to identify patients at risk for HCS prior to cancer onset requires in-depth family health history, which is often not collected, contains insufficient detail, or is not acted on by providers – these barriers disproportionately affect underserved populations. To address this care gap, we created a low-literacy, patient-facing HCS risk assessment tool.

**Methods:** We adapted two validated risk assessment algorithms - the Breast Cancer Genetics Referral Screening Tool (B-RST™ 3.0) and Prediction Model for gene Mutations (PREMM<sub>5</sub>), and developed a unique family knowledge screening module. Working iteratively with experts and patient stakeholders, we combined these three assessments into a low-literacy interactive electronic application. We recruited patients of two large health care systems to complete the tool; interactions with the tool were tracked and evaluated.

**Results:** We made two major modifications to the PREMM<sub>5</sub>: asking patients to complete family history, rather than to select an affected side, and reducing complex terms and computation (e.g. “*first degree* relatives”). B-RST™ 3.0 was adapted to present questions individually rather than in a table, to reduce graphicacy and numeracy demands. The entire tool includes pop-up literacy aids such as infographics and definitions.

A total of 688 individuals began the tool; 55% met study criteria for underserved populations (racial/ethnic minority and/or low SES). Only 6% of individuals (26% underserved) who began the tool did not complete it. Most (86%) incomplete interactions lasted under 5 minutes. Most people completed the tool without assistance (93%) spending a mean of 4.2 minutes (range: 0.4-53) doing so. Time spent was similar between underserved individuals (mean: 4.0 min, range: 0.4-24.1) and those who were not underserved (mean: 4.4 min; range: 0.4-53). Completion times for PREMM<sub>5</sub>, B-RST™, and the limited history assessment modules were 2.4 minutes (range: 0.3-49), 1.7 minutes (range: 0.4-39), and 1.5 minutes (range: 0.4-8.6), respectively.

**Conclusions:** Most patients quickly self-completed this risk assessment for HCS. Using a patient-facing

tool ensures standardized assessment and provides clear 'results' (high/low risk), which may facilitate appropriate referrals for genetics services.

## **PgmNr 651: Age-specific ovarian and breast cancer risks associated with inherited predicted pathogenic variants in *RAD51C* and *RAD51D*.**

### **Authors:**

Yang <sup>1</sup>; G. Leslie <sup>1</sup>; H. Song <sup>2</sup>; L. Castera <sup>3</sup>; G. Chenevix-Trench <sup>4</sup>; F. Couch <sup>5</sup>; M. Hoya <sup>6</sup>; O. Diez <sup>7</sup>; J.S. Dolinsky <sup>8</sup>; H. Ehrencrona <sup>9</sup>; W. Foulkes <sup>10</sup>; J. Garber <sup>11</sup>; S. Gayther <sup>12</sup>; L. Golmard <sup>13</sup>; S. Gruber <sup>14</sup>; E. Hahnen <sup>15</sup>; R. Janavicius <sup>16</sup>; A. Meindl <sup>17</sup>; H. Nevanlinna <sup>18</sup>; A. Osorio <sup>19</sup>; S.J. Ramus <sup>20</sup>; M. Robson <sup>21</sup>; C. Turnbull <sup>22</sup>; T.V.O. Hansen <sup>23,24</sup>; A. Vega <sup>25</sup>; P.P.D. Pharoah <sup>1,2</sup>; A.C. Antoniou <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Public Health and Primary Care, University of Cambridge, Cambridge, Cambridge, United Kingdom; 2) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 3) Laboratory of Cancer Biology and Genetics, Comprehensive Cancer Center François Baclesse, Caen, France; Inserm U1245, Rouen, France; 4) Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 5) Department of Laboratory Medicine and Pathology. Mayo Clinic. Rochester, MN: USA; 6) Molecular Oncology Laboratory CIBERONC, Hospital Clinico San Carlos, IdISSC (Instituto de Investigación Sanitaria del Hospital Clínico San Carlos), Madrid, Spain; 7) Area of Molecular and Clinical Genetics, University Hospital Vall d'Hebron. Hereditary Cancer Genetics Group. Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain; 8) Ambry Genetics, Aliso Viejo, CA, USA; 9) Division of Clinical Genetics, Department of Laboratory Medicine, Lund University and Department of Clinical Genetics and Pathology, Laboratory Medicine, Region Skåne, Lund, Sweden; 10) Departments of Human Genetics, Oncology and Medicine, McGill University, Montréal, QC, Canada; 11) Dana-Farber Cancer Institute, Boston, MA, USA; 12) Center for Bioinformatics and Functional Genomics and the Cedars Sinai Genomics Core, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 13) Department of Genetics, Institut Curie, Paris, France, 75005 and Paris Sciences Lettres Research University, Paris, France; 14) Departments of Medicine and Preventive Medicine, Keck School of Medicine, University of Southern California Norris Comprehensive Cancer Center, Los Angeles, CA, USA; 15) Center for Hereditary Breast and Ovarian Cancer, Center for Integrated, Oncology (CIO), University of Cologne, Faculty of Medicine and University Hospital Cologne, Cologne, Germany; 16) Vilnius University Hospital Santara Klinikos, Vilnius Lithuania; State Research Institute Innovative Medicine Center, Vilnius, Lithuania; 17) Department of Gynaecology and Obstetrics, University of Munich, Campus Großhadern, Munich, Germany; 18) Department of Obstetrics and Gynecology, Helsinki University Hospital and University of Helsinki, Finland; 19) Human Genetics Group, Human Cancer Genetics Programme, Spanish National Cancer Research Centre and Spanish Network on Rare Diseases (CIBERER) Madrid, Spain; 20) School of Women's and Children's Health, University of New South Wales Sydney, Australia; 21) Clinical Genetics Service, Department of Medicine. Memorial Sloan-Kettering Cancer Center. New York, NY: USA; 22) Division of Genetics and Epidemiology, The Institute of Cancer Research, London, UK; 23) Center for Genomic Medicine, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; 24) Department of Clinical Genetics, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark; 25) Fundación Pública Galega de Medicina Xenómica-SERGAS, Grupo de Medicina Xenómica-USC, CIBER de Enfermedades Raras, Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago de Compostela, Spain

---

Genetic testing through multi-gene cancer panels is widely available and has become an integral part of the genetic counselling and oncologic practice that is used to inform clinical management options. *RAD51C* and *RAD51D* are included on most cancer panels due the reported associations of predicted pathogenic variants (PPVs) in these genes with ovarian cancer. However, the available ovarian cancer relative risk (RR) estimates vary widely, are imprecise and the associations of *RAD51C/D* PPVs with breast cancer risk remain uncertain.

We used pooled data on 154 families with PPVs in *RAD51C* and 65 families with PPVs in *RAD51D*, ascertained through gene sequencing in 29 population- and family-based studies. Complex segregation analysis was used to estimate age-specific RRs and absolute risks of ovarian and breast cancer. We fitted genetic models to the cancer inheritance patterns in families, while adjusting for the mode of ascertainment for each family. The models allowed for the residual familial aggregation of both ovarian and breast cancer.

PPVs in both *RAD51C* and *RAD51D* were associated with ovarian cancer (*RAD51C* RR=8.29, 95% CI: 6.17-11.14,  $p=1.2\times 10^{-44}$ ; *RAD51D* RR=8.17, 95% CI: 6.00-11.12,  $p=1.2\times 10^{-40}$ ) and breast cancer (*RAD51C* RR=1.99, 95% CI: 1.39-2.85,  $p=1.9\times 10^{-4}$ ; *RAD51D* RR=1.93, 95% CI: 1.30-2.86,  $p=1.0\times 10^{-3}$ ). There was evidence that the ovarian cancer RRs were higher at ages 50 years or older for both *RAD51C* (RR=10.69, 95% CI: 7.55-15.14 for ages 50 or over versus RR=5.10, 95% CI: 2.83-9.20 for ages under 50) and *RAD51D* carriers (RR=11.52, 95% CI: 8.11-16.36 for ages 50 or over versus RR=3.45, 95% CI: 1.47-8.11 for ages under 50). There was a suggestion for higher breast cancer RRs at younger ages for both *RAD51C* and *RAD51D* PPV carriers, but these models did not fit significantly better than models with a constant RR. There was significant evidence that both ovarian and breast cancer risks for *RAD51C/D* PPV carriers vary by family history of the diseases. The estimated absolute risks of developing ovarian cancer by 80, based on the age-specific RR models, were 15% (95% CI: 11-21%) for *RAD51C* and 16% (95%CI: 12-22%) for *RAD51D* PPV carriers. The corresponding breast cancer risk was 21% (95% CI: 15-29%) for *RAD51C* and 20% (95%CI: 14-29%) for *RAD51D* PPV carriers.

These estimates will facilitate the genetic counselling of *RAD51C* and *RAD51D* PPV carriers and will inform the incorporation of *RAD51C* and *RAD51D* into cancer risk prediction models.

## PgmNr 652: Clinical implementation of Face2Gene facial photograph analysis in an adult, non-syndromic, cardiovascular genetic disease population.

### Authors:

E. Jordan; A. Morales; C. Hausfeld; R. Hershberger

[View Session](#) [Add to Schedule](#)

**Affiliation:** 0339 Biomedical Research Tower, 460 West 12th Avenue, Columbus, OH 43210, 0339 Biomedical Research Tower, 460 West 12th Avenue, Columbus, OH 43210, Columbus, Ohio.

---

Face2Gene (F2G) is a phenotyping application that uses facial analysis and a deep-learning network to identify patterns across hundreds of conditions to aid in diagnosis of genetic disease. From data extrapolated from facial photograph measurements and clinical features, F2G suggests differential genetic diagnoses. F2G has demonstrated most significant utility in syndromic, often pediatric, conditions with known or suspected dysmorphic gestalts. To our knowledge, F2G use in adult, primarily non-syndromic cardiovascular (CV) disease has not been published. The Ohio State University (OSU) CV genetics program has implemented F2G into clinical care in two outpatient clinics staffed by genetic counselors (GCs) and cardiologists. Of 331 patients seen over 17 months, 208 (63%) provided consent. Nine probands also had family members participate (N=13), therefore 94% of participants are unrelated. Photographs were obtained and securely stored in the F2G iPad app as a part of clinical workflow. Age, height, weight, ancestry and phenotypic features were collected by GCs and uploaded to F2G by a student assistant including: cardiomyopathy (CM) (40%), arrhythmia (13%), dyslipidemia (13%), aortopathies (8%), valve and vascular disease (5%), muscular and/or skeletal abnormalities (4%), other CV disease (12%), and other various features (5%). Of those with CM (N=153), 39% have hypertrophic CM, 33% have dilated CM, 9% have other CMs, and 19% have CM-related features. Due to limited sample size, statistical comparisons of gestalts for individual phenotypes were uninformative. CV genetic testing panels ordered often include some genes with syndromic associations. When variants are identified in syndromic genes, F2G is consulted. In addition, F2G suggested additional diagnoses in some cases that supported referral to additional genetics clinics. Based on 63% of patients opting to consent for the use of F2G, our experience supports the feasibility of F2G implementation in a largely adult, non-syndromic, CV genetic clinic setting. Further, associating genotype data with CV phenotypes and F2G photograph analysis has the potential to define previously undescribed gestalts, identify subclinical phenotypes, and serve as an interpretation tool in assessment of genetic variants. As the F2G system matures and additional CV phenotypes are collected within the F2G database, opportunities to conduct additional research and broaden the scope of F2G in new disease categories will present.



## PgmNr 653: *TNNI3* gene mutation as a cause of cardiomyopathy among Iranian patients.

### Authors:

S. Seyed Ali Akbar <sup>1,3</sup>; B. Rabbani <sup>1,2</sup>; S. Taghavi <sup>1</sup>; E. Siasi <sup>3</sup>; N. Mahdiah <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) genetic lab, shahid rajaei cardiovascular medical and research center, tehran, Select a Country; 2) Growth and development Research Center, Tehran University of Medical Sciences; 3) Islamic Azad University, North Tehran Branch, faculty of biological sciences

---

**Introduction:** Cardiomyopathy is common cause of cardiac death that is categorized to different types including HCM, DCM, LVNC and RCM. Over 20 genes detected to cause HCM; most of them are responsible to code sarcomeric proteins. More than 40 genes encoding cytoskeletal sarcomeric proteins, Z-disc have been described to cause familial DCM. The genetics of RCM is similar to HCM and sometimes this overlapping cause mixed HCM/DCM phenotype. Here, we study *TNNI3* mutations among Iranian patients with cardiomyopathy.

**Materials and Methods:** Total number of 30 patients affected with HCM, DCM and RCM were recruited in this study. Clinical data and EKG information were documented. All exons and introns of *TNNI3* gene were amplified and sequenced. Insilico analyses were applied for detection of novel variants.

**Results:** From the total number of 30, 13 patients were selected with dilated cardiomyopathy, 13 hypertrophic cardiomyopathies and 4 restrictive cardiomyopathies. After insilico analysis, two pathogenic variants were observed. Both were categorized as disease causing in Mutation taster (Free web-based application) and CADD score were calculated 38 and 12.60(CADD mutation version 1.4).

**Discussion:** *TNNI3* gene has been examined for the first time in a group of Iranian population in this study. Based on the Obtained information, *TNNI3* probably plays an important role in the incidence of cardiomyopathy. Few patients were recruited in this study; therefore, it is suggested that this study may be carry out on large populations.

## PgmNr 654: Deciding to pursue pediatric whole genome sequencing: Exploring the values of parents of children with cardiac disease.

### Authors:

K. Dunsmore <sup>1,2</sup>; L. Gotlib Conn <sup>3,4</sup>; M. Zubairi <sup>5</sup>; V. Venkataramanan <sup>1</sup>; E. Liston <sup>6,7</sup>; R. Kim <sup>6,7</sup>; R.Z. Hayeems <sup>1,2,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Child Health Evaluative Sciences, The Hospital for Sick Children, Toronto, ON; 2) Institute of Health Policy, Management and Evaluation, The University of Toronto, Toronto, ON; 3) Department of Anthropology, The University of Toronto, Toronto, ON; 4) Evaluative Clinical Sciences, Sunnybrook Research Institute, Toronto, ON; 5) Faculty of Health Sciences, McMaster University, Hamilton, ON; 6) Ted Rogers Centre for Heart Research, The Hospital for Sick Children, Toronto, ON; 7) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON

---

**Objective:** Health technology assessment calls for careful consideration of patient and societal values. In the context of genetic testing, existing research has demonstrated that personal and societal values influence how individuals and families orient and respond to tests on offer. Data that reflect on these values in the context of whole genome sequencing (WGS) and secondary variants (SVs) are only beginning to emerge. The objective of this study is to understand how personal, familial, and societal factors influence parents' decision to pursue WGS for their child.

**Methods:** Embedded in the Ted Rogers' Cardiac Genome Clinic (CGC) at the Hospital for Sick Children (Toronto, ON), parents of children with congenital heart disease who are offered WGS were eligible to participate. Prior to the receipt of WGS results, telephone-based semi-structured interviews explored personal, familial, and societal factors that influenced parents' decisions regarding WGS and the receipt of secondary variants. Interviews are audio-recorded and transcribed verbatim. Guided by the PEN-3 model that orients to the concept of culture in a multi-dimensional way, deductive and inductive coding identified core themes.

**Results:** We conducted 20 interviews with parents who were predominately healthy, English-speaking mothers of children with congenital heart defects or cardiomyopathies. During decision-making, parents value control and see primary and secondary variants as means to achieve control over their family's health and psychosocial aspects of their child's cardiac status. Parents' interest in WGS variants appears to be driven by their sense of responsibility to their child and their family. They express a moral obligation to seek out this information, no matter the emotional burden. For some, decisions to receive medically actionable secondary variants were value-laden and deliberated, while for others, decisions appeared to be value-neutral and somewhat automatic.

**Conclusion:** Findings from this study shed light on the acceptability of the technology from parents' perspectives and may inform education and counseling strategies.

## PgmNr 655: Genetic testing for stroke: Implications for sustainable interventions towards stroke prevention in West Africa.

### Authors:

E.O. Uvere <sup>1</sup>; R.O. Akinyemi <sup>2,3,4</sup>; J. Akinyemi <sup>5</sup>; A. Singh <sup>6</sup>; O.S. Arulogun <sup>7</sup>; C. Jenkins <sup>8</sup>; O.M. Owolabi <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, College of Medicine, University of Ibadan, Ibadan, Oyo State, Nigeria.; 2) Institute for Advanced Medical Research and Training, College of Medicine, University of Ibadan, Nigeria.; 3) Federal Medical Centre, Abeokuta, Nigeria.; 4) Centre for Genomic and Precision Medicine, College of Medicine, University of Ibadan, Nigeria.; 5) Department of Epidemiology and Statistics, College of Medicine, University of Ibadan, Nigeria.; 6) Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.; 7) Department of Health Promotion and Education, College of Medicine, University of Ibadan, Nigeria.; 8) Department of Nursing, Medical University of South Carolina, USA.

---

**Background:** It is crucial to assess stroke literacy as well as practices related to genetic testing for stroke among Africans in preparation for the ethical, legal and societal implications of the genetic revolution which has begun in Africa.

**Objective:** To assess the knowledge and practices of West Africans about stroke and genetic testing.

**Methods:** A comparative cross-sectional study was conducted among stroke patients and stroke-free controls recruited across 15 sites in Ghana and Nigeria under the SIREN study. Participants' knowledge of heritability of stroke, willingness to undergo genetic testing and methods of disclosing genetic test results were assessed using interviewer-administered questionnaire. Descriptive, frequency distribution and multiple regression analyses were performed.

**Results:** Only 49% of 2,029 stroke patients and 57% of 2,603 stroke-free individuals knew that stroke was a heritable disorder. Among those who knew, 90% were willing to undergo genetic testing. Knowledge of stroke heritability was associated with having at least post-secondary education (OR 1.51, 1.25-1.81) and a family history of stroke (OR 1.20, 1.03-1.39) while Islamic religion (OR=0.82, CI: 0.72-0.94), being currently unmarried (OR = 0.81, CI: 0.70-0.92), and alcohol use (OR = 0.78, CI: 0.67-0.91) were associated with lower odds of awareness of stroke as a heritable disorder. Willingness to undergo genetic testing for stroke was greater among respondents aged >45 years (78%) and married (73%) and was associated with having a family history of stroke (OR 1.34, 1.03-1.74) but inversely associated with a medical history of high blood pressure (OR = 0.79, 0.65-0.96). Only (40%, 17% & 14%) of stroke patients and (60%, 12% & 2.4%) of stroke-free individuals prefer to have their genetic test results disclosed to them in person, through telephone call and in the presence or through a family member respectively. Disclosure of genetic test results was significant (Fishers test<0.001) with age, educational and marital status of respondents.

**Conclusion:** Findings from the study shows that knowledge of stroke as a heritable disorder is sub-optimal. To further improve stroke knowledge, willingness to embrace genetic testing for and

prevention of stroke, there is need to target not just individuals with less formal education, history of high blood pressure, no family history of stroke but the wider population with contextually-driven, culturally-sensitive interventions.

## PgmNr 656: Newborn genomic screening for hypertrophic cardiomyopathy: Penetrance estimates from the PreEMPT model.

### Authors:

K.D. Christensen <sup>1,2</sup>; M. Gooch <sup>3</sup>; P.M. McMahon <sup>3</sup>; N.K. Stout <sup>2,3</sup>; J.M. Yeh <sup>2,4</sup>; A.C. Wu <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Division of Genetics, Brigham & Women's Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Department of Population Medicine, Harvard Pilgrim Health Care Institute, Boston, MA; 4) Division of General Pediatrics, Boston Children's Hospital, Boston, MA

---

### Purpose

Genomic screening can identify genetic risk factors in asymptomatic individuals, but few studies have projected whether the health benefits justify the financial costs and harms. We developed a simulation model to project the impact of genomic screening for conditions such as hypertrophic cardiomyopathy (HCM), where identification of disease-causing variants may inform screening and lifestyle recommendations and reduce sudden cardiac deaths. Here, we summarize preliminary model estimates of the number of newborns that would be identified with HCM variants and the associated penetrance through age 20.

### Methods

The Precision Medicine Policy and Treatment (PreEMPT) Model is a simulation model of population-based newborn genomic screening. The model integrates data from various sources to inform model inputs, including disease incidence, pathologic variant frequencies, and the proportion of affected individuals with pathogenic variants. Outputs include cumulative disease incidence and penetrance. We targeted genes associated with HCM based on gene-disease associations classified as definitive by a ClinGen Gene Curation Expert Panel. Variants were included if they were classified as likely pathogenic or pathogenic in ClinVar with a 2-star rating.

### Results

In a simulated cohort of 4 million newborns representative of an annual US birth cohort, we estimated that screening would identify 6,787 (95%UI: 6,197-7,296) newborns with genetic risk factors for HCM, and that 93 (95%UI: 56-164) of them would develop HCM by age 20 (estimated penetrance=1.4%, 95%UI: 0.9%-2.2%). Ongoing work incorporates benefits and risks that include early detection of disease, use of implantable cardioverter defibrillators, and reductions in sudden cardiac death. Sensitivity analyses will consider the impact of different reporting criteria and the likelihood that future applications may incorporate a greater number of genes and variants.

### Conclusion

The PreEMPT Model provides an innovative example of a model that simulates the impact of newborn genomic sequencing. Challenges in developing the model included deciding which genes and variants to target, estimating penetrance, and deciding what benefits and harms to include. Future results from the PreEMPT Model will provide insight about the cost-effectiveness of newborn genomic

screening not only for HCM alone, but in combination with newborn genomic screening for pediatric cancers, long QT syndrome, and other childhood-onset conditions.

## PgmNr 657: A 5-year retrospective study of individuals with phenylketonuria (PKU) treated at two specialized U.S. clinics.

### Authors:

J. Zhao <sup>1</sup>; D. Lamppu <sup>1</sup>; D. Kinch <sup>1</sup>; V. Anastasoie <sup>2</sup>; J. Baker <sup>3</sup>; K. DiBona <sup>2</sup>; J. Lindenberger <sup>3</sup>; M. McIllduff <sup>4</sup>; S. Watling <sup>4</sup>; A. Seymour <sup>1</sup>; H. Levy <sup>2</sup>; J. Vockley <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Homology Medicines, Inc., Bedford, MA; 2) Boston Children's Hospital, Boston MA; 3) University of Pittsburgh School of Medicine and UPMC Children's Hospital of Pittsburgh, Pittsburgh PA; 4) Boston Biomedical Associates, Marlborough MA

---

HMI-100-001 is a retrospective chart review of patients with a diagnosis of PKU due to Phenylalanine hydroxylase (PAH) deficiency using records spanning a 5-year period ending in November 2017 to capture real-world data associated with managing PKU under current standard of care.

PAH deficiency is an inborn error of metabolism due to mutations in the *PAH* gene, resulting in phenylketonuria (PKU). The mutations lead to absent or deficient *PAH* enzyme which converts Phe to tyrosine (Tyr), a precursor to multiple biologically important downstream neurotransmitters. Phenotypes range from mild HPA (Phe 120-360  $\mu\text{mol/L}$ ) to classical PKU (Phe over 1200  $\mu\text{mol/L}$ ). With the advent of newborn screening in 1963, managing the disease by low Phe diet in infants before clinical symptoms appear became the standard of care. Untreated PKU in children results in progressive irreversible neurological impairment. Current treatments do not address the underlying genetic defect.

The study was conducted at two U.S. clinics, a total of 152 patients (10-40 years old) were enrolled (65.8% with classical PKU). Data were collected from electronic medical records from baseline to -5 years (+/- 3 months) and characterized demographics, medical history, treatments and blood Phe. The number of patients with consecutive lab values decreased as the Phe threshold was lowered (Phe below 600, 360, 120 and 30  $\mu\text{mol/L}$ ). The data demonstrated decreased Phe control with age; mean Phe  $\pm$  standard error for patients 10-18 and over 18-40 was  $456.8 \pm 27.0$  and  $694.7 \pm 36.7$   $\mu\text{mol/L}$  respectively. 62.5% of patients were reported as having a history of at least one neuropsychiatric comorbidity, and adults were more likely than adolescents (69.5% vs. 54.3%). 90 of 98 *PAH* genotypes collected were distinct mutations; the 6 null-null genotypes were associated with classical PKU. Despite use of protein restriction, Phe concentrations over 360  $\mu\text{mol/L}$  were observed, particularly in classical PKU patients. Overall the demographics and clinical data were consistent across both sites.

These real-world data show that Phe levels were elevated, even when patients were on a Phe-restricted diet, and mostly above 360  $\mu\text{mol/L}$  (considered well-controlled based on current U.S. treatment guidelines). There remains an unmet need for therapies to control Phe concentrations without a Phe-restricted diet. As an autosomal recessive, monogenic defect, PKU due to PAH deficiency is a suitable condition for potential AAV-based gene therapy.

## PgmNr 658: A familial investigation of recurrent Osteogenesis Imperfecta (OI) over three decades.

### Authors:

C. Bascunana <sup>1,2</sup>; D. Rosenblatt <sup>1</sup>; F. Rauch <sup>1</sup>; P. Moffatt <sup>1</sup>; F. H. Glorieux <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McGill Department of Human Genetics, Montréal, QC, Canada; 2) Ovo Clinic, Montréal, QC, Canada

---

OI is a spectrum of bone genetic condition with variable severity. Before the advances of genomic, diagnostic was made via imaging and clinical examination. Then, biochemical analyses allowed further characterisation of physiopathology using protein electrophoresis. Today, OI is known to be genetically heterogeneous and is diagnosed by molecular sequencing. Although its inheritance is mainly autosomal dominant *de novo*, some autosomal recessive inheritance has been described, and recurrent diagnoses of autosomal dominant OI has been reported as well *via* the mechanism of germline mosaicism.

We present a case of recurrence of a lethal form of OI in siblings where 3 successive pregnancies were affected between 1977 and 1982. The first affected pregnancy was misdiagnosed as phocomelia, and low recurrence risk was given. The second pregnancy was properly diagnosed as affected with OI thanks to radiography imaging. The third affected pregnancy displayed abnormal skeletal ultrasonographic findings and the couple took the decision to have termination of the pregnancy. Autosomal recessive inheritance was mentioned to the couple, with a 25% recurrence risk. In 2017, a non-affected daughter of this couple is seen in Prenatal Genetics for her first pregnancy for which she would like be reassured regarding the risk of OI. In the medical chart of her mother, we discovered a result of biochemical analysis from 1986 using parental fibroblast and a sample from amniocentesis performed on the last affected pregnancy, leaning toward an autosomal dominant form of the condition from the electrophoresis pattern obtained. Surprisingly, a sample of DNA from the last affected pregnancy was stored since 1982. We wanted to be able to identify the causal mutation of recurrence of OI in this family, in order to offer proper genetic counselling to our patient. The foetal sample from 1982, sent for a molecular panel of genes associated to OI, revealed the heterozygous pathogenic novel mutation **COL1A1 (c.2336G>A)**.

This result confirmed the diagnostic of OI in its dominant form, and validated the explanation of germline mosaicism as the more likely to cause the familial recurrence. Thanks to the rise of genomic technologies, and the storage of biological material for more than 30 years, we resolved the uncertainty of a mother with a delicate reproductive past while reassuring her daughter, whom the recurrence risk of OI for all her pregnancies would be similar to general population.



## PgmNr 659: Role of genomic literacy in reducing the burden of common genetic diseases in Africa.

**Authors:**

G. Mboowa; I. Sserwadda

[View Session](#) [Add to Schedule](#)

**Affiliation:** Immunology and Molecular Biology, Makerere University, Kampala, Uganda

---

**Background:** In Africa, health practitioners and the public's current knowledge of genetics and genomics is still very low and yet this has potential to reduce the burden of common genetic diseases. Many initiatives have promoted genomic research, infrastructure, and capacity building in Africa. Now building on the critical mass of genomics achievements and collaborative projects already furnished on the continent through the H3Africa Initiative, African scientists are poised to move these initiatives to the next level while ensuring long-term sustainability by lobbying for funds from their governments and beyond. What remains to be done is to improve genomics literacy among populations and communities while utilizing an array of strategies. Genomic literacy and awareness are key in the formulating genomics related policies, genomics research and more importantly in the management of genetic diseases which includes diagnosis, prevention of complications and therapy. Africa is characterized by great cultural and language diversity thereby requiring a multidisciplinary approach to improving public and community genomic literacy and engagement. However, this is further complicated by having the fact that sub-Saharan Africa is comprised of countries with the lowest literacy rates in the world.

**Methods:** We applied the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to review genomic literacy in Africa using PubMed database.

**Results:** We found very limited evidence of genomic literacy for genetic diseases in Africa.

**Conclusion:** We propose a number of approaches that if embraced will significantly increase the genomic literacy and reduce the burden of genetic diseases in Africa namely; genomic counselling, revision of medical curricula, introduce genetics & genomics new graduate programs, genomics workshops & fellowships for healthcare workers, and community outreach programs delivered through a wide range of platforms.

## **PgmNr 660: Genetic analysis of children and adolescents living in Korean orphanage with behavior problems or developmental delay using chromosomal microarray and clinical exome sequencing.**

### **Authors:**

J. Kim; J. Cho; H. Kim

[View Session](#)
[Add to Schedule](#)

**Affiliation:** Department of Genetic Counseling Clinic, Konyang University Hospital, Daejeon, Korea

### **Purpose**

The aim of this study is to investigate genetic cause of developmental delay or behavior problems of residents of orphanage.

### **Methods**

Total 12 patients were referred to genetic counseling department of Konyang University Hospital from Korean orphanage facilities due to unexplained developmental delay or behavior problem (autism spectrum disorder (ASD) or violence) at between July 2017 and May 2019.

### **Results**

Mean age was  $13.8 \pm 5.7$  yrs (3-24 aged). Six patients were males. Six patients were females. All patients underwent chromosomal microarray (CMA). Seven patients had no abnormality in CMA. Seven variants from five patients were detected. Five variants were classified as benign CNV. Two pathogenic variants from two patients were 19p13.3 duplication and 16p14.11 duplication. Five of twelve patients were undertaken clinical exome sequencing (CES). Five variants from four patients were detected. One patient has no abnormality. Two variants were likely pathogenic variants. Three variants were variants of unknown significance (VUS). We could not co-segregation analysis because parents were not available for testing. Considering patients phenotype, we concluded that following gene variants were possible cause of their problems. Gene variant that founded are as follows: NPHP4, CD96 gene mutation (Nephronophthisis type 4, Opitz-Trigonocephaly syndrome), DNMT1 gene mutation (Autosomal dominant cerebellar ataxia, deafness, and narcolepsy), SRCAP gene mutation (Floating-harbor syndrome), RANBP gene mutation (Infection-induced encephalopathy 3).

Six patients among twelve patients were non-diagnostic.

### **Conclusion**

Total twelve patients with developmental delay and behavior problem including ASD were referred from the orphanage and was investigated for further diagnosis with CMA and CES. Diagnostic yield rate was 50% (six of total twelve patients). Genetic counseling was provided to caregivers for the understanding of biologic cause of their abnormal behaviors. Accurate genetic diagnosis enabled us to provide genetic counseling to caregivers of orphanage for better understanding of their problems and better care and management of them.

## PgmNr 661: Wolf-Hirschhorn syndrome: Natural history into adulthood, a preliminary study of 27 individuals.

### Authors:

J.C.. Carey <sup>1</sup>; A. Lortz <sup>2</sup>; A. Battaglia <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Pediatrics/Div Med Genetics, Univ Utah Hlth Sci Ctr, Salt Lake City, Utah.; 2) 4p- Support Group, Columbus, OH; 3) Dept of Developmental Neuroscience, Stella Maris Foundation, Pisa, Italy

---

Information on long-term outcome in Wolf-Hirschhorn syndrome (WHS) is very limited. Only five adult WHS are described in the medical literature, with age range between 24 and 37 years. We recruited 30 WHS individuals (22 females; 8 males), aged between 27ys and 54ys. 18 were personally observed at the Stella Maris Clinical Research Institute by A.B., and followed up between 5 and 20 years; and 12 were recruited from the 4p- Support Group-USA (9/12 had enough medical information to be included in the study); thus we are reporting on these 27. All are described as being happy, friendly, outgoing adults, liking to be around family and friends.

17/27 are close to requiring total care with assisted needs in feeding, diapering, bathing, dressing/undressing, and when walking needing assistance on uneven ground or transitioning from carpet to tile. 9/27 are partly self-independent, being able to feed and dress themselves, but needing supervision and some assistance with personal hygiene, and, at times, with walking on uneven ground. 1/27 is fully independent. All had variable degrees of developmental disabilities with a moderate-severe cognitive deficit. 18/27 enjoy good health, whereas 1/27 has diabetes type II, diagnosed at age 13; 1/27 has Raynaud disease, diagnosed at age 18; 1/27 had esophagitis at age 19; and 6/27 were diagnosed with hepatic neoplasms. All but one are seizures free. All are enrolled in a personalized rehabilitation program. Additionally we are collaborating with the CoRDS Registry at Sanford Research, South Dakota who have created a patient-reported database of information on individuals with rare diseases; the 4p- Support group has recruited many participants, including 18 persons > 18 years. Analysis of these persons with WHS is ongoing. Knowledge of the medical and developmental aspects of adults with WHS will inform health supervision.

## PgmNr 662: Biochemical genetics education: Legos for inborn errors of metabolism.

### Authors:

H.J. Lin <sup>1,2</sup>; P. Prasad <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, Harbor-UCLA Medical Center, Torrance, CA; 2) Department of Pediatrics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) NICHD/Perinatology Research Branch, Wayne State University School of Medicine, Detroit, MI

---

Earlier we showed how Lego bricks can be used to build atomically correct models of compounds relevant to medical biochemistry (Lin HJ, Lehoang J, Kwan I, Baghaee A, Prasad P, Ha-Chen SJ, Moss T, Woods JD. Lego bricks and the octet rule: molecular models for biochemical pathways with plastic, interlocking toy bricks. *Biochem Mol Biol Educ* 2018; 46:54-7. PMID 29105928). The models use “2 x 4” Lego bricks, in which the 8 studs represent the outer shell (or octet) of electrons around carbon, nitrogen, and oxygen atoms. Thus, Lego bricks model the chemical bonds in many small biological molecules, e.g., most amino acids, fatty acids, glucose (even the ring form), and intermediate metabolites. Joining 2 bricks with 2 studs indicates a single bond. Similarly, 4 studs model a double bond, and 6 studs form a triple bond.

Physicians in training are mostly unenthusiastic about reviewing metabolic pathways through names or chemical structures of compounds. But they quickly build Lego molecules. So we developed Lego teaching sessions to explain: fatty acid oxidation disorders; propionic and methylmalonic acidemias; and urea cycle defects. The sessions emphasize biochemical understanding to better grasp pathogenesis and treatment.

Our Lego sessions begin with fatty acid beta-oxidation to “acetate,” from which trainees build ketones -- acetoacetate, beta-hydroxybutyrate, and acetone. Later, they use Legos to see how propionate (e.g., from beta-oxidation of odd chain length fatty acids) normally converts to methylmalonate and then succinate. Legos show how propionate can disrupt the Krebs cycle and impede acetate catabolism -- causing ketone body accumulation and ketoacidosis. Next, trainees build the entire urea cycle with Legos -- including synthesis of *N*-acetylglutamate from “acetate” and glutamate. The models point out the causes of high ammonia in fatty acid oxidation disorders (less acetate available to form *N*-acetylglutamate) and in propionic acidemia (blocking of *N*-acetylglutamate synthase by propionate).

Trainees express satisfaction (even enjoyment) with reviewing biochemistry with Legos. They will easily go over the material with Legos almost daily, during a 2- or 4-week rotation in medical genetics. Our sessions have involved individual trainees, small groups (~10 people), and a group of nearly 30 people (with giant Legos for demonstration). Using Legos can be an engaging, chemically correct, simple, and fun way to teach on inborn errors of metabolism.

## PgmNr 663: Questionnaire survey for newborn screening including lysosome storage diseases.

### Authors:

M. Ozawa <sup>1</sup>; O. Migita <sup>1,2</sup>; Y. Abe <sup>2</sup>; N. Shimizu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, St. Marianna University, School of Medicine, Kawasaki, Kanagawa, Japan; 2)

2)Department of Clinical Genetics, St. Marianna University, School of Medicine, Kawasaki, Kanagawa, Japan

---

Among congenital diseases, there are some diseases which can prevent severe symptoms by early intervention. For those treatable disease, the longtime delay to diagnosis interferes early intervention, and it leads to treatment failure. Newborn screening is one of an effective and most important approaches to detect serious diseases on early or presymptomatic stage and lead to maximize efficacies of therapies.

Lysosomal storage diseases (LSDs) have promising therapeutic options, such as enzyme replacement therapies or pharmacological chaperones. But implementation of newborn screening (NBS) for LSDs is still debatable and some concerning points are existing. Some patients who have later-onset forms or attenuated forms of LSDs may not need to start treatment in their early life. It is still controversial issue, that NBS for LSDs creates unnecessary emotional and financial burden on the family. There are still few countries, which already provide nation-wide or region-wide neonatal screening program includes lysosomal storage diseases. Japan is one of the countries which are not including the lysosome diseases in nation-wide newborn screening. However, pilot screening studies have been started with expense of parents in some medical institutions. Before starting the new NBS tests including LSDs, families need to have knowledges of targeting diseases, and consensus of financial burden.

To address these issues, we have conducted questionnaires for assessing opinions of parents with their young children. Our aim of this questionnaire what type information and benefits expected for NBS before consent, what and when is the appropriate way to inform them, and how much financial burden is acceptable for families. We conducted a questionnaire survey on parents who visited our hospital for their newborn infant checkup. In our ongoing survey, 70% of the parents replied that they did not have any knowledge of LSDs, and 23% replied that they have been a little heard about LSD before. But after reading leaflet, 86% of parents replied that they would like to request the NBS for LSDs to their child, if they had known the LSD before NBS. They also commented that acceptable cost probably be around or below 10,000 Japanese yen. In order to avoid a delay in diagnosis, it is important to provide appropriate information of screening tests for LSDs.

## PgmNr 664: Evaluating the PROMIS scales as patient reported outcomes for acute intermittent porphyria.

### Authors:

H. Naik <sup>1</sup>; J. Overbey <sup>2</sup>; M. Balwani <sup>1</sup>; K. Anderson <sup>3</sup>; D. Bissell <sup>4</sup>; H. Bonkovsky <sup>6</sup>; J. Phillips <sup>7</sup>; B. Wang <sup>4</sup>; B. McGuire <sup>5</sup>; S. Keel <sup>8</sup>; C. Levy <sup>9</sup>; A. Erwin <sup>10</sup>; R. Desnick <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York.; 2) Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Preventive Medicine and Community Health, University of Texas Medical Branch, Galveston, TX; 4) Department of Medicine and Liver Center, University of California, San Francisco, CA; 5) University of Alabama, Department of Medicine, Birmingham, AL; 6) Department of Medicine (Section on Gastroenterology and Hepatology), Wake Forest NC Baptist Medical Center, Winston-Salem, NC; 7) Department of Internal Medicine, University of Utah, Salt Lake City, UT; 8) Division of Hematology, University of Washington, Seattle, WA; 9) Department of Medicine, University of Miami, FL; 10) Genomic Medicine Institute, Cleveland Clinic, OH

---

**Background:** Acute Intermittent Porphyria (AIP) is a rare inborn error of heme biosynthesis characterized clinically by life-threatening acute neurovisceral attacks. Few studies have assessed quality of life (QoL) in AIP patients and those that have had small sample sizes, and used QoL tools which may not have captured important domains. In this study, the NIH PROMIS scales were assessed in AIP patients to determine their sensitivity to disease severity, and assess their psychometric properties. These analyses measured whether the PROMIS scales are appropriate patient-reported outcomes (PROs) for this population and can be used as endpoints in future studies.

**Methods:** Baseline data from the US Porphyrias Consortium's Longitudinal Study were obtained for 259 AIP patients. These included detailed disease and medical history as well as PROMIS data. The following PROMIS scales were used: anxiety, depression, pain interference, fatigue, sleep disturbance, physical function, and satisfaction with social roles. The relationships between PROMIS domain scores and key clinical and biochemical features of AIP were assessed, and exploratory and confirmatory factor analyses (EFA and CFA) were conducted.

**Results:** PROMIS domain scores among AIP patients were significantly worse than the general population across all domains except depression. Each PROMIS domain discriminated well between asymptomatic and symptomatic patients with symptomatic patients demonstrating worse scores. Many important clinical variables were significantly associated with domain scores in univariable analyses, showing good response for the PROMIS scales, specifically pain interference and fatigue scales. Internal consistency was high for all the PROMIS scales. EFA revealed a five-factor model as the most appropriate, each factor consisting of a separate domain: pain interference, anxiety, depression, sleep disturbance, and fatigue. CFA of this five-factor model showed good fit.

**Conclusion:** Pain interference and fatigue were the most responsive scales in measuring QoL in this cohort. The five-factor model fit the data well, which is different from the model observed in the general population. These findings provide initial evidence that the PROMIS scales were valid and

reliable measures for clinical outcomes in AIP. Thus, the PROMIS scales provide an optimal assessment of PROs.

## PgmNr 665: Two decades follow-up of 123 patients with Williams-Beuren syndrome.

### Authors:

R.S. Honjo <sup>1</sup>; V.F. Monteleone <sup>1</sup>; E.A. Furusawa <sup>2</sup>; K.T. Faria <sup>3</sup>; A.M. Santos <sup>3</sup>; M.C.T.V. Teixeira <sup>3</sup>; D.R. Bertola <sup>1</sup>; C.A. Kim <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics Unit, Instituto da Criança, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, BR; 2) Nephrology Unit of Instituto da Criança, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, BR; 3) Universidade Presbiteriana Mackenzie, Sao Paulo, SP, BR

---

**Objective:** The objective of this study was to analyse the natural history of Williams-Beuren syndrome based on the clinical findings of 123 patients. **Methods:** comprehensive retrospective review of the medical records of patients with molecular diagnosis of Williams-Beuren syndrome followed in our genetics outpatient clinic in a period of 20 years. **Results:** All cases were sporadic except one case of monozygotic twins and one case of first-degree cousins. All of the patients had the typical facial characteristics of the syndrome. Most patients presented with cardiovascular abnormalities and 75.6% have some form of congenital heart disease (CHD), being supravalvar aortic stenosis the main one (found in 47.3% of patients with CHD). Among them, 21.5% needed heart surgery and/or cardiac catheterization, and one patient needed a heart transplant. Hypertension was present in 40% of patients and 19.2% of those had renal artery stenosis. Other vascular stenoses were also reported. Additional prevalent findings were vertebral spine abnormalities (71.5%), urinary tract abnormalities (73.2%) and hernias (45%). Additional findings included hypothyroidism, ocular findings, radio-ulnar synostosis, precocious puberty, constipation, feeding difficulties, hypercalcemia, and some rare atypical findings. Most presented developmental delays and some level of intellectual disability, however 3 patients had a borderline IQ and 2 had a low average IQ. Almost half (49.6%) of our patients are adults, with ages ranging from 18y to 38y (mean 18.1y; median 17y). None of the adults was reported as employed. **Conclusions:** Our cohort behaved predominantly according to published literature, however there were some atypical findings. To our knowledge this is one the largest Brazilian cohorts of Williams-Beuren syndrome patients, including adult patients, making it possible to describe the natural history of Williams syndrome. In the context of the implementation of a national policy for rare diseases it is increasingly important to better characterize this population in Brazil and to emphasize the importance of long-term follow-up with a multidisciplinary team.



## PgmNr 666: 22q11.2 deletion syndrome and coexisting conditions: An important prognostic, management, and genetic counseling consideration.

### Authors:

D.M. McDonald-McGinn <sup>1,2</sup>; M. Unolt <sup>3</sup>; T.B. Crowley <sup>1</sup>; D.E. McGinn <sup>1</sup>; J. Cohen <sup>1</sup>; A. Bailey <sup>1</sup>; M. Lambert <sup>1,2</sup>; S. McCormack <sup>1,2</sup>; S. Hopkins <sup>1,2</sup>; V. Vetter <sup>1,2</sup>; B.S. Emanuel <sup>1,2</sup>; E.H. Zackai <sup>1,2</sup>; T. Gambin <sup>4</sup>; M. Geremek <sup>4</sup>; B. Nowakowska <sup>4</sup>; J. Vermeesch <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, USA; 3) Ospedale Bambino Gesù and Sapienza University, Rome, Italy; 4) Institute of Mother and Child, Warsaw, Poland; 5) KU Leuven, Leuven, Belgium

---

**Background:** 22q11.2 Deletion Syndrome (22q11.2DS) is the most frequent copy number variant (CNV), affecting ~1/1000 fetuses and ~1/2000-4000 children, resulting in recognizable but variable findings across multiple organ systems. Patients with atypical features should prompt consideration of coexisting diagnoses due to the possibility of additional genome-wide mutations/CNVs which may be inherited, as well as, mutations/CNVs on the remaining intact chromosome 22q11.2 allele unmasking an autosomal recessive condition. Both occurrences compound symptoms impacting management and genetic counseling. **Methods:** Records on 1422 patients with a laboratory confirmed 22q11.2 deletion followed at the Children's Hospital of Philadelphia (CHOP) were reviewed to identify a dual diagnosis, including a subset of patients with atypical features, e.g. craniosynostosis, imperforate anus, limb differences, and sudden death, whose samples underwent next generation sequencing of the intact chromosome 22q11.2 allele as part of an ongoing IRB approved collaborative research protocol in Philadelphia, Leuven, and Warsaw. **Results:** 24 patients with 22q11.2DS were found to have a coexisting diagnosis representing 1.7% of our overall cohort. These included 8 patients with conditions unrelated to 22q11.2DS (Severe Combined Immunodeficiency, Trisomy 8 mosaicism, CHARGE syndrome due to a *CHD7* mutation, cystic fibrosis, a maternally inherited 17q12 deletion, G6PD deficiency, von Willebrand disease, and a 1q21.1 deletion) and 16 patients with a 22q11.2 deletion and a CNV/mutation on the remaining allele resulting in a previously undiagnosed autosomal recessive condition. The latter group included mutations or CNVs in *GP1BB*, *CDC45* (4), *LZTR1* (2), *SNAP29* (4), and *TANGO2* (5) which explained patient's atypical features in association with known conditions such as Bernard-Soulier, autosomal recessive Noonan syndrome, and CEDNIK syndrome, as well as, what we believe to be a new condition associated with a *CDC45* mutation, and *TANGO2* Related Disease unfortunately explaining the sudden death of a 5-year-old child. **Conclusions:** These findings support considering additional laboratory testing in all patients with 22q11.2DS, to ensure appropriate personalized care, as formulating medical management decisions hinges on establishing the correct diagnoses in their entirety, especially given that these findings are medically actionable, potentially altering long-term outcome and recurrence risk counseling.

## PgmNr 667: Clinical exome sequencing workflow for cerebellar ataxias: Increased diagnostic yield and identification of novel candidate.

### Authors:

P. Vishwakarma; S. Agarwal; 5 year

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India

---

### Clinical exome sequencing workflow for Cerebellar ataxias: increased diagnostic yield and identification of novel candidate

Priyanka Vishwakarma, Sarita Agarwal\*

Department of Medical Genetics, SGPGIMS, Lucknow

**Introduction:** Spinocerebellar ataxia (SCA) is a rare, genetically heterogeneous, adult onset, neurodegenerative disorder. It has 40 subtypes reported. Majority of SCA subtypes are caused by dominant CAG expansion mutations. Clinical features present by SCA patients are gait disturbance, movement problem, Nystagmus etc. Genetic studies of late onset sporadic ataxias are not routinely mentioned however, for unresolved cases NGS or clinical exome sequencing (CES) are available for a definitive diagnosis. **Materials and method:** In this study we recruited 50 SCA suspected patients for the routine genetic diagnosis of SCA (subtypes 1,2,3,6 and 7) by multiplex and TP-PCR method and we got 12 positive cases of SCA subtypes like 4 were positive for SCA1, 6 positive for SCA2 and 2 positive for SCA3 subtype. After this we found 38 patients who were uncharacterized for SCA in spite of this, they were presented with cerebellar ataxia features. Then we selected 10 patients from 38 uncharacterized cases of SCAs with autosomal dominant cerebellar ataxia for CES, because it was necessary to unravel the reason of problematic feature of patients. Additionally, we already excluded the presence of mutations in the genes causing FRDA by STR analysis in previous research projects. **Results:** we identified 3 cases of spastic paraplegia, all were male patients with a usual facial phenotype and they developed ataxia at early age of life with no significant family history and in initial clinical approach that ruled out common etiologies. Clinical exome sequencing was performed when the patients were 24, 24 and 21 years of age respectively. The results identified heterozygous pathogenic variant c.2014G-A (p. Gly672Arg), encode (SPG7 gene), and homozygous pathogenic variants c.869+1 G-T (5'Splice site), (SPG11 gene) and c.1A-G (p.met1?) (FA2H gene) in all 3 subjects respectively. Sanger sequencing in our laboratory confirmed all 3 variants. **Conclusion:** Our findings suggested that patients with progressive cerebellar ataxia would benefit diagnostically from CES irrespective of a positive family history or early age at onset. CES findings generally offer valuable information which can be used for clinical decision making and in defining new clinical & bioethical action like counseling.

## PgmNr 668: Systematic review and meta-analysis: Application of diagnostic sequencing in epilepsy, autism, and intellectual disability.

### Authors:

A. Stefanski <sup>1</sup>; Y. Calle-Lopez <sup>2</sup>; E. Pestana-Knight <sup>2</sup>; D. Lal <sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Medicine Institute, Lerner Research Institute Cleveland Clinic, Cleveland, OH 44195, USA; 2) Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, USA; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 4) Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, Germany

---

Understanding the underlying molecular pathology in neuropsychiatric disorders can affect patient management, predict recurrence risk, and enable clinical trial enrollment for novel precision medicine drugs. A comprehensive assessment of the field of clinical diagnostic sequencing across neuropsychiatric disorders is lacking. Here, we conducted a systematic literature analysis of clinical sequencing studies and found 8,396 articles. After additional filtering, we identified most studies for epilepsy (n=48), involving 15,277 patients. We also found 16 studies for intellectual disability, involving 2,317 patients, and 12 studies for autism, involving 1,578 patients. The diagnostic yield for autism was 0.26 (95% confidence interval [CI]: 0.19–0.33), for intellectual disability the yield was 0.25 (95% CI: 0.19–0.32), and for epilepsy, 0.23 (95% CI: 0.20–0.26). Uncorrected for pre-screening, the diagnostic yield for studies applying whole exome sequencing (WES) was 0.24 (95% confidence interval [CI]: 0.21–0.28), for gene panel sequencing the yield was 0.21 (95% CI: 0.18–0.23), and for whole genome sequencing (WGS), 0.19 (95% CI: 0.07–0.42). In the subset of epilepsy studies, the yield for combined generalized and focal epilepsies was 0.23 (95% confidence interval [CI]: 0.20–0.27), for generalized epilepsies the yield was 0.21 (95% CI: 0.15–0.29), and for focal epilepsies, 0.15 (95% CI: 0.08–0.25). We observed increasing numbers of clinical sequencing studies annually, particularly in research studies from Asia. In the combined list of 76 identified clinical sequencing studies, we did not find any publications from Africa, India, or Latin America. In contrast to earlier publications, more recent studies apply guidelines for variant interpretation and are more likely to report variants of uncertain significance. No study reported benign classified variants. In summary, clinical genetic testing represents a fast evolving, and increasingly useful approach in identifying the etiology of neuropsychiatric disorders. It is likely that the diagnostic yield will further increase with the broader application of more comprehensive testing (e.g., WES) and advanced variant interpretation methodologies. This systematic review provides a comprehensive overview of clinical sequencing studies of neuropsychiatric disorders, can guide policymaking, and help steer decision-making in patient management.

## PgmNr 669: Alexander disease: A rare case.

### Authors:

D. La Delfa

[View Session](#) [Add to Schedule](#)

**Affiliation:** American University of Antigua College of Medicine, Department of Internal Medicine, Kingsbrook Jewish Medical Center, Brooklyn, New York

---

Alexander disease is a rare autosomal dominant genetic disorder of the nervous system, classified as a leukodystrophy. The cause of Alexander disease is the destruction of myelin, a lipid layer which insulates the nerve fiber. This destruction decreases the rate of transmission of nerve impulses and leads to the overall functional impairment of the nervous system seen in clinical cases.

The true prevalence of Alexander disease is currently unknown, however, only approximately 500 cases have been reported since 1949. There are two subtypes of Alexander disease; Type I and Type II. Type I, infantile form, is typically diagnosed prior to age four while Type II, juvenile form, is diagnosed after age four. We present a case report of a 27-year-old African American female who was diagnosed with Infantile Alexander disease. She has required lifelong respiratory ventilation, nutritional supplementation via gastrostomy tube and prophylactic seizure medication. MRI of the brain showed persistent moderate leukoencephalopathy with abnormal patchy parenchymal enhancement involving the frontal lobe, pons, cerebellum, and basal ganglia bilaterally. This patient has doubled the estimated survival rate which makes this case even more rare.

Currently there is no treatment for patients with Alexander Disease. Supportive management should include nutritional therapy, antibiotic therapy for intercurrent infections, and antiepileptic therapy. It is crucial for affected patients to undergo preventative measures to avoid secondary complications. An MRI can assist in the confirmation of Alexander Disease. Specific findings can guide the diagnosis however physicians must keep in mind that other diseases can present similarly. Increased levels of  $\alpha\beta$ -crystallin, heat shock protein 27, and glial fibrillary acidic proteins were seen in all confirmed cases of Alexander Disease, making this the most specific test for establishing the diagnosis.

Although most cases present in early years of life, it is just as important for physicians practicing adult medicine to be familiar with the signs and symptoms of Alexander disease as it may mimic other neuromuscular diseases, such as Multiple Sclerosis.

*Key words: Alexander disease, leukodystrophy, glial fibrillary acidic protein, Multiple Sclerosis*

## PgmNr 670: A patient-centered perspective in genetic studies: Successful autism discoveries.

**Authors:**

Z. Talebizadeh; A. Shah

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pediatrics, Children's Mercy Hosp, Kansas City, Missouri.

---

Concerns over the need to improve translational aspects of genetics studies and engage community members in the research process have been noted in the literature and raised by patient advocates. To address this gap, we have launched an initiative called Autism Genetics Outcomes (AutGO). The overall goal was to establish partnership with a wide range of stakeholders and develop a new integrated concept called GO (i.e., research approaches that utilize data and principles of both genetics and outcomes research). To date, we have accomplished the following: identified inclusion criteria for GO hypotheses, developed a semi-structured literature review protocol to collect stakeholders' perspectives for developing such hypotheses, and identified 10 different topics related to autism that may lead to a GO hypothesis. We have also developed a GO hypothesis for a topic prioritized by our AutGO members (i.e., depression in autism). While there is a consensus among different stakeholders on the need to make a bridge between outcomes and genetics research, there is a paucity of a practical and meaningful framework to implement this much needed synergy.

By applying lessons learned from AutGO, we assessed autism genetic findings from a patient-centered perspective. To do so, we evaluated a few successful/replicated discoveries (related to PTEN, MET, and Fragile X genes) from the standpoint of both scientific impact and patient-centeredness. The following elements were distilled: (1) if/how patient and/or parent concerns contributed to the study inception/development process, and (2) if/how they contributed to developing treatments and improving patient health outcomes, and (3) highlight elements included in hypothesis generation, research design, and data evaluation that were key factors in making these discoveries successful.

Such a unique retrospective assessment of the genetic findings will facilitate developing a practical set of recommendations (conceptual framework) for the policy makers, funding agencies, and the research community to further promote designing research protocols that have a higher likelihood of contributing to translational research. Gathering and sharing such dual perspectives could facilitate igniting interest on the topic of making a bridge between outcomes and genetics research in autism, and promote building synergistic frameworks among relevant, but currently disconnected, initiatives that are aiming to increase patient-centeredness.

## **PgmNr 671: Developing genetic testing and genetic counselling in psychiatry in Europe: Perspectives after the first year of the EnGagE network.**

### **Authors:**

B. Chaumette <sup>1</sup>; A. McQuillin <sup>2</sup>; R. Moldovan <sup>3,4,5</sup>; M. Rietschel <sup>6</sup>; K. Tammimies <sup>7</sup>; D. Coviello <sup>8</sup>; M. Soto <sup>9</sup>; A. Cuthbert <sup>10</sup>; S. Schicktz <sup>11</sup>; J. Perry <sup>11</sup>; J. Pawlak <sup>12</sup>; D. Crepaz-Keay <sup>13</sup>; K. Koido <sup>14</sup>; L. Kapur-Pojkic <sup>15</sup>; F. Degenhardt <sup>16</sup>; Members of the EnGagE network

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute of Psychiatry and Neuroscience of Paris, INSERM U1266, Paris, Paris, France; 2) Division of Psychiatry – University College London - London - United Kingdom; 3) Department of Psychology - Babes-Bolyai University - Cluj-Napoca - Romania; 4) Division of Evolution and Genomic Sciences - School of Biological Science - University of Manchester - Manchester - United Kingdom; 5) Manchester Centre for Genomic Medicine - St Mary's Hospital - Manchester Academic Health Science Centre - Manchester University Hospitals NHS Foundation Trust - Manchester - United Kingdom; 6) Central Institute of Mental Health - Medical Faculty Mannheim - Heidelberg University – Mannheim - Germany; 7) Center of Neurodevelopmental Disorders, Department of Women's and Children's Health, Karolinska Institutet – Stockholm - Sweden; 8) Laboratory of Human Genetics - IRCCS Istituto Giannina Gaslini - Research Institute and Children Hospital – Genoa - Italy; 9) FEA Psiquiatria - Servicio Andaluz de Salud – Almería - Spain; 10) MRC Centre for Neuropsychiatric Genetics and Genomics - Division of Psychological Medicine and Clinical Neurosciences - Cardiff University – Cardiff - United Kingdom; 11) Department of Medical Ethics and History of Medicine - University Medical Center Goettingen – Goettingen - Germany; 12) Psychiatric Genetics Unit - Department of Psychiatry - Poznan University of Medical Sciences - Poznan - Poland; 13) Mental Health Foundation – London - United Kingdom; 14) Institute of Biomedicine and Translational Medicine - Centre of Excellence for Genomics and Translational Medicine - University of Tartu – Tartu - Estonia; 15) University of Sarajevo - Institute for Genetic Engineering and Biotechnology – Sarajevo - Bosnia and Herzegovina; 16) Universitätsklinikum Bonn - Institute of Human Genetics – Bonn - Germany

---

Recent major advances in psychiatric genetics have been driven by both major technological advances and the establishment of international consortia. Several results can already be translated from bench to the bedside. For example, according to the statement of the International Society of Psychiatric Genetics (<https://ispg.net/genetic-testing-statement/>), psychiatric genetic testing (psyGT) and identification of pathogenic copy number variations may help psychiatric patients to better understand and accept their diagnoses and could improve the psychiatric genetic counselling (psyGC) for familial recurrence risk. PsyGC is also efficient for alleviating emotional distress and guilt and in improving self-efficacy and empowerment of patients and their family. However, integrating this evidence into clinical practice requires a substantial research effort to overcome major clinical, ethical and practical challenges.

In 2018, the European Union funded a COST Action to build a research network called EnGagE: “Enhancing Psychiatric Genetic Counselling, Testing, and Training in Europe” (<https://www.cost.eu/actions/CA17130/>). EnGagE is a knowledge-sharing and expertise-enhancing

network. The objectives are to develop a framework to facilitate the implementation of PsyGC and PsyGT into routine clinical care. EnGagE aims to develop standardized practice recommendations and research protocols, share scientific knowledge and data, and provide standardized training in PsyGC and PsyGT. The EnGagE network now includes preclinical and clinical researchers from the fields of neuroscience, psychiatric genetics, psychosocial research, and ethics; clinicians from the fields of psychiatry, psychology, and medical genetics; genetic counsellors, and scientists from diagnostic genetic testing laboratories. Furthermore, patient and family advocacy groups are actively involved in EnGagE. Currently, the EnGagE network comprises 30 European countries and has recruited active participants in Australia, Canada, Russia, and the USA.

We will present the structure of the EnGagE network (especially its different taskforces), the achievements that have been made in its first year, as well as the timeline of future commitments. The specificities and the challenges of PsyGC and PsyGT will be discussed, both from a practical and ethical point-of-view. We will discuss the difficulties encountered in transferring scientific knowledge of psychiatric genetics into routine clinical practice in Europe.



## PgmNr 672: The experience of parenting a son with X-linked retinoschisis.

### Authors:

R. Nolen <sup>1</sup>; C. D'Amanda <sup>1,5</sup>; B. Biesecker <sup>2</sup>; C. Cukras <sup>3</sup>; P. Sieving <sup>4</sup>; A. Turriff <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ophthalmic Genetics & Visual Function Branch, National Eye Institute, Bethesda, MD; 2) RTI International, Washington, DC; 3) Division of Epidemiology & Clinical Applications, National Eye Institute, Bethesda, MD; 4) Director, National Eye Institute, Bethesda, MD; 5) Department of Medicine, Division of Genetics, University of Rochester Medical Center, Rochester, NY

---

**Introduction:** Parenting a child with a genetic condition has been associated with significant stress, grief and loss, and concern for the future. Prior studies of carriers of X-linked conditions have found that guilt is prominent. X-linked retinoschisis (XLRS), a childhood-onset retinal dystrophy, differs from previously studied conditions in terms of disease prognosis, caregiver burden, and lack of established carrier phenotype. This study explored the experiences and characteristics of parents of sons with XLRS.

**Methods:** Parents of sons with XLRS who were evaluated at the National Eye Institute (clinicaltrials.gov: NCT00055029) between December 2017 and January 2019 were eligible for this cross-sectional, mixed methods study. Semi-structured interviews were conducted by two researchers either in person or by phone. Interviews were transcribed, coded, and analyzed thematically. Participants also completed a survey composed of demographic questions and validated scales to assess dispositional optimism, anxiety, and personality traits.

**Results:** 11 mothers and 8 fathers participated in this study. Recruitment occurred until no new concepts emerged from the interviews. Reported optimism, anxiety, and personality traits were consistent with published normative data. Parents described a process of continuous adaptation. The initial diagnosis was characterized by grief, shock and devastation for most parents, which was exacerbated by reportedly unfavorable interactions with providers who lacked familiarity with XLRS. Maternal guilt was common, but usually improved with time. As parents adjusted to life post-diagnosis, they attempted to maintain a sense of normalcy while balancing the desire to protect their sons. Significant stressors included decisions around sports participation, driving, and school accommodations. Among all parents, the fear of retinal detachment was an ongoing source of concern. Most parents identified perceived benefits from their experiences, such as increased compassion or family cohesion.

**Conclusions:** Most parents viewed XLRS as a significant challenge in their sons' lives, but one which could be overcome with the right guidance. Clinical encounters may be enhanced for families with XLRS by providing quality information, preparing families for potential challenges, and by assessing and acknowledging the psychosocial impact of living with XLRS.



## PgmNr 673: Evaluation of depression in patients with inherited forms of vision loss: Results from a clinical practice improvement pilot project using the PHQ-9 questionnaire.

### Authors:

D. Blain <sup>1</sup>; A. Turriff <sup>1</sup>; D. Claus <sup>1</sup>; A. Bamji <sup>1</sup>; C. D'Amanda <sup>2</sup>; B. Brooks <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) OGVFB, Bldg 10, Rm 10N109, National Eye Institute, Bethesda, MD.; 2) Division of Genetics, Department of Medicine University of Rochester Medical Center 601 Elmwood Avenue, Box 641 Rochester, NY.

---

**Purpose:** Studies have documented the negative psychosocial impact of vision loss related to inherited eye conditions. In this context, a pilot practice improvement project was set up to evaluate whether a rapid assessment for depression during a clinical visit may facilitate identifying and appropriately referring at-risk patients. This project also allowed for a view into the depressive profiles of patients seen in clinic.

**Methods:** Nurse coordinators administered the Patient Health Questionnaire depression scale (PHQ-9) to 50 consecutive adult patients at the Ophthalmic Genetics and Visual Function Branch (OGVFB) clinic of the National Eye Institute (NEI) between January and March 2019. Patients scoring  $\geq 10$  on the scale were assessed by a genetic counselor that day, who followed up three months later by phone. Patient demographics and clinical findings were obtained from the NEI electronic medical record and correlated with the PHQ-9 scores for further characterization of the cohort.

**Results:** Data from 47 patients were included in the analysis. Median onset of visual symptoms was 24 years (range: birth to 60 years). On average, patients had lived 20 years with their symptoms. A majority (57%) did not have a family history of the condition. Most (96%) had a progressive or possibly worsening form of visual loss. Median Best Corrected Visual Acuity (BCVA) was 20/40 (range: 20/12.5 to Light Perception). PHQ-9 scores ranged from 0 to 17, with 41 patients (87%) falling into the Minimal or Mild depressive symptoms category (score  $\leq 4$ ). The average score was 3.6 (median: 2). No correlation was detected between scores and other variables such as BCVA, residual visual fields, age of onset, age at visit, length of diagnosis, sex and family history. Six participants (13%) had scores in the Moderate or Moderate to Severe category and were assessed by a genetic counselor. All had a history of depression and either had been or were under medical care for the diagnosis. Follow-up is ongoing.

**Conclusions:** The existing clinic infrastructure and brevity of the PHQ-9 questionnaire allowed for a rapid evaluation of patients for depressive symptoms. In this cohort, 13% were found to be at-risk, which is consistent with rates published for patients with chronic medical conditions, but lower than those reported in the literature for vision loss. The fact that all at-risk patients had a history of depression suggests that long-term follow-up may be needed for these patients.

## PgmNr 674: Greater than the sum: Thought experiment fusing HUGO Gene Nomenclature Committee assigned gene numbers into International Classification of Diseases for the coding of genetic variants.

### Authors:

H. Schlecht

[View Session](#)
[Add to Schedule](#)

**Affiliation:** Medicine, Baystate Health, Springfield, Massachusetts.

---

With the growing use of gene panels and next generation sequencing in the evaluation of human disease, gene-level description of illness is being realized on a more widespread basis. Despite this growing understanding, genetic variants and their attendant diseases remain difficult to study particularly due to a lack of pragmatic classification of genetic variants within medical nomenclature, *e.g.* International Classification of Diseases.

The internationally recognized HUGO Nomenclature Committee (HGNC) approves labelling of human loci to enable unambiguous scientific communication, including assignment of a unique HGNC ID number, *e.g.* *SYNGAP1* has been assigned the ID HGNC #11497. Even should the name or symbol of a gene change, the assigned HGNC ID remains stable. As a unique, durable identifier of a gene, it is an ideal label of pathogenic genetic variation. To code genetic variants in ICD-10, appending the HGNC ID to an ICD-10 prefix could be performed in the Q9 section (Chromosomal abnormalities, not elsewhere classified) of the Q chapter (Congenital malformations, deformations and chromosomal abnormalities), or to an unused U code chapter (U9 Germline Genetic Variants), *i.e.* Q91.1497 or U91.1497 for “*SYNGAP1* Pathogenic Variant Present, Germline or Postzygotic,” respectively. This methodology preserves the architecture of ICD-10 while utilizing an additional international system to remain consistent with the conventions of biomedical classification with numerous consequent benefits. Use of HGNC ID numbers within the Extension Codes of ICD-11 will then allow ongoing coding of pathogenic variants.

As such codes are entered into electronic medical records and billing databases, high-quality epidemiologic research and retrospective studies are possible. Moreover, in syndromes with multiple individual genetic causes, *e.g.* frontotemporal dementia or autism, gene-specific labelling can facilitate more detailed genotypic:phenotypic correlation or selective recruitment for clinical studies—a particularly important benefit with the rapid emergence of gene-specific therapies. Furthermore, this method can be adapted to add much-needed somatic mutation codes to ICD-10 as oncologic care increasingly considers personalized tumor genomics, *e.g.* U code chapter (U7 Somatic Genetic Variants), U79.047 “*PLAGL2* Pathogenic Variant Present, Somatic.” Use of the HGNC ID within ICD can begin a pragmatic and enduring method of incorporating genetic variation into medical coding.

## PgmNr 675: Phenotype-karyotype relationships in pediatric Turner syndrome.

### Authors:

K. Abihider<sup>1,4,5</sup>; R. Habiby<sup>2,3</sup>; W. Brickman<sup>2,3</sup>; S. Spencer<sup>4,5</sup>; M. Gutierrez-Kapheim<sup>6</sup>; A. Goetsch<sup>3,7</sup>; E. Leeth<sup>1,3,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Graduate Program in Genetic Counseling, Northwestern University, Chicago, IL, USA; 2) Division of Endocrinology, Ann & Robert H. Lurie Children's Hospital, Chicago, IL, USA; 3) Department of Pediatrics, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; 4) Division of Clinical Genetics, Northwestern University, Chicago, IL, USA; 5) Department of Obstetrics and Gynecology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; 6) School of Public Health, University of Illinois at Chicago, Chicago, IL, USA; 7) Division of Genetics, Birth Defects, and Metabolism, Ann & Robert H. Lurie Children's Hospital, Chicago, IL, USA

---

**Background:** Turner syndrome (TS) is associated with both phenotypic and karyotypic variability. Previous studies looking at relationships between karyotype and phenotype have been discordant. The aims of this study were to understand karyotype-phenotype relationships using data from a large pediatric TS cohort, stratify risk of TS-related morbidities across the different karyotype variations and enrich the evidence-base for TS pediatric management recommendations in order to improve prognostic genetic counseling post diagnosis.

**Methods:** Retrospective chart review was performed on 187 TS patients managed by a multidisciplinary clinic at Lurie Children's Hospital of Chicago. Inclusion criteria were a confirmed TS karyotype and at least two endocrinology visits. Karyotypes were categorized into five groups: 45,X; 45,X/46,XX mosaicism; karyotypes with Y chromosome material identified; structural X variants; and mosaicism with Triple X. All clinical data were dichotomized into normal and abnormal for comparative analyses.

**Results:** Statistically significant differences among the karyotype groups were observed in ten of thirty of the clinical variables including: age of diagnosis, congenital cardiac defects, abnormal EKG scores, beta-blocker use, abnormal low-density lipoprotein (LDL), diagnoses of autoimmune thyroiditis, deficient Vitamin D levels, estrogen use, spontaneous menarche, and spontaneous thelarche. Novel findings in this study included an increased risk for electrocardiographic abnormalities, beta-blocker use, and abnormal LDL levels in the 45,X group; and increased risk of Vitamin D deficiency in the mosaic 45,X with Triple X karyotype.

**Conclusions:** In this cohort, 45,X was typically the most severely affected and structural X karyotypes showed an increased risk of autoimmune thyroiditis and pubertal delay. Some of the novel findings in this study corroborate the previous assertion that 45,X karyotypes are the most morbid, particularly regarding cardiac findings. Mosaic Turner syndrome with 46,XX, Y chromosome material, and Triple X karyotypes have some more mitigated features, particularly regarding puberty measures, although evaluation for all Turner syndrome associated comorbidities is still warranted. While Turner syndrome remains variable in presentation of both phenotype and karyotype, there is evidence of relationships between the two, which may add value to counseling regarding prognosis for patients and families at the time of diagnosis.

## **PgmNr 676: Development of methods to expand the ClinGen Actionability Working Group framework to include assertions of the clinical actionability of genomic variation in the context of secondary findings.**

### **Authors:**

J. Hunter <sup>1</sup>; B. Bostwick <sup>2</sup>; O. Dagan-Rosenfeld <sup>3</sup>; K. Dipple <sup>4</sup>; A.K.M. Foreman <sup>5</sup>; A. Freed <sup>6</sup>; M. Gilmore <sup>1</sup>; A. Katz <sup>7</sup>; K. Lee <sup>5</sup>; N. Lindor <sup>8</sup>; K. Manickam <sup>9</sup>; C.L. Martin <sup>10</sup>; J. McConnell <sup>11</sup>; N. Meeks <sup>12</sup>; K. Mittendorf <sup>1</sup>; J. O'Daniel <sup>5</sup>; J. Posey <sup>2</sup>; E.M. Ramos <sup>7</sup>; S. Rego <sup>3</sup>; R.D. Steiner <sup>13</sup>; T. Trotter <sup>14</sup>; J. Vengoechea <sup>15</sup>; K. Wallace <sup>5</sup>; M. Weaver <sup>16</sup>; M.S. Williams <sup>17</sup>; A. Buchanan <sup>17</sup>; B. Powell <sup>5</sup>; K.A.B. Goddard <sup>1</sup>; on behalf of the ClinGen Resource

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Genetics, Stanford University, Stanford, CA; 4) Department of Medical Genetics, Seattle Children's Hospital, Seattle, WA; 5) Department of Genetics, University of North Carolina, Chapel Hill, NC; 6) Division of Medical Genetics, University of Washington, Seattle, WA; 7) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 8) Department of Health Sciences Research, Mayo Clinic, Scottsdale, AZ; 9) Division of Genetic and Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; 10) Autism & Developmental Medicine Institute, Geisinger, Danville, PA; 11) Department of Cardiology, GeneDx, Gaithersburg, MD; 12) Department of Clinical Genetics and Metabolism, Children's Hospital Colorado, Aurora, CO; 13) Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, WI; 14) Department of Pediatrics, John Muir Health, Concord & Walnut Creek, CA; 15) Division of Medical Genetics, Emory University, Atlanta, GA; 16) American College of Medical Genetics and Genomics, Bethesda, MD; 17) Genomic Medicine Institute, Geisinger, Danville, PA

---

The ClinGen Actionability Working Group (AWG) has developed a standardized, evidence-based framework to assess the clinical actionability of genes and associated disorders in the clinical context of returning secondary findings of genome-scale sequencing. This framework includes the generation of summary reports of clinical actionability based on available evidence for clinical outcomes and interventions that could improve future health outcomes in these conditions. A semi-quantitative metric is then applied to generate consensus scores for four domains of clinical actionability for specific outcome and intervention pairs: 1) severity of the outcome; 2) likelihood of the outcome to occur; 3) effectiveness of the intervention to prevent or minimize the outcome; and 4) nature of the intervention, akin to risk and burden of the intervention to the individual. These publicly available materials provide evidence-based resources to guide development and implementation of recommendations and policies regarding the return of secondary findings. Most recently, the AWG has begun to expand this framework by developing methods to generate AWG assertions on the clinical actionability of genomic variation as secondary findings. To inform this effort, a survey was developed to collect AWG member perspectives on the degree of actionability of 15 topics previously assessed by the AWG. For each topic, the relevant summary report and consensus scores were provided with a

0 to 100 sliding scale to capture the scorer's perspective of clinical actionability, where 0 indicated "not at all actionable," 100 indicated "very actionable," and 50 indicated an "actionability threshold." To date, 15 AWG members have completed this survey. The median score for all 15 topics was above the "actionability threshold." Two of the 15 topics are on the ACMG SF 2.0 list and scored relatively high on this scale compared with other topics. Interestingly, actionability scores were variable within each topic, with the inter-rater coefficient of variation ranging from 13.4 to 53.7 across topics. This variability highlights the complexities and challenges of making assertions of clinical actionability. Next steps to modify the methods to provide AWG actionability assertions will be discussed. Overall, actionability assessments of the ClinGen AWG provide a significant resource to aid stakeholders and decision makers as they determine best practices regarding return of secondary findings.

## PgmNr 677: Development of an educational tool: Approach to addressing the direct to consumer genetic testing trend.

### Authors:

H. Ayoubieh; T. Higashimoto; R. Forsyth; W. Mu; J. Bodurtha

[View Session](#) [Add to Schedule](#)

**Affiliation:** McKusick-Nathans Institute of Genetic Medicine and Department of Pediatrics. Johns Hopkins University, Baltimore, MD.

---

Direct to consumer (DTC) genetic testing (GT) for health conditions has become easily accessible to the general public via several models. The US FDA has cleared at least one company to test for specific pathogenic variants. The company asks that consumers consult with their doctors to discuss concerns related to the results. Consumers can also utilize 3rd party analysis for interpretation of their raw genetic data. Of the variants reported by this additional analysis, ~40% may be false positives, thus highlighting the importance of confirming the results through clinical GT. Many health-care professionals are not familiar with handling GT results, and there is no consensus information about methods to communicate and manage DTC GT. Medical literature regarding the approach to DTC GT includes case-based scenarios for individuals presenting to primary-care with DTC GT variants in BRCA1/2 genes, Alzheimer disease and polycystic kidney disease. To our knowledge, education about DTC GT in ambulatory undergraduate medical education has not occurred. Practical strategies in a more comprehensive framework to help health-care professionals, both PCPs and specialists to navigate these results are needed as evidence evolves. We describe an approach piloted in the ambulatory clerkship at Johns Hopkins SOM to address the DTC GT trend. We have begun to offer a 90-minute Genetics in Ambulatory Care education session that focuses on DTC GT as an optional part of our medical student and postgraduate trainee elective. Five learners have thus far participated in the module. Post session survey showed that all participants agreed that the session was relevant to their future careers and recommended offering the session to future learners. Based on their input, we identified gaps in the knowledge of this target group and created a preliminary educational tool/algorithm to approaching DTC GT results. 15 genetics professionals and trainees also contributed to the algorithm. Next steps for presentation: An electronic survey will be distributed to a larger group of genetics professionals and trainees to further analyze the tool. A group of non-genetics health professionals and patients will be later invited to pilot the tool. We plan to suggest strategies to aid health-professionals and patients navigate DTC GT results.

## PgmNr 678: Tangibles for genetics learning experiences.

**Authors:**

M.K. Konkel <sup>1</sup>; B. Ullmer <sup>2</sup>; A. Siqueira <sup>2</sup>; M. Nasiri <sup>2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Department of Genetics & Biochemistry, Clemson Center for Human Genetics, BDSI Core Faculty Member, Clemson University, Clemson, SC; 2) Human Centered Computing Division, School of Computing, Clemson University, Clemson, SC

---

Genetics has evolved from a niche discipline to standing as one of the five core recognized conceptual areas for biology students. One could argue that basic genetics literacy is instrumental for all students, with genetics analyses starting to become routine in health care across the United States and beyond. Moreover, consumer testing is becoming more popular, with many college level students and/or close relatives having used consumer-testing services. Toward reaching students from different backgrounds, learning styles, and with varying level of expertise, it is important to offer different ways to convey content that go beyond classical lectures. At the same time, while “flipped classrooms” have recently met some advocacy, they are very difficult to realize effectively and apply to large classrooms; and thus, consequently are often not appreciated by students. To address these challenges, we developed instrumented physical artifacts (or “tangibles”) to enhance the genetics learning experience. Starting with the NIH 3D Print Exchange, we are developing 3D printed objects for molecular genetics toward use in the classroom for exploration and explanation. We also encourage students (often as part of group projects) to print their own 3D models to broaden their learning experiences, enable students to learn new skills, and use the resulting tangibles as both passive and active elements of their presentations. (Our institution has makerspaces with student access.) Finally, we have also begun integrating sensing, display, and actuation electronics and software into 3D printed objects, toward further increasing the functional, representational, cognitive, and communicative potential of such practice.



## PgmNr 679: Genomics educational priorities, knowledge and needs among healthcare providers: Preliminary survey results.

### Authors:

D. Messersmith <sup>1</sup>; G.E. Fotso Chekam <sup>1</sup>; C. Daulton <sup>1</sup>; C. Easter <sup>1</sup>; J. Jenkins <sup>2</sup>; K. Calzone <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Education and Community Involvement Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland.; 2) Clinical Genomics Consultant, Harmony, FL.; 3) Center for Cancer Research, Genetics Branch, National Cancer Institute, NIH, Bethesda, MD.

---

**Background:** The implementation of genetics and genomics into clinical care has important implications for all healthcare providers, yet opportunities to upskill practitioners are limited.

**Aim:** To better understand the priorities, knowledge, and needs of healthcare providers for genomics educational resources.

**Methods:** We developed a survey incorporating both new and existing questions from validated instruments. Through collaborations with professional associations, the survey was distributed from July to September 2018 to healthcare providers.

**Results:** Data was analyzed from physicians (N=46), nurses (N=31), physician assistants (N=22), and pharmacists (N=48). Most respondents (78%; 115/147) thought it was very important for healthcare providers to become more educated about genetics and genomics of common disease. Almost all (97%; 142/146) thought that a potential advantage of integrating genetics of common disease into practice would lead to better treatment decisions and 91% (132/145) thought another advantage would be improved services to patients. A potential disadvantage noted by 78% (113/144) was that genetics/genomics may not be reimbursable/too costly. Another potential disadvantage noted by 67% (97/145) was an increase in insurance discrimination. When asked to rank the importance of incorporating genetics and genomics into their practice, 10% (15/145) reported low priority, 36% (52/145) moderate priority, and 54% (78/145) high priority. Regarding their genetics/genomics knowledge, 21% (31/146) rated knowledge as excellent, 53% (77/146) as good, and 26% (38/146) as poor. When asked "where do you go for genetics and genomics training and resources", the majority selected professional organizations, the internet, and peer-reviewed literature.

**Limitations:** Limitations of the data were a lack of diversity among respondents, low response rate, and a need for additional questions.

**Conclusions:** These results provide preliminary information on genomics priorities, knowledge and needs among healthcare providers. Because the majority indicated they go to their professional organizations for information about genetic and genomic training and resources, collaborating with these organizations is a potential educational and dissemination opportunity for the National Human Genome Research Institute.

**Next Steps:** We plan to distribute a revised, expanded survey to increase the overall response rate, the diversity of respondents, and to further explore our findings.



## PgmNr 680: Assessment of the knowledge and attitude regarding biobanking among undergraduate medical students in a Nigerian university.

### Authors:

O.I. Okoye <sup>1,2</sup>; A. Aghaji <sup>1,2</sup>; F.C. Maduka-Okafor <sup>1,2</sup>; N. Uche <sup>1</sup>; N.N. Uzokwe <sup>2</sup>; U. Ezemba <sup>3</sup>; O. Okoye <sup>1</sup>; R.E. Umeh <sup>1,2</sup>; C. Nkwocha <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ophthalmology, College of Medicine, University of Nigeria, Enugu, Enugu, Nigeria; 2) Bioethics Research Interest Group-UNN, Enugu, Nigeria; 3) Department of Ophthalmology, University of Calabar Teaching Hospital, Calabar, Nigeria

---

With the recent advances in genetic and biomedical technologies, there has been a dramatic increase in the value of stored biospecimens. Researchers are now developing protocols to examine such tissues. With the paradigm shift in healthcare towards personalized medicine, biobanks are becoming primary means of delivering personalized diagnostics and treatment. All these require interactions between many stakeholders, including healthcare staff. In Africa, genomic methods are gradually being adopted in biomedical research, though the status of genetics and genomic medicine is far from meeting the needs of the population. Even with the curricular reforms in medical education, little is known about physicians/medical students' knowledge and attitudes concerning medical genetics and genomics in Africa. Several studies have described the role of the public and patients in tissue banking issues and other related matters, but very few have also actually studied the role of healthcare staff and medical students. We assessed the knowledge of Nigerian medical students about biobanks and their willingness to donate their biospecimens for research. Using self-administered questionnaires, a cross-sectional study of the consenting respondents among 143 eligible senior medical students in a Nigerian University was conducted in May 2019. Data analysis was done with SPSS version 22. The sample included 79 males and 35 females, and the mean age was 23.7 years. Only 28.1% knew the difference between genetics and genomics, and only 14% were confident of their knowledge of genetics and genomics. The majority (73%) were either not aware or not sure of the term 'biobank', though 55.3% were willing to donate specimens for research; with the chief reason for willingness being altruism. The main reason for unwillingness to donate was fear/uncertainty about the intentions of the researchers. A majority (64%) had never participated in health research. Blood, sputum, urine and stool specimens were the specimens the respondents were most willing to donate. Only 7% of the respondents were willing to donate specimens from dead relations. Fifty-five (48.3%) students were unwilling to donate any organ for the medical care of needy patients. There is a low level of knowledge about genetics/genomics and biobanking among the students. Their attitude to donation of specimens for research is guarded. There is a need to improve biomedical research education and medical genetics/genomics teaching among medical students.

## **PgmNr 681: Genetics education pipeline for elementary, middle, high school (K-12) and undergraduates to enhance diversity in sciences, technology, engineering, and mathematics (STEM) careers at the Louisiana State University Health Sciences Center (LSUHSC).**

### **Authors:**

F. Tsien <sup>1</sup>; J. Peñas <sup>1</sup>; A. Umrigar <sup>1</sup>; M. Cuccia <sup>2</sup>; A. Augustus-Wallace <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Dept of Genetics, Louisiana State University Health Sciences Center, New Orleans, Louisiana.; 2) Health Policy & Systems Management, School of Public Health, and Louisiana State University Health Sciences Center, New Orleans, LA; 3) Department of Medicine & Office of Diversity and Community Engagement, School of Medicine, Louisiana State University Health Sciences Center-New Orleans, LA

---

The Louisiana State University Health Sciences Center in New Orleans (LSUHSC) created a genetics educational pipeline through the Science Youth Initiative (SYI) and the Research Experiences for Undergraduates (REU) programs. The long term goals are to increase diversity in STEM fields especially in genetics. Underrepresented groups as defined by the NIH and NSF include African American/Blacks, Hispanic/Latinos, Native American/American Indians, Pacific Islanders, Alaska Natives, women, persons with disabilities, first generation college students, students residing in rural areas, non-traditional students, and veterans. SYI instruction includes hands-on experiments, workshops, and educational videos which are coordinated with elementary, middle, and high school (K-12) teachers to maximize learning and complement their curriculum. The goals of the SYI are to improve academic achievements, introduce diverse role models, increase awareness of genetics and science careers, and provide LSUHSC trainees instruction in communication skills to the lay public. Topics range from strawberry DNA isolation (elementary schools), to forensics and HeLa cells (high schools). Since 2009, the SYI has reached > 2900 K-12 students, of which 72% belong to underrepresented groups. Diverse LSUHSC trainees from the LSUHSC Schools of Medicine, Graduate Studies, and Public Health help educate K-12 students. The REU program provides undergraduates scientific problem solving skills, state-of-the-art laboratory techniques, scientific communication skills, and resume writing instruction for competitive graduate school preparation. Since 2014, the NSF-funded REU program provided internships to more than 50 undergraduates. General topics include population genetics, molecular and cellular biology, chromatin instability, mutation analysis, translational research, and bioinformatics. Following the REU program, students co-author scientific publications and conference presentations, obtain laboratory employment, and are accepted to STEM programs. Participating schools include rural, urban, public, private, and historically black colleges and universities (HBCUs). Assessments are by qualitative and quantitative surveys and NSF-REU common assessment tools. Students are tracked via on-line reporting systems. The SYI and REU programs have been successful in providing opportunities in STEM especially in the field of genetics, to K-12 and undergraduate students to help them achieve career goals.

## PgmNr 682: Genomics2020: An overview of the ongoing NHGRI strategic planning process.

**Authors:**

A. Wise; National Human Genome Research Institute

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, Maryland.

---

The National Human Genome Research Institute (NHGRI) has a rich tradition of strategic planning that dates back to the Human Genome Project. The significant advances in genomics since the publication of NHGRI's last strategic plan in 2011 illustrate the need for an updated vision about the future of genomics for 2020 and beyond. In February 2018, NHGRI launched a new round of strategic planning, one that will yield publication of a new strategic plan in Fall 2020. The new plan will detail the most compelling areas of genomics research and their application to human health and disease; it will also inform NHGRI's scientific priorities and shape its research portfolio, ensuring that the Institute continues to be a driving force at the *Forefront of Genomics*. To formulate relevant and forward-reaching goals, NHGRI is fostering partnerships with research, healthcare, education, and policy communities, as well as the public. Starting in 2018, NHGRI has engaged a wide variety of stakeholders through workshops, town halls, advisory meetings, meeting sessions, and webinars. Over the next year, NHGRI plans to use the received input to develop draft tenets for the emerging plan as well as collect additional input through more focused activities. We continue to encourage the genomics community to participate in the process via [genome.gov/genomics2020](http://genome.gov/genomics2020), [genomics2020@nih.gov](mailto:genomics2020@nih.gov), and [#genomics2020](https://twitter.com/genomics2020). This poster, presented by NHGRI staff, provides an overview of the ongoing NHGRI strategic planning process to date.

## PgmNr 683: Integrating modern genetics, genomics, and bioethics into high school curricula: The *Teaching the Genome Generation* program.

### Authors:

C. Wray <sup>1</sup>; S. Wojiski <sup>2</sup>; M. McKernan <sup>1</sup>; D. Waring <sup>3</sup>; K. Bass <sup>4</sup>; A. Kieffer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Education, The Jackson Laboratory, Bar Harbor, Maine.; 2) Genomic Education, The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut; 3) Personal Genetics Education Project, Department of Genetics, Harvard Medical School; 4) Rockman et al., San Francisco, California

---

Teaching the Genome Generation (TtGG) is a high school teacher professional development program (PD) and curricular and laboratory resource effort that supports classroom instruction across three broad topics: molecular genetics, bioinformatics, and bioethics. The three instructional themes are aligned with Next Generation Science Standards. Modular lesson plans, laboratory protocols and instrument and reagent support are provided to teachers who participate. TtGG is unique because it uses voluntary, anonymous sampling and isolation of high school student genomic DNA for use in subsequent genotyping demonstrations. Bioinformatics exercises are designed to reinforce content presented in the laboratory exercises and highlight the deep levels of DNA sequence variation that exists in the human genome. Bioethics lessons, from the Personal Genetics Education Project (pgEd), align with and support the molecular genetics arm of the program and model how to introduce and facilitate discussion on controversial and sometimes difficult subjects with high school students. To date, 163 teachers have taught over 10,000 high school students content provided by the TtGG program. Lessons and laboratories are flexible and are taught across a range of high school courses; 40% of students are exposed to the curriculum in their introductory biology course, while 22% and 19% of students participate in the curriculum in honors or AP biology, respectively. The remainder of students experience the curricular elements in a biotechnology or other science course. Participating teachers from New England implement lessons and labs at a high rate; however, incorporation of bioinformatics lessons in high school classes lags behind laboratory and bioethics classroom integrations. In 2018, new case-based and narrative bioinformatics lessons were introduced in order to increase integration of bioinformatics into high school classes. Evaluation data collected over several years indicate that TtGG has increased teachers' abilities to integrate complex concepts of genetics and genomics into their high school classes. Self-efficacy items show increased teacher confidence when tested before the professional development (PD) and at the time of the school implementation, sometimes months later. For example, respondents from the 2017 PD cohort (n=27) are 17% more confident on tasks related to use of bioinformatics tools at the time of implementation than before the PD.

## PgmNr 684: An opportunity to restore scientific justice? Willingness to participate in genetic research in two rural Puerto Rican communities.

### Authors:

T. Dye <sup>1</sup>; C. Vélez-Vega <sup>2</sup>; J. Pérez-Ramos <sup>1</sup>; N. Cardona-Cordero <sup>1</sup>; D. Vega-Ocasio <sup>1</sup>; S. McIntosh <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Obstetrics and Gynecology, University of Rochester, Rochester, NEW YORK.; 2) School of Public Health, University of Puerto Rico, San Juan, PUERTO RICO; 3) Public Health Sciences, University of Rochester, Rochester, NEW YORK.

---

**Background:** Historic bioethical abuses in the conduct of science and medical intervention in Puerto Rico has led to an atmosphere of scientific distrust and subdued interest in being included in research. As new, large research initiatives have created Puerto Rican-centered infrastructures, researchers are increasingly interested in engaging communities - equitably - in genetic research. Such engagement could help accomplish distributive justice in that communities may ultimately benefit from the research where they participate. We ascertained attitudes and intent to participate in genetic research in two rural communities of Puerto Rico where new tissue-based genetic research was being initiated.

**Methods:** This study used an in-person cross-sectional design to ascertain community members' attitudes toward genetic research and likelihood of participating in genetic research studies in two rural community health centers in the northern Karst region of Puerto Rico. We ascertained the relationship between willingness to participate in genetic research and psychometric, attitudinal, demographic, and general health indices.

**Results.** Overall, 44.8% indicated they were very/likely to participate in genetic research. Attitudes toward genetic research were strongly associated with likelihood of participating in such research: the Composite Attitudes Score was approximately 30% higher for participants very/likely to participate in genetic research versus others ( $p < .001$ ). Respondents indicating they were very/likely to participate in genetic research were significantly more likely to trust researchers ( $p < .001$ ); indeed, respondents indicating they trusted researchers were more than six times more likely to participate in genetic research than were respondents not indicating trust (84.6% v. 12.5%). Additionally, respondents perceiving that research could benefit them/ their family ( $p < .005$ ) were more likely to indicate they would participate in genetic research.

**Discussion.** Likelihood of participating in genetic research varied sharply by demographics, attitudes toward genetic research, and by psychosocial classifications in this sample of rural Puerto Ricans. Ensuring an informed understanding of what genetic research entails and overcoming "lack of trust" in researchers (*No confío en los investigadores*) - one of the strongest determinants of participation in research identified in this study - are critical in successfully partnering with communities.

## PgmNr 685: Scoping review of participant-centric initiatives in medical research.

### Authors:

N. Hamakawa <sup>1</sup>; R. Nakano <sup>2</sup>; A. Kogetsu <sup>1</sup>; J. Kaye <sup>3</sup>; B. Yamamoto <sup>2</sup>; K. Kato <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Ethics and Public Policy, Graduate School of Medicine, Osaka University, Japan; 2) Graduate School of Human Sciences, Osaka University, Japan; 3) Centre for Health, Law and Emerging Technologies, Nuffield Department of Population Health, University of Oxford, UK

---

The importance of patient and public involvement in medical research is increasingly recognized. There is a discernable movement away from a paternalistic model of medical research to one in which patients are actively involved in the research process to better reflect their experiences and needs. Participant-centric Initiatives (PCIs) are defined as “tools, programs, and projects that empower participants to engage in the research process” using digital technologies. Studies point to the effectiveness of using information and communication technology (ICT) for medical research via PCI platforms. Not only do they allow more patients to participate in research, but interactions between research participants and researchers are facilitated. However, there were no reports on the development of PCIs since 2012, and existing reports were limited in scope. Furthermore, previous studies had not sufficiently clarified how patient involvement is facilitated by PCIs.

The objective of this study was to clarify the status of PCIs and the key features of participant involvement using ICTs. In previous studies, patient involvement was broadly defined across a spectrum of distinct levels of involvement activity. Therefore, we focused on the question of what kinds of patient/participant involvement are facilitated by PCIs, aiming at clarifying the characteristics of different levels of participant involvement activity.

By applying a scoping review methodology, a literature search was conducted in 2017. After screening 6,568 documents and websites, 152 documents were found relevant to PCIs. By examining the abstracts of those documents, we extracted activities related to PCIs and analyzed their characteristics and key features. We found that they can be classified into two distinct groups in terms of the levels of participant involvement. Group 1: Patients are actively involved in the research as partners, by making decision or leading research activities. Group 2: Patients participate in the research by providing feedback or input, or to gain knowledge about research and ask questions to researchers. The preliminary results revealed that terminology usage and definitions of involvement are varied and inconsistent. By comparing these groups, we illustrate that the active involvement facilitated by PCIs can be achieved by continuous dialogue between participants and researchers, and priority setting by participants.

## **PgmNr 686: Regulating preimplantation genetic diagnosis across the world: A comparison of trends and international policy perspectives.**

### **Authors:**

R. Isasi; M. Ginoza

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, University of Miami Miller School of Medicine, Miami, Florida.

---

Preimplantation genetic diagnosis (PGD) is a reproductive technology which, in the course of in vitro fertilization (IVF), allows prospective parents to select their future offspring based on genetic characteristics. During the last decade, the regulation of PGD has followed mostly a linear path, with incremental changes driven by scientific advances or by court driven alterations to legal practice. But also, shift has been prompted by greater societal uptake. No longer considered an experimental practice, stakeholders (e.g. scientists, patients, etc.) across the world have challenged the status quo, recognizing that both technological and policy developments do not take place in a moral vacuum. They have done so under the flag of reproductive freedom, equitable access and welfare considerations, amongst other ethical concerns. Despite the passage of time, the ubiquitous presence of eugenics seems to constantly cast a shadow over technologies that open the door for selecting the genetic characteristics of a human offspring. For this reason, we can predict that despite greater societal acceptance, widening the permissibility criteria for PGD will remain contested and a matter of continue social and policy debate.

In this presentation, we examine – from a comparative perspective - variations in policy approaches to the regulation of PGD. We draw on a representative sample of 19 countries (Australia, Austria, Belgium, Brazil, Canada, China, France, Germany, India, Israel, Italy, Japan, Mexico, Netherlands, Singapore, South Korea, Switzerland, United Kingdom, and the United States) to provide a global landscape of the spectrum of policy and legislative approaches (e.g. restrictive to permissive, public vs. private models). We also explore central socio-ethical and policy issues and contentious applications as for instance, permissibility criteria (e.g. medical necessity), non-medical sex selection and reproductive tourism. Finally, we further outline genetic counseling requirements across policy approaches.



## **PgmNr 687: GenEtiquette: Developing guidelines for the conduct of genomic research in collaboration with Aboriginal and/or Torres Strait Islander peoples of Queensland, Australia.**

### **Authors:**

G. Pratt; M. Vidgen; S. Kaladharan; D. Whiteman; C. Brown; J. Pearson; N. Waddell

[View Session](#) [Add to Schedule](#)

**Affiliation:** QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

---

Precision medicine serves to benefit detection, diagnosis and personalising of the health care experience. It is widely recognised that there is significant bias in genomics databases towards people of European ancestry. Aboriginal and/or Torres Strait Islander residents of Queensland, Australia experience significantly higher mortality and morbidity and lower life expectancy compared with non-indigenous Queenslanders. With a growing emphasis on genomics in healthcare, there is a risk that advances in genomics will not benefit Indigenous Australians. In 2018, QIMR Berghofer undertook to develop guidelines to support genomic researchers to partner with Indigenous Queenslanders. Development of the guidelines was informed by community engagement across the health service, research and ethics sector. Led by an Aboriginal man, the project was developed and delivered by a team with complementary skills in stakeholder engagement, epidemiology, medicine, genomics, ethics and biotechnology. Governance of the project was overseen by a project steering committee comprising noted Indigenous experts from the university, health service, research, policy and ethic sectors of Queensland. Community and stakeholder forums availed an opportunity to provide input, feedback and advice about the topics of genomics, genetics and precision medicine and to directly inform the development of the guidelines. The guidelines provide practical suggestions for genomic researchers to evidence a commitment to Indigenous research and to respect the right to self-determination and empowerment of Aboriginal and/or Torres Strait Islander leadership and voice. Consultation and engagement are essential to the uptake and translation of key learnings, ensuring that consumers, families, communities, service providers, policy personnel, researchers and clinicians are informed about, respected and empowered to make decisions and influence health action. The Aboriginal and/or Torres Strait Islander community of Queensland, Australia showed overwhelming support for our engagement and the opportunity to learn more about genomics and precision medicine. The guidelines serve as a launch pad to support relationships, conversation and partnership in the area of Indigenous genomics in Queensland, Australia. Key learnings serve to benefit the international community about the value of deliberate and empowering strategies to encourage public discourse about genomics and precision medicine.



## **PgmNr 688: Regulation of international direct-to-participant genomic research: Results of a 32-country study.**

### **Authors:**

M.A. Rothstein <sup>1</sup>; M. Zawati <sup>2</sup>; B.M. Knoppers <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute for Bioethics, Health Policy & Law,, Univ Louisville School of Medicine, Louisville, Kentucky.; 2) McGill University Montreal, Canada

---

### Regulation of International Direct-to-Participant (DTP) Genomic Research: Results of a 32-Country Study

Mark A Rothstein, Ma'n H Zawati, Bartha Maria Knoppers

The online recruitment of participants is helping to facilitate genomic research on rare disorders. With the development of cooperative arrangements between researchers and disease-specific patient organizations, interested patients and caregivers can be easily recruited, without relying on traditional recruitment models that depend on physicians and hospitals. Internet recruitment has the further advantage of allowing prospective participants in remote locations to take part in research. Such participants may supply samples for genome sequencing and provide access to electronic health records. International direct-to-participant (DTP) research promises to greatly increase the number and diversity of participants for research on rare disorders, but participant recruitment across international borders is often thwarted by prohibitions on research without ethics review in each participating country. The process of obtaining ethics review in multiple jurisdictions, while an essential part of traditional research, can quickly become burdensome and infeasible. Researchers may be hesitant to undertake such effort only to enroll a few participants in each studied country. Thus, although global recruitment is a promising strategy, much work remains in devising methods that are legally compliant and ethically acceptable for this novel research. There is an urgent need to understand the international regulatory landscape in order to foster greater potential for a global DTP genomic research model. In this presentation, we will outline the results of an NIH-funded study in 32 countries on six continents. Drawing on the reports of our local expert collaborators, we will describe the state of the law and genomic research practices, and discuss the measures needed to ensure the successful implementation of international DTP genomic research.

## **PgmNr 689: The latent class analysis to clarify the factors associated with the structure of the public attitude toward the genome research at the survey in 2016 in Japan.**

### **Authors:**

Z. Yamagata <sup>1</sup>; K. Muto <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department Hlth Sci, Univ Yamanashi, Chuo, Yamanashi, Yamanashi, Japan; 2) The Institute of Medical Science, The University of Tokyo, Tokyo, Japan,

---

[Aim] The aim of this study was to clarify the structure of the public attitude toward the genome research by the latent class analysis.

[Methods] The nationwide surveys about the attitude toward the genome research were conducted in 2016 in Japan. The participants were comprised of 2,000 people (age, 20-69), selected from the Japanese general population by using the two-step stratified random sampling method.

[Results] Five clusters were assumed as an explanation model of six variables related to the knowledge of genome and attitudes toward genomic research promotion about three themes; basic genome research, genome research related to agriculture and medicine at the survey in 2016. They were able to be named "Group of aggressive promotion" (41.5%), "Group of passive support" (20.8%), "Group not making judgment" (19.2%), "Group making prudent judgment" (18.5%), and "Group not interested in genome". The results in 2008 and 2009 were almost the same as that in 2005. It is possible to forecast to which cluster to belong according to respondent's attribute, and we can forecast the reaction to other questions by using a cluster oppositely. For examples, "Group of aggressive promotion" is the layer of a high academic background, and is positive to donate their blood for the genome research, "Group making prudent judgment" is high academic background persons as same as "Group of aggressive promotion", and is interesting in the science and technology, but is negative to the blood donation for the research.

## **PgmNr 690: “Additional findings”: Expected and unexpected benefits of All of Us for federally qualified health centers (FQHCs) and community-based organizations (CBOs).**

### **Authors:**

J. Yu <sup>1,5</sup>; R. Chang Weir <sup>2</sup>; S. Grayson <sup>3,4</sup>; N.A. Garrison <sup>5</sup>; M. Doerr <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) University of Washington School of Medicine, Box 356320 1959 NE Pacific St HSB RR307, Seattle, WA 98195; 2) Association of Asian Pacific Community Health Organizations, 101 Callan Avenue, Suite 400, San Leandro, CA 94577; 3) Institute for Public Health Genetics, University of Washington, 1959 NE Pacific St HSB, Seattle, WA 98195; 4) Sage Bionetworks, 2901 3rd Ave #330, Seattle, WA 98121; 5) Treuman Katz Center for Pediatric Bioethics, Seattle Children’s Research Institute, 1900 Ninth Ave, M/S JMB-6, Seattle, WA 98101

---

The All of Us Research Program has the potential to enable precision medicine for all people in the U.S. Likely, this vision will be realized over the course of decades and depends on the engagement of FQHCs and CBOs to recruit and retain a diverse cohort, and to strengthen newly formed institutional partnerships. These institutions are not only critical linchpins for equitable inclusion but are key stakeholders who anticipate and experience benefits and risks of All of Us participation in the here-and-now. Building on findings from interviews with FQHC medical directors and CBO executive directors prior to All of Us, we conducted two related qualitative research projects that inform our understanding of organizational and community benefits of the All of Us Research Program. Interview and focus group participants (n=10) from All of Us-funded and non-funded FQHC members of the Association of Asian Pacific Community Health Organizations suggest that experienced and expected benefits include resources for building IT and EHR infrastructure, improving patient portal access and use, activating patients on their health conditions, and increasing opportunities for conducting community-based research through partnering with local investigators. Interviews (n=16) with leaders of minority health CBOs, citizen scientists, and patient-led researchers, suggest that supporting communities to engage in big data research may be a welcomed ancillary benefit with impacts beyond precision medicine. These potential impacts include more authentic relationships with researchers and tangible benefits from community-led research that support advocacy and agency as patients and communities. We conclude that while participants recognized the value of equitable precision medicine based on the availability of more inclusive data, more proximal and practical benefits appear to drive both interest and commitment to All of Us. Understanding and articulating the full range of expected and unexpected proximal benefits and even risks of All of Us to FQHCs and CBOs may support organizational commitments by clarifying its value, improve community engagement including subsequent participant recruitment and retention, and demonstrate a commitment to enduring partnership.

## PgmNr 691: Clinical exome and chromosomal microarray reanalysis: Balancing the burden on laboratories, clinicians, and families.

### Authors:

E.T. DeChene <sup>1</sup>; M.A. May <sup>3</sup>; L.B. Lulis <sup>1</sup>; E. Bedoukian <sup>2</sup>; L. Medne <sup>2</sup>; I.D. Krantz <sup>2</sup>; J.R. Murrell <sup>1</sup>; L.K. Conlin <sup>1</sup>; M.C. Dulik <sup>1</sup>; N.B. Spinner <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genomic Diagnostics, Department of Pathology & Laboratory Medicine, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Roberts Individualized Medical Genetics Center, Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) UPMC-Magee Pittsburgh Clinical Genomics Laboratory, Pittsburgh, PA

---

In a recent policy statement on re-contact after revision of genomic results, the ACMG states that responsibility for re-contact is shared between ordering clinicians, clinical laboratories, and patients, and indicates that although there is an ethical obligation to attempt to re-contact patients with new medically actionable results, the patient has the right to decline such re-contact. As our understanding of the molecular basis of genetic disease increases, so does the likelihood of finding a diagnosis through case-level reanalysis of previously non-diagnostic genomic tests. Studies estimate that at least 60-70% of clinical exomes are inconclusive or negative after initial testing, and subsequent reanalysis identifies a diagnosis in ~10-15% of cases. Thus, many clinical laboratories are performing clinician-requested and/or laboratory-initiated case-level reanalysis on previously completed non-diagnostic cases. Clinician-requested reanalysis (CRR) is often ordered at a follow up visit, allowing the patient/family to affirm their interest in reanalysis and their understanding of potential outcomes. In contrast, laboratory-initiated reanalysis (LIR) is usually instigated unbeknownst to patients and clinicians as part of laboratory quality improvement efforts or routine reanalysis. Although this method may be more streamlined and fruitful than CRR, it can be challenging due to the resource requirements for variant analysis and results reporting. In addition, LIR can increase the burden on clinicians to provide results, particularly if they are no longer in contact with the family, and on families who may have to deal with complex emotional ramifications. The family may no longer be seeking a diagnosis or may be facing medical/emotional struggles that outweigh the desire for new information. Thus, while an increased diagnostic rate leads to improved medical management and a shortened diagnostic odyssey, reanalysis can have disadvantages. Through surveys and/or interviews with laboratory personnel and frequent orderers of genomic testing at CHOP, our qualitative study explores laboratories' and clinicians' experiences with identifying, receiving and sharing results from reanalysis. Our goal is to understand the impacts and identify potential enhancements in processes and communication between the three groups. It is essential to balance the improvements in diagnosis and medical care with the potential repercussions for clinicians, families, and laboratories.

## PgmNr 692: An analysis of the factors affecting family adaptation for families of people with rare diseases.

**Authors:**

I.H. Choi; H.W. Yoo

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medical Genetics Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.

---

Family burden, family resilience, and patient support group activities affect the adaptation for the patients with rare genetic diseases and their family members.

In the current study, the relationships were assessed between variables that may affect family adaptation for patients with rare diseases and their families and the mediating effects of family resilience, and the moderating effects of patient support group activities between family burden and family adaptation.

A survey of 245 patients with rare diseases and their families was conducted.

The results of the study can be summarized as follows:

First, regarding the links between family burden, family resilience, patient support group activities, and family adaptation, a significant negative correlation was found between family burden, family resilience, and family adjustment. A significant positive correlation was found between family resilience, patient support group satisfaction, and family adaptation.

Second, regarding the comparison between rare disease groups, the group with neuromuscular disorders and those having problems with treatment, prognosis, and management had a heavier family burden and lower levels of family resilience and family adaptation. They also showed higher levels of patient support group participation and satisfaction.

Third, the lower the level of family burden, the higher the level of family resilience and family adaptation for patients and their families, while the higher the level of family resilience, the higher the level of family adaptation for patients with rare diseases and their families.

These results can be utilized as basic data for defining guidelines and future perspectives regarding support and the provision of higher-quality care at school, hospital, and national level. They will also be helpful in improving family adaptation in the case of patients with rare diseases and their families, as well as in developing educational programs to facilitate family adaptation.

## PgmNr 693: A novel approach to supporting employee benefits programs for genetic wellness.

### Authors:

C.A. Fine; B.J. Swope; E.S. Gordon; L. Meyers; S.M. Weissman; J. Bailey; A. Fan; E. Jordan; B. LeLuyer; A.F. Rope; H. Shabazz; S.B. Bleyl

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genome Medical, 701 Gateway Blvd. Suite 380, South San Francisco, CA 94080

---

Vying for top talent in a tight labor market is difficult for employers as the economy grows and competition increases. Thus, employee packages have started incorporating benefits called “genetic wellness” as a result of an explosion of affordable genetic tests offering medically actionable insights. Employers offering these benefits typically do not partner with genetic clinicians, but rather with a single genetic laboratory offering a specific test. We describe an employee benefit program offering medical genetics services through Genome Medical, a nationwide genomics medical practice, and a proactive genetic screen for cardiac, cancer and other conditions. This program was offered during a three-month timeframe with a registration cap of 200 employees. It included mandatory viewing of an educational webinar of material typically discussed in a genetic counseling session. Employees were required to complete a medical and family history questionnaire which was assessed by a certified genetic counselor (CGC), and had the option of a pre- and/or post-test virtual one-on-one CGC session. Genetic results were interpreted by a CGC in the context of each employee’s reported histories, and a results-based clinical action plan was provided to each participant. 200 employees registered for the program. 189 (95%) individuals watched the webinar. 181 (91%) completed the required questionnaire. Only 25% requested a pre-test CGC session. 170 individuals pursued proactive screening with some of the following motivations: understand a condition in family (42%), help family members (14%), and/or take medical action (12%). 15/75 (20%) individuals were found to have medically actionable genetic variant(s) for hemochromatosis, familial hypertrophic cardiomyopathy, a hereditary cancer syndrome, and/or malignant hyperthermia. 74/170 (44%) individuals were found to be a carrier for at least one condition. Only 20% of participants requested a post-test CGC session. Satisfaction data indicated >80% of responding participants did not prefer an in-person pre-test CGC session; 92% did not prefer an in-person post-test CGC session, program expectations were met in 92% of respondents and 91% found the program to be valuable. Almost 100% of respondents would recommend this program to a coworker. Assisting employers in helping their employees obtain personalized actionable genetic insights can be done in a responsible, scalable and efficient manner, with a high rate of satisfaction.

## PgmNr 694: Clinical characteristics of pregnant women who chose not to have prenatal genetic testing after careful genetic counseling for advanced maternal ages: From an experience in a single institute.

### Authors:

E. Kise <sup>1</sup>; M. Ishikawa <sup>1</sup>; T. Kojima <sup>2</sup>; N. Kikuchi <sup>3</sup>; Y. Fukushima <sup>4</sup>; T. Kosho <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Medical Genetics, Shinshu university hospital, Matsumoto, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Obstetrics and Gynecology, Shinshu University Hospital, Matsumoto, Japan; 4) Shinshu University School of Medicine, Matsumoto, Japan

---

**Background:** In Shinshu University Hospital, three-staged careful genetic counseling (GC) is provided on pregnant women for the decision whether or not to have amniocentesis for advanced maternal ages (AMA): (1) pre-counseling session by certified genetic counselors (CGC) by phone or after prenatal checkups; (2) GC by CGC, collecting detailed clinical and psychosocial information through listening to their narratives; followed by GC by clinical geneticists with CGC, providing comprehensive medical and social information and making free discussion about what is the value and true meaning of the testing for the women and their husbands; (3) post-counseling session by CGC by phone or after prenatal checkups. The purpose of this study is to demonstrate the prevalence of having amniocentesis after such GC and factors influencing the decision whether or not to have amniocentesis.

**Materials and methods:** Pregnant women receiving GC for AMA in our institute from April 2013 to March 2019 were recruited. Detailed information mainly focusing on decision-making process was collected through their medical records and thoroughly reviewed.

**Results:** In total, 230 pregnant women received GC: 102 (44%) had prenatal testing and/or screening (amniocentesis, n=85; maternal serum screening, n=13; NIPT, n=3; and maternal serum screening and amniocentesis, n=1) and the remaining 128(56%) did not. Among 85 pregnant women having amniocentesis, one miscarried, and two had fetuses diagnosed with trisomy 21. In those who had testing and those who did not, the median ages were both 39 years old, the median numbers of previous pregnancies were both 1, and the median numbers of previous miscarriages were both 0. In those who had testing, 24 (23.5%) had histories of infertility treatment, whereas, in those who did not, 51 (39.8%) had these histories. Reasons for not choosing prenatal testing were described as “refusal to terminate their fetuses” and “information that any pregnant women could be mothers of babies with congenital disorders or handicaps and that there is sufficient medical and social support system in Japan”.

**Discussion:** Sincere discussion based on comprehensive information through careful GC provided by CGC and clinical geneticists in our institute might have been related to the low prevalence (44%) of prenatal testing mainly consisting of amniocentesis, though experiences of infertility treatment might have had some effects on the choices.



## PgmNr 695: The Genome Empowerment Scale (GEmS): An assessment of parental empowerment in families with undiagnosed disease.

### Authors:

A. McConkie-Rosell <sup>1</sup>; K. Schoch <sup>1</sup>; J. Sullivan <sup>1</sup>; H. Cope <sup>1</sup>; R. Spillmann <sup>1</sup>; C.G.S. Palmer <sup>2</sup>; L. Pena <sup>3</sup>; Y.-H. Jiang <sup>1</sup>; N. Daniels <sup>1</sup>; N. Walley <sup>1</sup>; K.G. Tan <sup>1</sup>; S.R. Hooper <sup>4</sup>; V. Shashi <sup>1</sup>; Undiagnosed Diseases Network (NIH Common Fund, Bethesda, MD, USA)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Duke University Medical Center, Durham, NC.; 2) Department of Psychiatry and Biobehavioral Sciences, UCLA Semel Institute, 760 Westwood Plaza, Room 47-422, Los Angeles, CA 90095, USA; 3) Division of Human Genetics, Cincinnati Children's Hospital Medical Center Department of Pediatrics, University of Cincinnati School of Medicine; 4) Department of Allied Health Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

---

While genomic sequencing (exome and genome- ES and GS) has the potential to diagnose children with difficult to diagnose phenotypes, it demands a high level of parental involvement in understanding and emotionally managing information, and engaging with healthcare providers. Thus, objective measures to assess parental perspectives of sequencing are essential to optimize the process and psychosocial outcomes. We present the preliminary validity and reliability of the Genome Empowerment Scale (GEmS), developed using a theoretical model of healthcare empowerment. To evaluate its psychometric properties, we enrolled 158 parents of 117 children with an undiagnosed condition undergoing ES and/or GS. Parents completed the GEmS, measures for criterion validity and for depression and anxiety. Using Principal Components Analysis with Varimax rotation, resulted in a four factor solution: 1) Meaning of a diagnosis; 2) Emotional management of the process; 3) Seeking information and support and 4) Implications and planning. Reliability and validity analyses demonstrate that the GEmS has good psychometric properties. The Cronbach Alpha estimates ranged from 0.83-0.67. We found reasonable standard error of measurements, temporal stability, and the GEmS aligns with our targeted criterion measures providing preliminary evidence for convergent validity. The inter-relationships among the four factors revealed a profile, which has the potential to identify parents who may be at risk for a poorer outcome as well as areas that may benefit from targeted genetic counseling interventions. As an objective measure of parental healthcare empowerment for genomic sequencing, the GEmS can be utilized for future research and translational applications.



## PgmNr 696: An integrated end-to-end platform for engaged population-scale research.

**Authors:**

C. Neben; S. Topper; A. Zhou

[View Session](#) [Add to Schedule](#)

**Affiliation:** Color, Burlingame, California.

---

Human subjects research is a partnership. It cannot begin without the generosity and trust of participants, and it will not be successful without researchers' attention focused on their scientific and clinical questions.

Participants expect to be acknowledged as interested parties as the project progresses. That interest can extend to study results, their personal data, clinically relevant incidental findings in that data, and derivative sample usage. Researchers recognize a duty to consider these expectations, but considerable logistical and ethical concerns exist. These include ensuring the analytic and clinical validity of data reported to individuals, supporting communication, education, data sharing and clinical routing, capturing and honoring informed consent, and constraining costs.

Recent projects such as the *All of Us* Research Program have made great contributions to the current thinking around human subjects research and genomic return of results. Here, we describe the essential components of an integrated end-to-end platform for engaged population-scale research. We describe our process to drive recruitment and participant engagement, the logistics and laboratory infrastructure, and clinical validation and return of results for several studies, including two StandUpToCancer Dream Teams: MAGENTA and GENERATE.

The Color web-platform provides customizable study-specific landing pages, supports consent and data gathering workflows, and scheduled interactions to engage the participants over the course of the study. Integrated logistics support distributed sample collection and tracking. Our state-of-the-art CLIA/CAP next-generation sequencing laboratory provides efficient and affordable data assets, including clinical, deep whole genome sequencing (WGS), research-grade low-coverage WGS, and customized analytics. Finally, the clinical infrastructure ensures that incidental findings are confirmed, clinically interpreted using best practices, and returned to participants and their healthcare provider using clear language, and supported by Color's Genetic Counseling service.

Taken together, Color's research platform powers responsible and efficient large-scale research studies by fully supporting a clinically valid return of results component. Our unique capabilities to provide software innovations and educational opportunities ensures the success of studies in the eyes of the researchers and coordinators as well as those of the studies' participants.

## PgmNr 697: Pseudo-dominance in inherited retinal dystrophies: The need to consider common pathogenic variants in recessive disease genes as causative of multi-generational retinal dystrophy.

### Authors:

E. Place; N. Wagner; K. Bujakowska; R. Huckfeldt; J. Comander; E. Pierce

[View Session](#) [Add to Schedule](#)

**Affiliation:** Ocular Genomics Institute, Department of Ophthalmology, Massachusetts Eye & Ear Infirmary, Boston, Massachusetts.

---

**Background:** Inherited retinal dystrophies (IRDs) are genetically heterogeneous, with over 250 genes identified to cause autosomal dominant (AD), autosomal recessive (AR) and x-linked (XL) disease. For families with multi-generational retinal dystrophy, AD or XL inheritance is initially considered. When initial analysis does not identify a likely causative variant in an AD or XL gene, other genetic explanations are considered including new disease gene or a non-coding mutation in a known IRD gene. Before considering these possibilities, AR inheritance in all affected generations should be considered as multiple pathogenic alleles of the same gene may be involved. To determine the prevalence of AR mutations as causative of pseudo-dominant IRD, a cohort of patients with a known family history of retinal disease was genetically analyzed. **Methods:** Next-generation sequencing (NGS) of over 250 known IRD genes was performed for > 1000 IRD probands. Clinical diagnosis was based on a complete ophthalmic evaluation. In probands with a reported multi-generational family history of retinal disease and no pathogenic or likely pathogenic variants in known AD or XL disease genes, an AR explanation of disease was considered. Targeted NGS was performed on available family members, which included copy number analysis.

**Results:** We identified 8 pseudo-dominant families with multiple generations of AR disease. In 5 families with macular degeneration the disease was due to mutations in *ABCA4*. In multiple families, the common pathogenic mutation p.(Gly1961Glu) was identified, which has an allele frequency of 0.0046% in gnomAD (10 homozygote reported). In one family, where disease was due to mutations in *BBS1*, the common mutation p.(Met390Arg) was involved, which has a gnomAD allele frequency of 0.0016%. Finally, we identified two families in which the family history was suggestive of XL disease or AD disease with reduced penetrance, and yet causal mutations in *C2orf71* and *ABCA4* were identified. **Conclusions:** Mutations in AR genes are an important cause of IRD in families with multi-generational disease. Certain AR pathogenic mutations can be relatively frequent in the general population, thus contributing significantly to pseudo-dominant inheritance. Given the high carrier frequency of some disease-causing IRD alleles, the re-occurrence of AR may be higher than initially estimated, which needs to be considered during genetic counseling of probands with AR disease.

## PgmNr 698: Reporting of variants in genes with limited, disputed, or no evidence for a Mendelian condition among GenomeConnect participants.

### Authors:

J.M. Savatt<sup>1</sup>; D.R. Azzariti<sup>2</sup>; W.A. Faucett<sup>1</sup>; D.H. Ledbetter<sup>1</sup>; V.R. Miller<sup>3</sup>; E. Palen<sup>1</sup>; H.L. Rehm<sup>2,4,5</sup>; J. Rhode<sup>3</sup>; J.A. Vidal<sup>3</sup>; E.R. Riggs<sup>1</sup>; C.L. Martin<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Geisinger, Danville, Pennsylvania, USA; 2) The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 3) Invitae, San Francisco, California, USA; 4) Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; 5) Harvard Medical School, Boston, Massachusetts, USA

---

The rapid evolution of DNA sequencing technologies has escalated the detection of genomic variants. However, a variant or gene's relationship to disease may still be unclear, making variants of uncertain significance (VUS) commonplace. VUS results can be challenging for genetic counselors and patients alike. One aim of the NIH-funded Clinical Genome Resource (ClinGen) is to establish the clinical validity of gene-disease relationships for use in genomic medicine and genetic counseling. Per ACMG/AMP sequence interpretation guidelines, an established gene-disease relationship is necessary to determine a variant's pathogenicity, and without such, a variant should only be classified as a VUS. To date, 743 gene-disease pairs have been curated by ClinGen; a gene can be curated for more than one disease. Here, we report on clinically reported variants in gene-disease pairs with limited, disputed, refuted, or no reported evidence that were shared by participants in GenomeConnect, the ClinGen patient registry, to characterize how variants in such genes have been reported.

Of 969 variants evaluated, 1.5% (n=15) were in a gene-disease pair that ClinGen curated as limited, disputed, refuted, or no reported evidence. Of these variants, the majority were reported as a VUS (n=14) and one was reported as pathogenic. These variants were reported via disease specific panels (n=10), familial cascade testing (n=1), or exome sequencing (n=4). Variants reported via disease specific panels or familial testing were likely assumed to have a clearer disease association by the laboratory given the testing type. From exome, none of the variants detected were described as being in genes of uncertain significance, meaning the laboratory likely thought there was a disease relationship. Overall, the way these variants were reported suggests that they were classified based on the data available regarding the variant alone, and not due to the gene-disease relationship. This indicates that variants in genes curated by ClinGen as having little or no evidence may be reported as though they are disease causing. Pairing a gene's relationship to disease and the strength of that relationship impacts genomic medicine by informing multiple parts of the genetic counseling process including test selection, counseling for a variant's significance, and familial cascade testing. An understanding of how genes with limited or no evidence are being reported can inform genomic medicine moving forward.

## PgmNr 699: Delivering high-quality, scalable genetic counseling in international settings: Case studies from Trinidad and Tobago and the United Arab Emirates.

### Authors:

L. Servais <sup>1</sup>; M. Stamp <sup>1</sup>; C.L. Neben <sup>1</sup>; C. Lai <sup>1</sup>; G.T. Parkinson <sup>2</sup>; A.Y. Zhou <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Color Genomics, Burlingame, CA.; 2) Yale New Haven Hospital, New Haven, CT

---

**Background:** Genetic counseling (GC) is an important component to help individuals understand genetic testing results and their implications for their medical care. GC is often provided through in-office consultation. However, as genetic testing continues to become more affordable and accessible, the field must re-evaluate how to deliver high-quality, scalable GC across diverse populations. Color has provided telephone-based GC to over 15,000 individuals, including 1048 individuals from 51 countries outside the US. Here, we present two case studies on GC in two very different international settings: United Arab Emirates (UAE), a well-resourced, digitally savvy, country with a total GDP of \$432.612B, and Trinidad and Tobago (T&T), a less well-resourced country with a total GDP of \$22.438B.

**Methods:** The T&T cohort received genetic testing with a 30-gene NGS panel for hereditary cancer that was offered through a high-risk study. The UEA cohort received genetic testing for hereditary cancer risk as well as a 30-gene NGS panel for hereditary cardiovascular disease that was offered as an employee benefit. In both cohorts, GC was conducted by telephone, with translation services available. Color GCs also met with healthcare providers (HCPs) from T&T and UAE to better understand their healthcare systems to ensure cultural competency.

**Results:** In the T&T cohort, 141 individuals received GC, 16% had a positive result. 99% were female, and the majority were age 35-64 years. Many individuals did not have consistent access to email or a computer, and as such, consults were scheduled by telephone. Results and notes from GCs were emailed or sent via post. In the UAE cohort, 52 individuals received GC, 85% had a positive result. 42% were female, and the majority were age 25-44 years. In contrast to the T&T cohort, all individuals in the UAE cohort had access to email or a computer and were able to schedule consults and view results online. In both cases, all GC was delivered successfully.

**Conclusions:** Providing GC in an international setting requires consideration of the unique needs of each country including time zone and language, as well as cultural competency. Leveraging technology can help enable access to high-quality genetic counseling in countries with disparate levels of healthcare. Finally, building partnerships with local HCPs is crucial to ensure a warm handoff for individuals, particularly those with a positive result to receive appropriate follow-up care.

## PgmNr 700: Uptake of services among patients referred for hereditary cancer genetic counseling.

### Authors:

E.S. Gordon; C.A. Fine; B. Swope; L. Myers; S.M. Weissman; J. Bailey; A. Fan; E. Jordan; B. LeLuyer; A.F. Rope; H. Shabazz; S.B. Bleyl

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genome Medical, South San Francisco, CA

---

Recent studies have found that despite clear guidelines, patients are under referred for genetic counseling for hereditary cancer risks (Childers et al 2017). However the impact of low referral is further compounded by low uptake of genetic counseling among those who are referred. Published data suggests that in the majority of studies uptake of referrals for hereditary cancer counseling is less than 50% (7.3% to 44%) (Niendorf et al 2016, Kne et al 2017, Rahm et al 2007). Here we describe our approach to referrals, data on patient follow through and proposed opportunities for improvement. Genome Medical, a nationwide telegenetic service, contracted with a large health system to provide cancer genetic counseling services. To date, we received 407 referrals in an 11 week period. All patients were referred to genetics within the health system previously, but due to staffing shortages did not see a genetics provider. Some referrals were up to 12 months old. All patients referred to Genome Medical were contacted within one week of receiving the referral to attempt to schedule. Two different approaches were taken to this outreach. For the first six weeks patients received a phone call, followed by two emails, all 24 hours apart with 49% of referred patients scheduling. In order to try to increase follow through, the outreach approach was adapted for the following five weeks, with patient's receiving two emails followed by a phone call over a one week period. In total, 207/407 (51%) of patients referred chose to schedule an appointment and 93% of scheduled patients completed their appointment. No statistically significant difference in the uptake of appointments was observed with the change of approach. Additional work is needed to evaluate barriers and to drive further uptake of referrals. Barriers under investigation include: time since original referral for genetic services; understanding of the reason for referral or perceived utility; concerns regarding cost; completion of a genetic counseling visit elsewhere, and discomfort with telemedicine. Potential interventions to improve uptake include: decreased time between initial referral and first point of outreach; additional education about the cost of the appointment and utility of genetic counseling; and improved ease of use.

## PgmNr 701: Returning results in the genomic era: Comparison of protocols implemented in the eMERGE Network.

### Authors:

G.L. Wiesner<sup>1</sup>; K.A. Leppig<sup>2</sup>; A.K. Rahm<sup>3</sup>; P. Appelbaum<sup>4</sup>; S. Aufox<sup>5</sup>; C. Blout<sup>6</sup>; K.D. Christensen<sup>6,7</sup>; W.K. Chung<sup>4</sup>; E.W. Clayton<sup>1</sup>; R.C. Green<sup>6,7</sup>; M.H. Harr<sup>8</sup>; N. Henrikson<sup>2</sup>; C. Hoell<sup>5</sup>; I.A. Holm<sup>7,9</sup>; G.P. Jarvik<sup>10</sup>; I.J. Kullo<sup>11</sup>; P. Lammers<sup>12</sup>; E.B. Larson<sup>2</sup>; N.M. Lindor<sup>13</sup>; M. Marasa<sup>4</sup>; M.F. Myers<sup>14</sup>; C.A. Prows<sup>14</sup>; J.D. Ralston<sup>2</sup>; H.M. Rasouly<sup>4</sup>; R.R. Sharp<sup>11</sup>; M.E. Smith<sup>5</sup>; S. Van Driest<sup>1</sup>; J.L. Williams<sup>3</sup>; M.S. Williams<sup>3</sup>; J. Wynn<sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt University Medical Center, Nashville, Tennessee.; 2) Kaiser Permanente of Washington, Seattle, WA; 3) Geisinger, Danville, PA; 4) Columbia University, New York, NY; 5) Northwestern University, Chicago, IL; 6) Brigham and Women's Hospital, Boston, MA; 7) Harvard Medical School, Boston, MA; 8) Children's Hospital of Philadelphia, Philadelphia, PA; 9) Boston Children's Hospital, Boston MA; 10) University of Washington, Seattle, WA; 11) Mayo Clinic, Rochester, MN; 12) Meharry Medical College, Nashville, TN; 13) Mayo Clinic Arizona, Scottsdale, AZ; 14) Cincinnati Children's Hospital Medical Center, Cincinnati, OH

---

**Background:** Despite the rapid increase in genomic sequencing, there are few practical recommendations to guide healthcare systems (HCS) when returning unsolicited or unanticipated genomic results. One aim of the 3<sup>rd</sup> phase of the electronic Medical Record and Genomics Network (eMERGE3) is to understand the implementation of genomic medicine by sequencing over 100 genes in 25,000 participants and returning the results of actionable variants to consenting participants and their health care providers (HCP). We sought to compare the return of results (RoR) protocols at the eMERGE3 sites with the goal of developing recommended RoR guidelines for HCS.

**Methods:** Each of the 10 eMERGE3 sites independently developed processes for RoR and discussed the plans as part of the eMERGE3 RoR Workgroup. The Consolidated Framework for Implementation Research (CFIR) was used to characterize the planned RoR processes at each site and to compare these plans across three essential domains: disclosure to the participant, notification of the health care provider, and integration of results into the electronic health record (EHR).

**Results:** A high degree of variability in RoR protocols was found across the 10 institutions. 8 sites enrolled adults and 2 sites enrolled pediatric participants, and most planned on returning the pathogenic or likely pathogenic results from the genes on the eMERGE platform. Three main RoR pathways were identified, with half of sites (n=5) planning to disclose to the participant, followed by HCP notification and uploading to the EHR. Three sites planned to disclose to the participant, followed by EHR uploading and then notifying the HCP, and 2 sites uploaded to the EHR, then notified the participant and the HCP. Modifications were made in all three pathways during the initial stages of eMERGE3 to accommodate unusual or unforeseen events, including participant death, loss of contact, sample integrity, gender mismatch and variant mosaicism.

**Conclusion:** The RoR processes in the eMERGE3 network are heterogeneous and reflect the “real world” of genomic medicine in which RoR procedures are shaped by the needs of the patients and institutional environment. This study provides early empirical evidence about the structure and framework of the RoR process that can be used by HCS considering genomic medicine programs and for future studies examining the implementation of genomic data in healthcare.

# PgmNr 702: Perspective of Nigerian health care providers, researchers, and bioethicists on germline editing technology.

## Authors:

O. Obak <sup>1</sup>; E. Obak <sup>1</sup>; C. Ugorji <sup>2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) University of Calabar teaching Hospital, Calabar, Nigeria; 2) Institute of Human Virology Nigeria

## Introduction

Gene editing technology is one of the new technologies useful for genomics research and therapy. Despite the ethical and safety concerns that needs to be addressed, its rapid wide spread by scientists, drug companies and the general public enthuse overs its potential benefits. In this study, we evaluate the perspective of Nigerian health care providers, bioethicists and researchers on the use of germline editing technology.

## Methodology

We utilized the consensus opinion summarized by Professor Charis Thomson of the proceedings at the International Summit on Human Gene Editing to develop questionnaire on attitude to genome editing. After a brief introduction to the technology, we asked Nigerian scientists, health care providers, bioethicists and researchers how much they agreed with the consensus opinions using Likert scale. The questionnaires were distributed both directly and electronically (SurveyMonkey, Twitter and Facebook) to respondents. We analyzed our responses using Stata®.

## Findings

There were 85 respondents, mean age (SD) of 34.4 (8.6) years and almost half were doctors (49.4%) while others were researchers, nurses, medical laboratory scientists and others. Some 57.0% (49 of 85) had no previous knowledge of gene editing technology. Most 85.9% agreed or strongly agreed that gene editing can be used to eliminate serious genetic conditions if it is likely to be safe and effective. 59.5% of the respondents support or strongly support placing a moratorium on gene editing technology while the technical issues are resolved, 21.4% disagreed or strongly disagreed. Some 63.1% support or strongly support placing a moratorium on gene editing technology for ethical reasons while 19.1% disagreed or strongly disagreed. Some 41.2% disagreed or strongly disagreed with trusting scientists to take decisions on when to use gene editing technology while 25.9% were neutral and 32.9% were willing to trust scientists.

## Conclusion

Majority of the Nigerian health care providers, scientists and bioethicist had no awareness of gene editing technology but accepted its use after a brief education and under different circumstances. Most of them wanted a moratorium on its use until technical and ethical issues are resolved. Minority were willing to trust scientists alone to take decisions on the use gene editing technologies.



## PgmNr 703: A case of mosaic Turner syndrome - a value of examination different germ layers and combination of karyotyping and FISH.

### Authors:

J. Stellmachova <sup>1</sup>; V. Curtisova <sup>1</sup>; K. Adamova <sup>1</sup>; D. Vrbicka <sup>1</sup>; P. Vrtel <sup>1</sup>; R. Vrtel <sup>1</sup>; M. Prochazka <sup>1</sup>; E. Klaskova <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics University Hospital and Palacky University Olomouc, Czech Republic; 2) Department of Pediatrics University Hospital and Palacky University Olomouc, Czech Republic

We present a case of a prenatal detection Turner's syndrome in pregnancy with increased risk of trisomy 21 (1: 20, NT 6.9mm, maternal age 35 years). Amniocentesis was performed, fetal karyotype was 45,X [28]/46,XX [22] (56% of monosomy X). The mother decided to continue her pregnancy.

The child was born at 33 + 5 wg for pathological CTG, was hypoxic postnatally, and required 48 hours oxygen therapy. Small coarctation of aorta was detected during newborn screening, conservative approach was recommended by cardiologist. No other congenital malformations were found, further postpartum adaptations were uncomplicated.

Postnatally following investigations were performed on cells from three germ layers utilizing G banded karyotype on 50 metaphases from peripheral lymphocytes. FISH cells obtained from three different germ layers: uncultivated lymphocytes from peripheral blood (mesoderm), buccal smear (ectoderm) and smear from the root of the tongue (endoderm). FISH examination of 250 nuclei with centromeric probes (Cytocell).

### Results:

Karyotype: 45,X [2]/46,XX [48] (4% of monosomy X)

FISH native peripheral blood (mesoderm): ish (DXZ1x1)[34]/(DXZ1x2)[216] (13.6% of monosomy X)

FISH buccal smear (ectoderm): ish (DXZ1x1)[60]/(DXZ1x2)[190] (24% of monosomy X)

FISH root of the tongue (endoderm): ish (DXZ1x1)[23]/(DXZ1x2)[227] (9.2% of monosomy X)

The karyotype still represents gold standard for determination of numerical chromosomal aberrations. In case of mosaic finding the proportion of abnormal cells may differ according to the method used and to the tissue examined. In case of monosomy X the proportion of monosomic cells may appear non significant (less than 5%) but expanded testing as demonstrated above may reveal much larger proportion of monosomic cells with an outcome for genetic counselling and further management.

Early diagnosis of Turner syndrome enables timely multi-disciplinary follow-up, which includes an endocrinologist (growth and puberty monitoring, growth hormone therapy), cardiologist, psychologist and pregnancy planning.

**Supported by MH CZ - DRO (FNOI, 00098892)**

## PgmNr 704: Daytime napping is an addictive behavior lowering general health rating for middle-aged and old adults: A pan-phenotype mendelian randomization study based on UK Biobank.

### Authors:

L. Chen <sup>1</sup>; P. Chen <sup>2,3,5</sup>; M. Zhang <sup>3</sup>; Y. Wu <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) College of Clinical Medicine, Jilin University, Changchun, Jilin Province, China; 2) The Key Laboratory of Pathobiology, Ministry of Education, College of Basic Medical Sciences, Jilin University, Changchun, Jilin Province, China; 3) Department of Genetics, College of Basic Medical Sciences, Jilin University, Changchun, Jilin Province, China; 4) University Archives, Jilin University, Changchun, Jilin Province, China; 5) Department of Pathology, College of Basic Medical Sciences, Jilin University, Changchun, Jilin Province, China

**Purpose:** As a common practice, daytime napping has been taken worldwide. However, it is still controversial as to whether daytime napping is beneficial for our health since some studies unveil its function in improving memory and cognitive abilities while others demonstrate its deleterious effects on metabolism. Thus, We have performed a pan-phenotype mendelian randomization analysis to detect all the potential phenotypes that are in a causal relationship with daytime napping and finally reveal the important role daytime napping play in health.

**Methods:** Mendelian randomization, a randomized trial based on Mendel's law of inheritance, was utilized to grope for the potential causal links. The data source came from UK Biobank repository, whose participants aged from 40~69 years old. A high throughput method was developed to deal with it. In addition, a false discovery rate was employed to add power to our multiple statistical tests.

**Results:** Our findings have unveiled the causal relationship between daytime napping and other five categories of phenotypes including metabolism, psychiatry/pain, anthropometry, lifestyle and blood count among adults. The daytime napping can raise the both systolic ( $\beta=0.207$ ,  $Se=0.053$ , IVW P value= $9.98 \times 10^{-5}$ ,  $FDR=2.89 \times 10^{-2}$ ) and diastolic ( $\beta=0.226$ ,  $Se=0.056$ , IVW P value= $4.67 \times 10^{-5}$ ,  $FDR=1.95 \times 10^{-2}$ ) blood pressure, along with elevating the body fat percentage ( $\beta=0.234$ ,  $Se=0.054$ , IVW P value= $1.37 \times 10^{-5}$ ,  $FDR=1.11 \times 10^{-2}$ ) including arms, legs and trunk. In addition, increased daytime napping frequency can make people feel more tired and lethargic ( $\beta=0.266$ ,  $Se=0.047$ , IVW P value= $1.32 \times 10^{-8}$ ,  $FDR=2.99 \times 10^{-5}$ ) as well as decrease the daily activity, for example lowering usually walking pace ( $\beta=-0.157$ ,  $Se=0.027$ , IVW P value= $4.22 \times 10^{-9}$ ,  $FDR=1.19 \times 10^{-5}$ ). Moreover, there exists direct evidence that increased daytime napping frequency can lower overall health rating ( $\beta=0.156$ ,  $Se=0.041$ , IVW P value= $1.21 \times 10^{-4}$ ,  $FDR=3.13 \times 10^{-2}$ ), a phenotype in UK Biobank.

**Conclusion:** Generally, daytime napping can be denoted as a kind of addictive behavior for adults since it can trap nappers into a vicious circle, where an increased daytime napping frequency can cause increased tiredness and lethargy, resulting in more daytime napping in turn. Moreover, such addictive behavior has a detriment effect on metabolism, potentially leading to cardiovascular diseases.

## PgmNr 705: Establishing molecular medicine laboratories and generating population specific DNA sequence data in the developing country.

### Authors:

M.C. Ergoren <sup>1,2</sup>; G. Mocan <sup>1,3</sup>; S.G. Temel <sup>4</sup>; M. Dundar <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Biology, Faculty of Medicine, Near East University, Nicosia, Cyprus; 2) Research Centre of Experimental Health Sciences, Near East University, Nicosia Cyprus; 3) Medical Pathology, Faculty of Medicine, Near East University, Nicosia, Cyprus; 4) Medical Genetics, Faculty of Medicine, Bursa Uludag University, Bursa, Turkey; 5) Medical Genetics, Faculty of Medicine, Erciyes University, Kayseri, Turkey

---

The first human genome sequence was established at the beginning of the third millennium. Since then advanced life sciences has been shaping economy, societies, bodies and mind. Globally, biodiversity is concentrated around the equator, but the scientific institutions generating DNA sequence data tend to be clustered in developed countries. A lack of funding and insufficient scientific knowledge of the governments, the financial, technical and logistical difficulties are major hurdles to establish medical laboratories and producing DNA sequence data in the developing countries.

Here, we introduce our case for North Cyprus in illustrating the problems of adjusting to the 'new molecular bioscience', since it is an example of a developing country that is striving to progress from an agrarian to an educational and research technology development society.

Cyprus, which is the third largest island in the Mediterranean Sea, has been at the crossroads of multiple civilizations throughout human history. Today, Turkish Cypriots are located in the *de facto* state that comprises the northeastern portion of the island of Cyprus. North Cyprus is recognized only by Turkey. Due to its lack of recognition, ~350.000 population of Northern Cyprus is heavily dependent on Turkey for economic, political and military support.

Despite being unrecognized, Near East University greatly contributes the economy and its recognition by placing thousand of international students in North Cyprus. Recently, the university have been concentrating on research and development. Since 2015, molecular medicine laboratories have been established as well as molecular medicine postgraduate program to generate scientists which having many international students in the program. Currently, basic molecular techniques are quite well established.

In the laboratory, novel disease-causing mutations were identified and many cases were unsolved. Lack of genetic databases in the region, the laboratory set up the capacity to produce DNA sequence data expands the number of contributors to DNA sequence databases, which is beneficial for science generally and may be helpful for any number of bioinformatic analyses.

Our data is not only helping us to contribute population specific allele frequencies of variants, it also

strengthens scientific institutions in the country. In conclusion, our establishment and success raise awareness of the North Cyprus government and address the importance of education and research.

## PgmNr 706: Scalable IRB framework for returning genetic/genomic test results to research patients in community healthcare settings.

### Authors:

E. Kurnat-Thoma <sup>1</sup>; C. Allen <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Symptom Science Center Symptom Management Branch NIH/NINR 3 Center Drive Bethesda, MD 20892; 2) Intramural Research Program NIH/NINR Building 10 Clinical Center Bethesda, MD 20814

---

**Introduction/Background:** In the era of transparency, big data, and expansion of genetics/genomics (G/G) technologies into learning healthcare systems—communication of individual findings to patients who donate biospecimens for clinical research is an emerging area of health policy focus.

**Problem Statement:** Non-CLIA laboratory results, tests with uncertain value, tests lacking sufficient analytic and clinical validity, clinical utility, and regulatory compliance represent multiple competing stakeholder priorities for Institutional Review Boards (IRBs) when assessing plans for the return of G/G research test information to patients. Recent initiatives such as the NIH Precision Medicine Initiative *All of Us* and the NHGRI's electronic Medical Records and Genomics (eMERGE) Network provide exemplars of delivering individual G/G results to patients within the context of Institutional Review Board (IRB) oversight and funding mechanisms affiliated with major tertiary care facilities. However ethical, legal social and policy (ELSI) implications are presented when expanding similar efforts into community health care hospitals and systems which lack comparable institutional infrastructure capable of supporting complex laboratory, regulatory, clinical research, data management, and governance oversight requirements.

**Methods:** A focused review of the literature concerning the return of individual G/G research results was conducted. Available IRB translation resources and interventions for community hospital network settings were identified. A simplified evaluation and implementation framework specific to the level of IRB policy, IRB protocol review, and return of individual G/G research results was conceptualized.

**Conclusion:** Critical logistical elements of IRB oversight for genomic medicine research protocols that involve individual return of G/G research tests in community hospital and network settings need to be simplified to be compatible with available resources.

**Recommendation:** As Precision Medicine translation initiatives grow beyond academic medical centers and diffuse into more limited-resource community health care settings, efforts should be made to develop scalable IRB implementation frameworks for smaller hospitals and health care systems that are not excessively burdensome.

## PgmNr 707: Population-based genetic screening: Does diagnostic information enter into the medical record?

### Authors:

G. Elhanan <sup>1,2</sup>; W.J. Metcalf <sup>1,2</sup>; H. Reed <sup>1,2</sup>; B. Lipp <sup>1,2</sup>; E. Smith <sup>1,2</sup>; N. Washington <sup>3</sup>; K. Dunaway <sup>3</sup>; A. Boltz <sup>3</sup>; C. Rowan <sup>1</sup>; N. Slotnick <sup>1</sup>; J. Lu <sup>3</sup>; J.J. Grymski <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Renown Institute of Health Innovation, Reno, NV; 2) Desert Research Institute, Reno, NV; 3) Helix Opco, LLC, San Carlos, CA

---

Population-based genetic screening can be a valuable screening tool for relatively common genetic risk conditions such as hereditary breast and ovarian cancer (HBOC). However, underlying the assumption that screening the general population is effective is the premise that the vast majority of screened individuals will take consistent and persistent recommended actions regarding positive findings, starting with reporting such findings to their healthcare providers.

We report here our early findings on the follow-up of genetic carrier status into the medical records of population-based genetic screening participants. As part of the Healthy Nevada Project (HNP) in Northern Nevada (Renown Health, Reno), 23,713 unselected participants underwent independent clinical exome sequencing from March 15 to Sept 30, 2018 (135 HBOC, 46 Lynch syndrome (LS), and 117 familial hypercholesterolemia (FH) carriers). Overall, 243 of the carriers had populated medical records with Renown and for 177 (78 HBOC, 33 LS, 66 FH) of carriers, genetic findings were delivered by qualified Renown physicians or outside genetic counselors. Medical record diagnoses were retrieved at a minimum period of three months after delivery of findings. Twenty-one of HBOC carriers had an appropriate diagnosis code (ICD9/10) documenting the genetic finding in their medical record and for 13 (18.6%), HNP testing was the precipitator of the documentation. Five of LS carriers had an appropriate ICD code, two (6.7%) of which were contributed due to HNP testing. Twelve of the FH patients had an appropriate ICD code, 10 of which were in response to HNP testing (15.6%).

Our results underscore the potential of population-based genetic screening for broader identifications of individuals at risk and likely under-represent the rate of provider notification by participants. However, the setting of the HNP in northern Nevada is unique, with Renown being the dominant primary care provider. Therefore, the low rate of diagnostic documentation in Renown's medical records may have implications regarding the effectiveness of voluntary, independent population-based genetic screening, even for life threatening conditions. We plan to review these findings on a periodic basis and survey HNP participants for the actual rate and reasons for non-reporting of genetic findings. Such feedback is essential to maximize the value of independent population-based genetic screening as a preventive tool.

## PgmNr 708: The potential economic burden of genetics on healthcare.

### Authors:

S. Hebring<sup>1</sup>; J. Mayer<sup>1</sup>; L. Bastarache<sup>2</sup>; S. Bland<sup>2</sup>; C. Weng<sup>3</sup>; H. Hakonarson<sup>4</sup>; G. Hripacsak<sup>3</sup>; E. Larson<sup>5</sup>; J. Denny<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Marshfield Clinic, Marshfield, Wisconsin.; 2) Vanderbilt University, Nashville, Tennessee; 3) Columbia University, New York City, New York; 4) Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; 5) Kaiser Permanente Washington, Seattle, Washington

---

Everyone carries genetic variants that influence disease risk, but it is unknown how these variants influence healthcare costs at a population level. To address this, we conducted a genome-wide association study (GWAS) on healthcare utilization with further focus on common and rare variants that influence human traits. Because actual healthcare costs may differ by who is paying, what is actually recovered, and can vary dramatically by service and institution, we calculated a Medicare Risk Adjustment Model (MRAM) score as a proxy. MRAM is a standardized model developed by Centers for Medicare & Medicaid Services designed to predict future costs and is used to set premiums for Medicare insureds. MRAM scores were calculated for nearly 20,000 Medicare eligible individuals across eight separate healthcare systems that participated in the eMERGE Network. Input variables for MRAM scores include International Classifications of Disease Codes, age, and sex.

Using a linear mixed model from GCTA, approximately 3.8% of the predicted variance for MRAM scores was explained by thousands of common variants known to be associated with human traits according to the GWAS Catalog. This is compared to 14.3% of the predicted variance explained by the genome. No single common variant reached genome-wide significance. Conversely, multiple statistically significant associations passed a conservative experiment wise Bonferroni correction when focusing on known trait specific variants. For example, when evaluating 263 rare non-monomorphic variants associated with predominantly Mendelian disorders, a genotype consistent with alpha-1 antitrypsin deficiency (AATD) was associated with healthcare utilization ( $P=1.89e-4$ ). Those diagnosed with AATD often had costly liver and lung ailments (e.g., liver transplants) based on manual review of medical records. Moreover, when generating weighted polygenic risk scores (PRS) for 735 common traits, education status ( $P=1.76e-5$ ) was significantly associated with MRAM scores. Several other phenotypes approached statistical significance including type II diabetes ( $P=9.58e-5$ ) and obesity ( $P=1.31e-4$ ).

In conclusion, these results emphasize that both common and rare variants contribute to healthcare utilization at the population level. Importantly, these results also underscore the importance of maintaining and strengthening legislation, such as the Genetic Information Nondiscrimination Act, to protect all since everyone carries disease risk alleles.

## PgmNr 709: Impact of non-clinical genetic trait insights on clinical engagement in participants after clinical genetic testing.

### Authors:

J. Van Den Akker; A. Boateng-Kuffour; L. Servais; L.E. Jones; C.L. Neben; L. Ryan; A.Y. Zhou

[View Session](#) [Add to Schedule](#)

**Affiliation:** Color Genomics, Burlingame, California.

---

**Background:** Patient recall, or a patient's ability to recollect pertinent medical information, is key to an effective patient-provider communication and adherence to prevention and treatment plans. However, multiple studies report that up to 40-80% of patients are unable to recollect information from their medical encounters. To improve patient engagement with their genetic results, Color offers 'Discovery', a platform that engages clients and provides them with non-clinical genetic insights such as ancestry, bitter taste perception, lactose intolerance, and alcohol flush response. In this study, we sought to assess the impact of Discovery on engagement with clinical results after genetic testing.

**Methods:** We analyzed de-identified data from a consecutive cohort of 12,410 Color clients who consented to participate in research. Clients received a clinical report from Color for genetic risk associated with hereditary cancer, hereditary cardiovascular disease, or medication response. Engagement metrics included the following actions: viewing a clinical report for the first time, re-reviewing an existing clinical report, downloading a clinical report(s), and sharing a clinical report(s) with their healthcare provider.

**Results:** 25.3% of clients in the cohort opted to participate in Discovery. Of these, 91.8% viewed at least one and 70.3% viewed at least six non-clinical genetic trait insights. Within 30 minutes of viewing their trait insights, 28% of Discovery participants re-reviewed at least one clinical report for hereditary cancer, hereditary cardiovascular disease, or medication response. 98.0% of Discovery participants viewed their clinical report(s) within 30 days after their release, as compared to 93.6% of non-Discovery participants. Discovery participants were also more likely to re-review their clinical report(s) after their initial view (70.6 vs 47.9%), share their clinical report(s) with their healthcare provider (8.9% vs 5.0%), and download their clinical report(s) (21.1 vs 16.2%).

**Conclusion:** Color Discovery is an effective engagement tool that prompts participants to take clinical actions such as reviewing, downloading, and sharing their clinical reports with their healthcare provider. Overall, Discovery participants took more clinical actions than non-Discovery participants. Our results demonstrate that this engagement platform may be an effective way to engage clients to drive better medical information recall and ultimately, adherence.



## PgmNr 710: Willingness to pay for whole genome sequencing after testing.

### Authors:

N. Aggarwal <sup>1</sup>; K. Christensen <sup>2</sup>; J. Robinson <sup>3</sup>; C. Blout <sup>2,4</sup>; R. Green <sup>2,5</sup>; A. McGuire <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical College of Georgia, Augusta, GA; 2) Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Baylor College of Medicine, Houston, TX; 4) Massachusetts General Hospital, Boston, MA; 5) Partners Center for Personalized Genetic Medicine, Broad Institute of MIT & Harvard, Boston, MA

---

**Background:** With increasing availability of whole genome sequencing (WGS), it is crucial to examine its economic impact in ways that consider the personal utility it may provide. Willingness to pay (WTP) is a common approach to measure personal utility, but its use to assess genomic testing raises concerns about whether valid data can be collected from people who are unfamiliar with the technology. To address this, we collected WTP data from patients who did or did not receive WGS in randomized trial.

**Methods:** The MedSeq Project was a randomized clinical trial of WGS. Adults from primary care clinics (mean age 54, 58% female) and from cardiology clinics with cardiomyopathy diagnoses (mean age 55, 43% female) were randomized to review family histories alone or with WGS. Prior to and 6 months after receiving reports, participants stated the maximum amount they are willing to pay for WGS, as well as their likelihood to pursue WGS via insurance- or self-payment on 4-point scales. Analyses determined WTP changes over time and variation by experience.

**Results:** At 6 months, the mean likelihood to pursue WGS if insurance covered payment was 3.5 on a 4-point scale, and ratings did not differ between participants who did and did not receive WGS ( $p=0.72$ ) or change from baseline ( $p=0.41$ ). The mean likelihood to pursue WGS through self-payment was 2.3 on a 4-point scale, lower than the likelihood if insurance paid ( $p<0.001$ ). Ratings for self-pay did not differ between participants who did and did not receive WGS ( $p=0.73$ ) or change from baseline ( $p=0.26$ ). Six months after receiving care, WTP decreased an average of \$532 from baseline ( $p<0.001$ ). Changes in WTP did not vary between those who did and did not receive WGS ( $p=0.41$ ). Analyses restricted to participants who received WGS showed no associations between identification of monogenic disease risks or polygenic risk predictions and changes in WTP or likelihood of pursuing testing.

**Conclusions:** Participants were willing to pay less for WGS after participating in the MedSeq Project, although changes did not appear to be affected by the experience of WGS. Participants rated their likelihood to pursue WGS the same at the end of the study as at enrollment, even if they had to self-pay for testing. Findings suggest that researchers can collect valid assessments of WTP even if individuals do not receive WGS. Future work will explore reasons why WTP for WGS decreased.

## PgmNr 711: Parental perspectives on return of secondary findings from a research biobank.

### Authors:

J.J. Connolly <sup>1</sup>; H.S. Hain <sup>1</sup>; M.H. Harr <sup>1</sup>; H. Hakonarson <sup>1,2</sup>; Electronic Medical Records and Genomics (eMERGE) Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

This study examined parental perspectives across several dimensions and how they may differ by sociodemographic factors in a targeted genetic testing, research setting. The aim of the study was to determine differences and associations in perspectives in a baseline survey, prior to return of results from a targeted sequencing panel, enriched for genes on the American College of Medical Genetics incidental findings panel. These results are intended to be used to promote health utilization for genetic testing and other healthcare services centered on the child and their family.

Participants completed an online self-reported survey focused on challenges faced in and the impact on participants of return of results, specifically attitudes to testing, family sharing, decision conflict, feelings about genomic results, information seeking, privacy, as well as a range of health/psychosocial outcomes. Sociodemographic parameters were collected as well for comparison purposes.

Ancestry affected responses for attitudes toward genetic testing, family sharing and information seeking. The health status of participants and/or their children affected responses for attitudes toward genetic testing, information seeking, privacy concerns and literacy. Information seeking was different in groups based on sex and religion as well. Attitudes toward genetic testing were more favorable in participants with a lower education level. Learning written medical information differed by religiosity, income, education, self-reported health and insurance. Confidence about filling out medical forms varied by income level and presence of a genetic disorder in participants or their children.

Overall, participant attitudes were favorable towards research genetic testing. Sociodemographic differences observed in literacy, privacy, and information seeking survey questions could guide educational materials and practices used to personalize genetic counseling sessions.

## PgmNr 712: Exploring women's experiences of non-invasive prenatal testing in a private health care setting in the Western Cape, South Africa.

### Authors:

K. Barrow <sup>1</sup>; T. Wessels <sup>1</sup>; K. Fieggen <sup>1</sup>; M. Schoeman <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Human Genetics, Department of Pathology, Faculty of Health Sciences, University of Cape Town, South Africa; 2) Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University and Tygerberg Hospital, Cape Town, South Africa

---

Non-invasive prenatal testing (NIPT) is the most accurate prenatal screening test that is used to screen for common chromosomal anomalies including Down syndrome (DS). NIPT can be done from 8-10 weeks gestation with a simple blood test which makes use of cell free DNA in maternal plasma. It is recommended that all women undergoing NIPT receive adequate pre-test counselling. Currently in South Africa, NIPT is only available in the private health care setting. Very little qualitative research has been performed to understand the experience and psychological impact of NIPT. This study aims to explore the experiences of women who have undergone NIPT in a private health care setting in Cape Town, South Africa.

This study used a qualitative exploratory design, which allowed for investigation of the topic. The sample population included women who had NIPT in a private health care setting in Cape Town. Participants were recruited through private practice genetic counsellors (GCs). Semi-structured interviews were conducted to collect the research data. Interviews were recorded and transcribed verbatim and the data was analysed using thematic analysis.

A total of 10 participants were interviewed, all received low risk NIPT results. Five themes were identified in the data namely, 'It's not just a DS test', 'Emotional rollercoaster', 'Making the choice', 'Support' and 'Genetic counselling'. There are medical and emotional consequences of NIPT which may be more complicated than traditional prenatal screening. Women were anxious to learn of their 'increased risk' for chromosomal abnormalities, particularly those other than DS. The early timing of NIPT and information it could provide were factors valued by all women in decision-making. Participants' initial anxiety was followed by reassurance. Women equated this reassurance with the reliability of diagnostic testing and often extended it to the overall health of the fetus. Sex chromosome testing was poorly understood. Most women did not initially disclose their increased risk with family and friends due to the tentative nature of the pregnancy. Women all had some external support but also relied on their own resilience. As mothers, they felt responsible for the health of the fetus. Genetic counselling helped contextualise risk, disseminate information and provided emotional support.

This data provides healthcare professionals with some insight into how they can implement a better NIPT service.

## PgmNr 713: Public involvement in global genomics research.

**Authors:**

J.S. Nunn <sup>1,2</sup>; S. Kent <sup>1</sup>; V. Sundararajan <sup>1</sup>; J. Tiller <sup>2</sup>; P. Peter Fransquet <sup>2</sup>; P. Lacaze <sup>2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Public Health, La Trobe University, Melbourne, Australia, Australia; 2) Public Health Genomics, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia

---

Public involvement in research occurs when the public, patients, or research participants are actively contributing to the research process. Public involvement has been acknowledged as a key priority for prominent human genomics research initiatives in many different countries. However, to date, there has been no detailed analysis or review of the features, methods, and impacts of public involvement occurring in human genomics research projects worldwide. We reviewed the reported public involvement in 96 human genomics projects (initiatives), based on a database of initiatives hosted by the Global Alliance for Genomics and Health, according to information reported on public domain websites. To conduct the scoping review, we applied a structured categorization of criteria to all information extracted from the search.

We found that only a third of all initiatives reported public involvement in any capacity (32/96, 33%). In those reporting public involvement, we found considerable variation in both the methods and tasks of involvement. Some noteworthy initiatives reported diverse and comprehensive ways of involving the public, occurring through different stages of the research project cycle. Our findings suggest there would be intrinsic value in having more public involvement occur in human genomics research worldwide. We also suggest that more systematic ways of reporting and evaluating involvement would be highly beneficial, to help develop best practices.

We are now working to apply these findings to four prominent genomics research projects in Australia, to co-design their studies and explore people's views about involvement. These projects include a large cohort study of >15,000 healthy, elderly research participants, the Australian Indigenous Precision Medicine project, a group of patients and families affected by a rare immunological disorder, and an extended family of donor-siblings who share the same sperm-donor father. We have assessed the impacts of involving people at various stages of each project, from study design, to recruitment to ethical oversight and governance. We have created a standardised reporting framework to evaluate impacts across these studies.

## **PgmNr 714: Preparation for a child with a genetic condition: What is gained from prenatal diagnosis when parents choose not to terminate?**

### **Authors:**

M. Ford <sup>1</sup>; S. Ossler <sup>2</sup>; L. Cohen <sup>3</sup>; M. Michie <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Case Western Reserve University, Cleveland, OH; 2) DDC Clinic for Special Needs Children, Middlefield, OH; 3) Sema4 Genomics, Stamford, CT

---

Some critics of prenatal genetic testing have argued it is primarily used to enable elective terminations. Yet many women and families say they choose prenatal testing for a very different reason: “preparation.” Although literature on prenatal testing often cites preparation as a reason genetic tests are offered or accepted, this concept is not well researched, and no clear definitions or recommendations exist for prenatal preparation. Here we compare “preparation” as it is used in clinical literature with a qualitative study of parental experiences after prenatal diagnosis of aneuploidy. Semi-structured phone interviews explored the experiences of 29 participants who received a prenatal diagnosis of trisomy 18 (n=14) or trisomy 21 (n=15) three or fewer years ago, recruited via national patient advocacy groups. Data were analyzed using an approach based on grounded theory and constant comparative analysis. Our extensive literature review found that preparation has been used to refer to at least 3 overlapping ways in which prenatal information may change parents’ approach to the rest of the pregnancy: (1) clinical activities during or immediately following pregnancy; (2) social and informational support, including interactions with support groups; and (3) psychological adjustment to a revised vision for the future child and family. In interviews, parents discussed many of the same themes, but characterized and prioritized them somewhat differently. Practical activities reported during pregnancy did include clinical decision-making, but also included modifying pregnancy and delivery plans to accommodate personal and family needs, especially when fetal or newborn death was likely. Connecting with other parents with similar experiences took priority for most interviewees, and a lack of timely referral to such connections was one of the most frequent disappointments the interviewees voiced about their clinical providers. Finally, psychological preparation included not only individual re-visioning of the future, but also carefully planned communication strategies with friends, extended family, and other children. In assessing clinical providers, parents prioritized support for values-based decisions throughout pregnancy over clinical knowledge. These findings provide insight into the experiences of families, suggest actions providers can take to improve patient care, and inform ethical conversations about the translation and scope of prenatal genetic tests.

## PgmNr 715: Implementation of exome sequencing in a diverse pediatric population: Experiences from UCSF.

### Authors:

H. Hoban <sup>1</sup>; S. Rego <sup>1</sup>; T. Yip <sup>1</sup>; S. Ackerman <sup>2</sup>; B. Anguiano <sup>1</sup>; F. Chen <sup>1</sup>; N. Sahin Hodoglugil <sup>1</sup>; B. Koenig <sup>3</sup>; P.-Y. Kwok <sup>4</sup>; M. Norton <sup>5</sup>; N. Risch <sup>1</sup>; A. Slavotinek <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of California, San Francisco (UCSF), Institute for Human Genetics; 2) UCSF Department of Social and Behavioral Sciences; 3) UCSF Institute for Health and Aging; 4) UCSF Cardiovascular Research Institute; 5) UCSF Department of Obstetrics, Gynecology and the Reproductive Sciences

---

The Program in Prenatal and Pediatric Genomic Sequencing (P3EGS) at UCSF is one of six sites in the NIH-funded Clinical Sequencing Evidence-Generating Research (CSER) Consortium studying the utility of integrating genomic sequencing into the clinical care of diverse and medically underserved (MU) patients. The pediatric arm of P3EGS is currently enrolling patients presenting with intellectual disability, metabolic disease, epilepsy, or multiple congenital anomalies.

Since recruitment began in October 2017, we have enrolled 217 children and their biological parents from three hospitals in San Francisco and Oakland, California, that serve diverse and MU communities. 87% of participants have MediCal (CA's Medicaid program) and about one third live in a zip code classified as a MU area. 53% of parents self-identify as Hispanic/Latino(a); 13% as white; 12% as Asian; 8% as Middle Eastern or North African; 5% as Black or African American; 2% as other groups; and 7% as more than one race/ethnicity category. The preferred language of almost half (47%) of enrolled families was a language other than English. Informed consent and results have been communicated via medical interpreters in Spanish, Farsi/Dari, Arabic, Punjabi, Mam and others. Exome results have been returned to 134 families with 37 definitive/probable positive results (28%). 176/217 (81%) of families requested return of secondary findings and 3/176 (1.7%) received positive secondary findings.

During the implementation of P3EGS we addressed several challenges that arose in the process of conducting research with diverse and MU populations. One eligibility criterion is a non-diagnostic microarray, which may not be covered by insurance (especially MediCal), and can require multiple clinic visits. We established a process to obtain pre-authorization for microarrays in order to minimize visits and reduce burden on families. We also adapted our consent form to accommodate lower literacy and non-English speaking populations with a focus on concise, simple language around complex data sharing policies. Lastly, we incorporated flexible survey administration procedures that offer families a choice of completing surveys in-person, through mail or e-mail, thereby minimizing the burden on participants. While offering exome sequencing to MU populations presents unique challenges, our preliminary results show that P3EGS is effectively enrolling and addressing the needs of a wide range of patients and their families.

## PgmNr 716: Methods for identifying health care attributable to unexpected genomic findings.

### Authors:

Z. Mackay <sup>1</sup>; D. Dukhovny <sup>7</sup>; R. Parad <sup>8</sup>; K. Phillips <sup>3,4</sup>; A. Beggs <sup>9</sup>; R. Green <sup>2,5,6</sup>; K. Christensen <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Boston University School of Medicine, Boston, MA, USA; 2) Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 3) Department of Clinical Pharmacy, Center for Translational and Policy Research on Personalized Medicine (TRANSPERS), University of California San Francisco, San Francisco, CA, USA; 4) Philip R. Lee Institute for Health Policy and Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA, USA; 5) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 6) Partners HealthCare Personalized Medicine, Boston, MA, USA; 7) Department of Pediatrics, Oregon Health & Science University, Portland, OR, USA; 8) Department of Pediatric Newborn Medicine, Brigham and Women's Hospital, Boston, MA, USA; 9) The Manton Center for Orphan Disease Research, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA

---

Existing recommendations for clinical genomic sequencing include screening for secondary findings that may be unrelated to test indications, and many advocates envision a future where results are used routinely not only to screen for genetic risk factors, but also to improve medication choices, inform reproductive decisions, etc. The variety of health-relevant information generated by genomic sequencing creates challenges to tracking health care utilization and assessing the full range of clinical benefits that genomic sequencing can provide. In the BabySeq Project, a multi-site randomized clinical trial, we implemented six different approaches to identify health care services that were motivated by newborn genomic sequencing findings over a 10-month period, including (1) reviews of medical records based on recommendations from a genetics specialist during disclosure sessions; (2) reviews of medical records based on risks identified on disclosure reports; interviews of (3) parents and (4) providers; surveys of (5) parents and (6) providers. Preliminary results show that parent interviews were most comprehensive for capturing common services, including 13 of 15 specialty visits (87%) and 21 of 21 cardiac screening tests (100%) for the newborn. Parent interviews were less effective than other approaches for identifying infrequent services such as follow-up genetic tests and lab tests of newborns and other family members. Findings highlight the challenges of assessing downstream medical services that result from unexpected genomic findings. They also demonstrate the benefits of using multiple approaches to assess follow-up care that was received not only by the patient, but by other family members. We discuss these benefits against the conceptual strengths and limitations of each approach, including their feasibility to implement, their ability to track follow-up care that occurs outside of study sites, the ability to account for loss to follow-up and incomplete or biased recall. We also discuss the ability of each approach to capture services that may have been 'avoided' by enhancing prevention.



## PgmNr 717: Effective recruitment and retention of diverse populations in NYCKidSeq.

### Authors:

J.E. Rodriguez <sup>1,2</sup>; J. Lopez <sup>5</sup>; E. Maria <sup>5</sup>; N.M. Yelton <sup>1,2</sup>; G. Bertier <sup>1,2</sup>; N. Kelly <sup>5</sup>; M. Ramos <sup>6</sup>; J.A. Odgis <sup>1,2</sup>; K. Donohue <sup>1,2</sup>; K. Gallagher <sup>5</sup>; M. Sebastin <sup>5</sup>; N. Ioele <sup>5</sup>; K. Brown <sup>5</sup>; S.A. Suckiel <sup>1,2</sup>; M. Robinson <sup>9</sup>; G.A. Diaz <sup>7</sup>; N.S. Abul-Husn <sup>1,2</sup>; M.P. Wasserstein <sup>5</sup>; B.D. Gelb <sup>8</sup>; C.R. Horowitz <sup>6</sup>; E.E. Kenny <sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.; 2) Center for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY.; 3) The Center for Statistical Genetics, Icahn School of Medicine at Mount Sinai, New York, NY.; 4) The Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY.; 5) Montefiore Medical Center/Albert Einstein College of Medicine, Bronx, NY.; 6) Center for Health Equity and Community Engaged Research, Icahn School of Medicine at Mount Sinai, New York, NY.; 7) Department of Genetics and Genomic Sciences, Icahn School of Medicine, New York, NY.; 8) Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY.; 9) Bethel Gospel Assembly, New York, NY.

Genomic research heavily skews toward participants of European descent, and there is a gap in understanding the impact of genetic testing in underrepresented and diverse populations. Here, we outline effective recruitment and retention strategies of diverse populations into NYCKidSeq, a clinical trial offering whole genome sequencing (WGS) to 1,100 children and young adults with undiagnosed neurologic, cardiovascular disorders or primary immunodeficiencies. The goal is to compare WGS to standard targeted gene panels for diagnosis, and to test participants' understanding of genetic test results using a bi-lingual health app. The patient population is from predominantly low-income and minority communities from Harlem and the Bronx. As these families are particularly vulnerable, recruiting them warrants the design of culturally sensitive, ethically sound recruitment strategies.

Researchers worked with patients, advocates and clinicians as part of a genomic stakeholder board to develop approaches to recruitment and retention and they reviewed all study materials. All study coordinators are bilingual, and a committee of individuals from diverse Latin nations translated study materials into Spanish, including recruitment and educational materials and study surveys. We also continuously engage providers and clinical staff. Our team put in place flexible solutions to accommodate the needs of families who live far from the hospital, are unavailable during work hours, or in which parents are separated. We also use continuous engagement and retention strategies such as mailing customized holiday and birthday cards, and making regular phone-calls throughout the duration of the trial.

To date, we have recruited 63 patients, and their families at 11 clinics throughout Manhattan and the Bronx: 17% Black and 60% are Hispanic/Latinx (half of whom completed the baseline survey in Spanish). Their median household income range was \$25K to \$50K, and less than 60% of parents completed a bachelor's degree. Common reasons among 10 parents who refused to participate were lack of time, concerns with life insurance, or belief that the test is unnecessary.

We have developed, are implementing, and will test the impact of strategies developed in partnership with patients, advocates and clinicians, to enhance recruitment of underrepresented, minority populations into genomics research. These strategies may help mitigate the risk of propagating health



disparities among the US population.

# PgmNr 718: Effect of genetics clinical decision support tools on healthcare providers' decision making: A systematic review.

## Authors:

A. Sebastian <sup>1,2</sup>; J. Carroll <sup>1,3</sup>; L. Oldfield <sup>4</sup>; E. Uleryk <sup>5</sup>; Y. Bombard <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) University of Toronto, Toronto, Ontario, Canada; 2) St. Michael's Hospital, Li Ka Shing Knowledge Institute, Toronto, Ontario, Canada; 3) Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada; 5) EU Consulting, Toronto, Ontario, Canada

**Background:** Deciding when to order genetic tests, refer to genetics specialists, or change treatment or surveillance based on genetic risk, is particularly challenging for non-genetics healthcare providers. Clinical decision support (CDS) tools have been suggested as a potential solution because they provide patient-specific risk assessment or management recommendations. This systematic review synthesized evidence on whether using CDS tools affected appropriate changes in genetics-related patient management by non-genetics healthcare providers, compared to standard clinical care without using CDS.

**Methods:** A comprehensive search in Medline, Embase and CINAHL from the year of database inception to March 6 2019 yielded 2254 articles, plus an additional 11 articles from handsearching. Two independent reviewers screened abstracts and full texts for articles that quantitatively compared management outcomes when non-genetics clinicians (clinicians other than medical geneticists, laboratory geneticists or genetic counsellors) used a CDS tool in actual patient care. 9 articles (3 RCTs, 5 pre-post studies and 1 cohort study) were included in the final review. Effect sizes were calculated, and a narrative synthesis was performed for the 9 articles. Study-level bias was assessed using the Joanna-Briggs Institute Checklist appropriate for study design. High risk of bias did not lead to study exclusion.

**Results:** CDS tools focused mainly on pharmacogenetics or cancer, with two tools on thrombophilia and one on cystic fibrosis. CDS tools had a small positive effect on changes in management based on clinical guidelines or best practices that was statistically significant compared to pre-implementation but not significant compared to control groups. These findings are limited by low statistical power, sources of bias in the comparative data such as no controls or blinding, and significant heterogeneity of interventions. Certain predictors of CDS tool success, such as an active design and integration into the clinical workflow, were absent in almost all tools evaluating management outcomes.

**Conclusions and Significance:** This systematic review is the first to review quantitative evidence of CDS tools on genetics management decisions by non-genetics providers. CDS tools show promise because they show a small effect leading to changes in management, but more studies explicitly evaluating tools and bigger sample sizes are needed to establish effectiveness.

## **PgmNr 719: A study on data sharing policies among human gene/variation databases: Data accessibility, usability and disclaimers.**

### **Authors:**

C. Yamasaki; N. Ohashi; K. Kato

[View Session](#) [Add to Schedule](#)

**Affiliation:** Graduate School of Medicine, Osaka University, Osaka, Japan

---

The human gene and variation data are used by a large number of users on wide range of human-related researches from basic sciences to applied and clinical researches. The advances on human genome research, especially personal genome analysis and studies on personalized medicine made those databases essential infrastructure for modern biomedical research. There are several human gene and variation databases, and the numbers of such databases are still increasing year by year. However, the data sharing policies and guidelines for users are different in each database and decided by each database almost independently, which may be one of the causes for difficulties of international collaborations and data-sharing.

In this study, we investigated data-sharing policies of human gene and variation databases, specially focusing on the policies for their data accessibility, usability and disclaimers. We selected and referred to total of 20 human gene and variation databases, 9 from USA, 6 from Europe and 5 from Asia. Next, we extracted and compared the database policies in terms of descriptions of data accessibility, usability and disclaimers for non-profit and commercial users. Then, we examined and summarized the conditions and restrictions for non-profit and commercial users. Here, we report the summary of the current situation and the problems on the international collaboration and data-sharing for all users, and propose what kind of data-sharing policies are required for future researches.

## PgmNr 720: Challenges and prospects of genomic researches in developing countries: Insights from African researchers.

### Authors:

O. Adewole <sup>1</sup>; T. Adewole <sup>2</sup>; S. Ottaru <sup>3</sup>; J. Edem-Hotah <sup>4</sup>; D. Kemevor-Asema <sup>5</sup>; J. Jumbo <sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine, Obafemi Awolowo University/Teaching Hospitals, Ile Ife, Osun, Nigeria; 2) Family Medicine, Obafemi Awolowo University/Teaching Hospitals, Ile Ife, Osun, Nigeria; 3) Shree Hindu Mandal Hospital – Mwanza, P. O. Box 2066, Mwanza, Tanzania; 4) Nursing, College of Medicine and Allied Health Sciences University of Sierra Leone Freetown Sierra Leone; 5) Obstetrics and Gynaecology, Korle Bu, Accra, Ghana; 6) Medicine, Niger Delta University, Amassoma, Bayelsa State,

---

**Introduction:** Africa unlike other parts of the world is lagging behind in the adoption of genomics. The implication of this, if not addressed is that global health inequities gap may continue to widen. In order to mitigate against this, it is important to understand the challenges and perception of African researchers about genomic studies.

**Aims/Objectives:** This study seeks to unravel the challenges being faced by African researchers in conducting genomic studies and their perceptions of the role genomics in healthcare.

**Methods:** We conducted a qualitative study, involving in depth interviews of 40 mid career and established researchers/physicians across three African Countries (Nigeria, Tanzania and Ghana). The locations were purposively selected. Snowball sampling technique was employed to select the respondents. The data was imported into ATLAS.ti computer software for analysis. Data were analyzed using inductive coding and results were presented thematically using network diagrams.

**Results:** The findings revealed that majority of the respondents are interested in conducting genomic studies, but have never been opportuned. Lack of facilities , non availability of funding , ignorance of the importance of genomics and lack of mentors were some of the reasons identified as being responsible. There are no curriculum for genetics in majority of the respondents medical schools. Most commonly conducted genomics studies were related to infectious disease and were mostly conducted outside their home institutions. The researchers believed that genomics can greatly impact national health care development and disease outcome. They also believed that developed countries are not doing enough to honestly encourage genomic studies by African researchers.

**Conclusion:** The study concluded that if genomics medicine will positively influence healthcare in Africa, there is a need to ensure that African researchers can conduct genomic studies in their native countries as well.

## **PgmNr 721: Description of an active educational system supporting the implementation of a personalized medicine program, in a large multi-site rural hospital.**

### **Authors:**

A. Aifaoui<sup>1</sup>; N.J. Petry<sup>1,4</sup>; A. Hutchinson<sup>1</sup>; A.J. Schultz<sup>1,3</sup>; J.F. Baye<sup>1,2,3</sup>; A. Massmann<sup>1</sup>; J. Van Heukelom<sup>1</sup>; C. Hajek<sup>1,3</sup>; R.A. Wilke<sup>1,3</sup>; M. Moore<sup>1</sup>; B. Tucker<sup>1</sup>; L. Mullineaux<sup>1</sup>; E. Royer<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Sanford Health, Sioux Falls, South Dakota.; 2) SDSU College of Pharmacy and Allied Health Professions, Brookings, South Dakota.; 3) USD Sanford School of Medicine, Vermillion, South Dakota.; 4) NDSU College of Pharmacy, Fargo, North Dakota.

---

Precision medicine has become more prevalent in healthcare over the last decade. Pharmacogenetics (PGx), an emerging field of precision medicine, has become an important tool that enables clinicians to better predict the efficacy and toxicity of specific medications. The objective of this poster is to describe our unique and robust educational program created to support the implementation of our PGx program in a large rural setting.

As a large health care institution consisting of 44 hospitals and 289 clinics in nine states, Sanford Health serves a predominantly rural population with 1,400 physicians providing care for more than 2 million patients. Imagenetics (our personalized medicine program) began offering PGx testing in 2015. Education is an essential part of PGx implementation. Our education initiatives aim to address a large audience, including patients, providers, pharmacists, laboratory personnel, nurses, and other healthcare professionals.

To maximize patient education, we developed a comprehensive series of educational material, including patient pamphlets and flyers, a public website, and various lectures to the community and various patient groups. Our public website contains storyboard-type videos and explanations that are clear and easy for patients to understand. We have found that patient education and awareness facilitated the acceptance and adoption of this new type of testing by helping to introduce, and explain the benefits of a guided treatment plan tailored to each patient's unique genetic makeup. Our employee educational program uses various tools and platforms to deploy customized educational material in order to reach a diverse audience of healthcare professionals. Our training material includes mandatory webinars, live lectures, and clinic in-services supplemented by informational handouts. We developed a centralized informational resources website that contains videos and downloadable educational materials.

The Sanford PGx educational program thus contains a wide spectrum of components to address patient and healthcare professional educational needs. We developed tools for our institution to improve the awareness about the use and interpretation of PGx testing among our healthcare professionals and the public community. This poster aims to share our experience. We also hope to provide guidance for other institutions seeking to develop an educational process and materials for a PGx program in their clinical practice.

## PgmNr 722: The triad approach: Stimulating learning in the undergraduate genetics classroom.

**Authors:**

M. Eslinger; J. Torre

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Chemistry & Life Science, US Military Academy, West Point, New York

---

National education efforts emphasize the need to practice the process skills of science rather than strictly focus on discipline-specific content. Developing these vital skills early within undergraduate education promotes cognitive progression and conceptual application; however, they require deliberate cultivation, particularly with early-stage scientists. We developed the “Triad Approach” for use in the genetics classroom as a departure from traditional lecture to encourage translational skills. The Triad leverages active-learning through iterative cycling of lectures, journal clubs, and case studies on specific genetics course related themes. Thus, students observe, analyze, and apply content three times from unique perspectives, resonating with multiple learning styles. Following Triad iterations, students were also assessed with comprehensive problem sets, requiring interpretation of data, analysis of case studies, and mining primary sources for genomic-related information. Through a systematic comparison of data from direct and indirect mechanisms (self-assessments, grades, and surveys) we show marked improvements in learning and scientific literacy. Pre- and post-course self-assessments show a 0.55 average normalized gain in quantitative knowledge (modeling); 0.33 gain in critical thinking; 0.40 gain in quantitative reasoning (math); and a 0.63 gain in analytical genetic skills (-omics). Importantly, student report a 0.70 average gain in their ability to lead scientific discussions on genomics and bioethics. Similarly, we compare cumulative pre-course GPA to genetics course performance to identify students who “over” or “under” perform through this approach and relate this to preferences for traditional learning or active learning strategies. Finally, we see higher reported means for critical thinking and learning indicators in the genetics cohort compared to students across the university. Moreover, the departure from memorization of foundational knowledge enhances their ability to apply knowledge in new ways. Together, we present this effective teaching style as a mechanism for faculty to stimulate independent student learning while leveraging contemporary tools for data acquisition and analysis, all important skills desirable to improve retention and conceptual application.

## PgmNr 723: Student genomics: Integrating DNA sequencing into a middle school classroom.

### Authors:

A. Gener<sup>1,2,3,4</sup>; K. Harris<sup>5,7</sup>; M. Wertz<sup>6</sup>; S. Richards<sup>8,9</sup>; J. Malhan<sup>10</sup>; N. Moreno<sup>5</sup>; L. Ortiz-Velez<sup>10</sup>; R. Britton<sup>10</sup>; B. Perez-Sweeney<sup>5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Integrative Molecular and Biomedical Sciences Program, Graduate School of Biomedical Sciences, Baylor College of Medicine, Houston, Texas, USA.; 2) Medicine/Nephrology, Baylor College of Medicine, Houston, Texas, USA.; 3) Department of Genetics, MD Anderson Cancer Center, Houston, Texas, USA.; 4) School of Medicine, Universidad Central del Caribe, Bayamón, Puerto Rico, USA.; 5) Center for Education Outreach, Baylor College of Medicine, Houston, Texas, USA.; 6) Baylor Academy at James D. Ryan Middle School, Baylor College of Medicine, Houston, Texas, USA.; 7) Baylor College of Medicine Biotech Academy at Rusk, Baylor College of Medicine, Houston, Texas, USA.; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA.; 9) UC Davis Genome Center, UC Davis, Davis, California, USA.; 10) Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, Texas, USA.

---

**Background:** While many partnerships exist between research institutions and secondary schools, few opportunities exist for young learners to become involved in established research projects. New technologies such as 3rd generation DNA sequencing are often perceived to be out of reach of new faculty, or postdocs and predoctoral trainees who cannot write their own grants. With slight modifications, projects can be adapted as learning opportunities for learners at all stages of development. In this classroom pilot, we adapted ongoing DNA sequencing projects at Baylor College of Medicine to meet learning objectives for a flexible 7th grade science elective. So, **how well would DNA libraries prepared in the classroom by 7th graders compare to those of a “proficient user?”**

**Methods:** Meeting every other day for 55-minute periods (4 periods total over 2 weeks), students were taught about DNA, biological information content, and the manual dexterity required to process Biosafety Level 1 DNA pellets from cultured *Lactobacillus reuteri* for downstream multiplexed sequencing on the MinION Mk1B device (Oxford Nanopore Technologies, Oxford, UK). Students were also taught the technology behind this sequencer, and spaced retrieval was used to reinforce concepts. The libraries produced in the classroom were compared to the libraries produced in a typical laboratory by a trained professional.

**Results:** Students completed a majority (53%) of procedures needed to create the libraries, and the quality of these libraries exceeded the quality of libraries from a proficient user. Data from this sequencing run has contributed to multiple manuscripts.

#### 7th Grade Classroom Libraries

1.) 4 bacterial pellets (bar02-05) (“Student Genomics”; in prep)

#### Control “proficient user” Libraries

2.) Transgenic mouse genomic DNA (bar01, 12) (in prep)

3.) Single plasmid (bar 10, 11) (Gener, 2019. BioRxiv. <https://doi.org/10.1101/611848>)

#### 4.) Multiple plasmids (bar06-09) (“Physical Multiplexing”; in prep)

**Conclusions:** To our knowledge, this pilot is the first time students at this level were able to contribute to ongoing DNA sequencing projects, serving as a proof-of-concept for future collaborations between research institutions and secondary schools. While self-contained, similar work may be completed within the current bandwidth of most sequencing centers, demonstrating that real data may be recovered from educational activities.



## PgmNr 724: Challenges and prospects of genetic education in primary and secondary science education in Japan by international comparative study.

### Authors:

R. Kawaguchi <sup>1</sup>; F. Takada <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Kitasato University Hospital, Sagamihara, Kanagawa, Japan; 2) Department of Medical Genetics and Genomics, Kitasato University Graduate School of Medical Sciences, Sagamihara, Kanagawa, Japan

---

**Purpose** Recently, genetic tests which look at human genes or chromosomes have become close to Japanese citizens. That makes them more interested in Genetics. On the other hand, it is pointed out that the contents of science education aren't enough for citizens to understand Genetics (Ikeuchi, 2012, 2015). In this study, we focused on science education in primary and secondary education which played an important role as an education of the public. Purposes of this study were understanding the present situation, extracting problems, and considering improvement plans of genetic education in Japan.

**Method** In this study, we classified "Genetics" as "Heredity" and "Variation". We selected items about heredity and variation using science curriculums in primary and secondary education which were published from the education ministry of Japan, USA (California, Massachusetts), UK (England), Australia, Canada (Ontario), Singapore, Philippine and International Baccalaureate Organization (IBO). And then, we investigated differences of genetic education among Japan, foreign countries and IBO.

**Results** Problems of current curriculums include only discussing the law of segregation in middle school, and not using the term of "mutation" in secondary education. Despite this, there is enough information taught that students can understand the link between genetics and diversity. We confirmed that genetic education in Japan is shorter than other countries and other countries deal with diversity in primary education. Genetic diseases are taught in some countries. Therefore, it is a global problem that most curriculums do not include information on various genetic diseases.

**Discussion** To improve the genetic literacy of citizens, it is important to do genetic education to all the people common in primary and secondary education. We think that it is necessary to do gradually genetic education from primary education over the multi-year and include contents which students can understand the basis of genetics link to human genetics.

**Conclusion** In this study, we understood the present situation and considered improvement plans of genetic education in Japan. It is difficult to state that genetic education in Japan is enough. We think that Japan should consider improvement of educational curriculum refer to global genetic education.

## PgmNr 725: SoapyCilantro<sup>(R)</sup> AFLP assay: A facile genomics education tool for students, STEM educators, and health care professionals.

### Authors:

P. Mahajan; M. Ohman; A. Hoerres; A. Golbach

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Pharmaceutical & Administrative Sciences, Drake University College of Pharmacy & Health Sciences, Des Moines, Iowa.

---

**Background:** Genomics is the foundation of precision medicine, the future of human health care. Since 2003, voluminous information on structure and function of human genome has been gathered. However, translation of this technical knowledge into routine health care applications has progressed at a relatively slower pace. Lack of trained health care professionals is deemed a major barrier. Experts recommend inclusion of genomics in pharmacy, nursing, medical and allied health care education. Moreover, genomic analysis of students' own DNA has been shown to improve learning outcomes and student interest in precision medicine. Despite these pedagogical benefits, this approach has an inherent ethical limitation. These methods use gene-phenotype pairs with health implications for participants, necessitating genetic counselling. Lack of access to genetic counselling precludes many institutions from offering genomic education and training. We have developed a facile DNA genotyping assay to alleviate this ethical concern further enhancing the educational value of the experience.

**Methods and Results:** We employed the 'soapy cilantro' phenotype for genetic variations rs72921001 and rs78503206, as a tool to teach effects of single nucleotide polymorphism (SNPs) on human traits. Specific PCR primer sets were designed to amplify genomic DNA containing these SNPs associated with the 'soapy' cilantro taste. Variations in DNA sequence were detected by amplicon fragment length polymorphism (AFLP) using specific restriction enzymes.

Primer sets we designed successfully amplified DNA regions containing rs72921001 and rs78503206. Agarose gel electrophoresis of amplicons before and after restriction digestion agreed with *in silico* predicted results and confirmed the presence of specific genetic variation. Using this method, we detected 'soapy' genotype in 12 out of 118 DNA samples from adult human volunteers. Our method shows a 100% concordance for the 'soapy' genotype with the taste-test phenotype. These results, along with advantages and limitations of our protocol are discussed.

**Implications:** Our method is a safe, facile AFLP genotyping assay for a human phenotype. It alleviates ethical limitations about sharing participants' genomic information. It is being used for genomic education of students from middle school to professional school, as well as for continuing education of STEM teachers, health care professionals and community members alike.

## PgmNr 726: Influence of genetic education in the practices of health-care workers in University College Hospital, Ibadan, Nigeria.

### Authors:

O.E. Olagunju <sup>1</sup>; M.O. Akinye <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Nursing Science, Faculty of Basic Medical Sciences, Obafemi Awolowo University, Ile-Ife, \*Not Applicable, Nigeria; 2) University College Hospital, Ibadan. Oyo State. Nigeria

---

Geneticist and genetic counsellors are scarce in developing nations. Providing access to genetic information and appropriate use of genetic technologies becomes prerogative duties of available health care practitioners in the health centers. Thus, this study assessed health-care workers' knowledge of genetic education and the influence it had on their practices of genetic services in a foremost tertiary hospital in Nigeria.

A descriptive cross-sectional design was employed. Sample size was calculated using Cochran formula. Multistage sampling technique was used to select 320 health workers (Nurses, Physicians, Pharmacists and Laboratory scientist). A 29-item questionnaire was used for data collection. It had three sections. Section A explored the socio-demographic details of the respondents, section B assessed their level of knowledge of genetic education and past training curricula while section C determined their current practices of genetic services. Ethical approval was obtained before data collection. Data was analyzed using SPSS version 21.0. Results were presented in descriptive and inferential statistics (chi square) used to test for hypothesis at p-value of 0.05.

Result revealed that more than half (56.6%) of the respondents were below 40 years of age, with majority having first degree in their field (77.2%) and married (71.6%). Less than half (48.2%) had more than ten years of work experience, while (55.6%) were Nurses and had worked in the field of clinical genetics (57.8%). Few (15.9%) of the respondents had poor knowledge of genetic education, and 42.5% had good knowledge. Among the items assessed and the percentage of correct responses were, unit of inheritance (60.3%), mutation definition (72.2%) and list of diseases with genetic components (75.3%). More than half (57.8%) of the respondents reported good practice while 42.2% had poor practice of genetic services. Educational level, profession and years of experience had significant association ( $p < 0.05$ ) with knowledge of genetic education. Also, educational level, profession, availability of medical genetic unit and experience in clinical genetics were significantly associated ( $P < 0.05$ ) with practice of genetic services.

This study concluded that provision of facilities for the practice of genetics in various department is needed to improve practitioners' knowledge and practice of genetics services.

## PgmNr 727: Integrating an alternative genetic services delivery model using an online platform.

### Authors:

D. Riconda <sup>1</sup>; S. Huguenard <sup>2</sup>; P. Magoulas <sup>2</sup>; S. Nassef <sup>1</sup>; S. Darilek <sup>2</sup>; A. Lewis <sup>2</sup>; H. Streff <sup>2</sup>; A. Wallis <sup>2</sup>; C. Holder <sup>2</sup>; S. Lalani <sup>2</sup>; I. Machol <sup>2</sup>; C. Shaw <sup>2</sup>; S. Chargois <sup>2</sup>; K. Worley <sup>2</sup>; L. Rosales <sup>2</sup>; B. Lee <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Health Professions, Baylor College Medicine, Houston, Texas.; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

---

The Department of Molecular and Human Genetics at Baylor College of Medicine has created and implemented a hybrid genetic service delivery model called Consultagene™ (www.Consultagene.org). Consultagene™ is a secure multimedia platform that offers a variety of educational videos, genetic counseling, and peer-to-peer genetic services. The videos enhance traditional face-to-face encounters, and the platform provides a venue for tele-genetic counseling services. For the past two years, a Whole Exome Sequencing (WES) video was integrated into the clinic workflow to assist with consenting patients for WES. Pre- and post-video surveys demonstrate improved patient knowledge, understanding, and comfort level with explaining WES to others after viewing the video (i.e. patient comfort explaining WES improved from 1.82 (N=243) on a Likert scale of 1 to 5 to 3.42 (N=196) after viewing the WES video.

As a result of this favorable patient feedback, we have now implemented an expanded array of videos and educational “journeys”, which are guided client experiences customized to the referral indication, in additional settings including cancer. Here we present our ongoing experience with our hybrid service delivery model. The videos, including: Basics of Genetics, BRCA1 & BRCA2 Sequencing, and Multigene Panel Testing, are assigned for viewing based on patients’ indications prior to their appointments. Survey data is collected on their experiences with videos as well as experiences with the tele-genetic counseling process for those who received tele-genetic counseling. Survey respondents who viewed the Basics of Genetics video (N=33) rated the informativeness of the video as 4.05 on a scale of 5. Comfort level explaining genetic testing for BRCA1 and BRCA2 variants to others improved from 2.38 (N=24) to 3.5 (N=22) out of 5 after viewing the video, and improved from 2.2 (N=10) to 3.78 (N=9) after viewing the multigene panel video.

This preliminary data suggests that the videos effectively supplement genetic counseling by improving patient knowledge and understanding of relevant genetics concepts. In addition, we have accumulating evidence that patients have more focused questions as a result of having viewed the videos in advance, therefore, improving clinical efficiency. Given the improvement in patient understanding of genetics concepts via video education, we plan to generate additional data from newly developed prenatal and preconception educational journeys.

## PgmNr 728: Exploring genetics through genetic disorders: Developing and testing an NGSS-aligned high school curriculum unit.

### Authors:

L. Stark <sup>1,2</sup>; M. Malone <sup>2</sup>; S. Homburger <sup>2</sup>; K. Fenker <sup>2</sup>; D. Dritz-Esser <sup>2</sup>; M. Kelly <sup>2</sup>; R. Perkins <sup>2</sup>; K. Breitenbach <sup>2</sup>; J. Peterson <sup>2</sup>; K. Pompei <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, University of Utah, Salt Lake City, Utah; 2) Genetic Science Learning Center, University of Utah, Salt Lake City, Utah

---

We have developed an NGSS-aligned high school genetics unit, *Exploring Genetics Through Genetic Disorders*. In this 3-week unit, students explore foundational genetic concepts on the Learn.Genetics website, then they apply those concepts to explore how a specific allele causes the symptoms of a genetic disorder. They emerge with a concrete understanding of how variations in genes—through effects on proteins, cells, tissues, and organs—lead to phenotypes. Resources include in-depth information on 5 alleles for each of 5 disorders: alpha-1 antitrypsin deficiency, cystic fibrosis, hemoglobin disorders, hemophilia, and Marfan syndrome. The unit addresses the 3 dimensions of the Next Generation Science Standards (NGSS): Disciplinary Core Ideas (HS-LS1 and LS3); Science Practices (Developing & Using Models, Analyzing & Interpreting Data); and Crosscutting Concepts (Cause & Effect, Structure & Function).

Our iterative process for developing and refining the unit involved an alpha test with two biology teachers, followed by revisions, and a pilot test with two others the following year. During the pilot phase, we collected data through: (1) a pre/post student knowledge gain assessment with 27 multiple choice and 2 open-response items, developed and validated using our published methods (Bass, Dritz-Esser & Stark, 2016); (2) a teacher end-of-implementation survey, designed to understand teachers' experiences with the unit, perceived student learning and engagement, and feedback for making revisions; and (3) classroom observations.

The 80 students who completed both the pretest and the posttest showed statistically significant learning gains,  $t(79) = 7.44$ ,  $p < .001$ , with a Cohen  $d$  effect size of .82. Surveys indicated that both teachers plan to teach the entire unit in sequence with the addition of other lessons; that students learned all the unit's intended learning outcomes either very well (genes code for proteins, most traits are influenced by both genes and the environment) or somewhat well (mutation is a natural process that generates variation in DNA sequences); and that the disease-causing allele investigation was effective for student engagement.

We will revise the unit then conduct an RCT in September 2019 with 6-8 teachers, comparing learning gains of students who experience the curriculum either with or without the disease-causing allele investigation. We will present preliminary RCT results along with pilot test results.

## PgmNr 729: ClinGen community curation: Crowdsourcing curation efforts from geneticists to citizen scientists.

### Authors:

C. Thaxton <sup>1</sup>; E. Kearns <sup>1</sup>; K. Toner <sup>2</sup>; D. DeCristo <sup>1</sup>; C.L. Kurtz <sup>1</sup>; S. Rao <sup>3</sup>; D. Azzariti <sup>4</sup>; M. DiStefano <sup>5</sup>; C. Douglas <sup>1</sup>; J. Goldstein <sup>1</sup>; J. McGlaughon <sup>1</sup>; D. Ritter <sup>6</sup>; R. Patel <sup>6</sup>; M.W. Wright <sup>7</sup>; E.R. Riggs <sup>2</sup>; L. Milko <sup>1</sup>; ClinGen Community Curation Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, Univ North Carolina, Chapel Hill, North Carolina.; 2) Geisinger Health, Danville, PA, USA; 3) Innovation Center for Biomedical Informatics, Georgetown University, Washington D.C., USA; 4) Broad Institute, Cambridge, MA, USA; 5) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA, USA; 6) Baylor College of Medicine, Baylor University, Houston, TX, USA; 7) Stanford University School of Medicine, Stanford University, Stanford, CA, USA

---

With the current advancements in molecular diagnostics and precision medicine, new genes and genetic variants are discovered daily. Recent reports highlight that ~15,000 genes are associated with human disease. Genetic variants may number in the tens to thousands per gene, indicating the involvement of millions of variants with disease. Current variant numbers in ClinVar, a variant database, indicate half a million disease-causing variants, with numbers projected to rise in the coming years. Though automation of evidence curation by machine learning approaches has the potential to improve and accelerate curation, it will not replace the need for manual curation and interpretation of complex data in the foreseeable future. ClinGen has developed and implemented standards and infrastructure to support expert clinical annotation and interpretation of genes and variants; however, a significant crowdsourcing effort is needed to support the exponential increase in the number of genes and variants requiring interpretation. In order to establish a community of volunteer curators, from clinical geneticists to citizen scientists, ClinGen established the Community Curation Working Group (CCWG). Since forming in Fall 2018, CCWG outreach efforts have matched over 200 volunteer genetics professionals from 19 countries with their preferred ClinGen curation effort (gene, variant, dosage, somatic cancer, and actionability) and clinical disease area via an online survey on the ClinGen website. In addition, the CCWG is developing a web-based volunteer annotation program to engage citizen scientists who wish to contribute to the overarching genomic knowledge base of clinical evidence, such as human phenotyping and disease ontologies, variant identifiers, and tags corresponding to ACMG and AMP sequence variant pathogenicity criteria. Here, we will present an overview of the engagement process, curation results, and volunteer feedback. Global participation is the cornerstone of sustainable and widely utilized gene and variant curation and expert interpretation, and the growing awareness of ClinGen activities with the development of web-based curation interfaces is cultivating a grassroots interest in a “crowdsourcing” effort that will enable large-scale enhancement and acceleration of ClinGen’s mission to “curate the clinical genome.”

## **PgmNr 730: Genetic education: Preparing Physician Assistants (PAs) for the genetic patient.**

### **Authors:**

L. Tribble; W. Patterson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Greenwood Genetic Center, Greenwood, South Carolina.

---

### **Background:**

In today's practice of medicine, there is growing acceptance of the role that genetics and genomics play in providing effective patient care. Given the shortage of clinical geneticists and genetic counselors, there is a need to bring other health care providers into the care of genetic patients. The challenge is to provide the genetic education to enhance providers' knowledge, application and practice of genetics and genomics.

### **Objectives:**

To address this challenge and shortage, the Greenwood Genetic Center's (GGC) Division of Education designed and implemented a pilot program targeting physician assistant students. The objectives were to provide background on basic processes, genetic testing, family history, and when to consider referring to genetic professionals. With this information, it is anticipated that when in practice, these students will have a better understanding and knowledge of what to look for and how to better manage their patients.

### **Methods and Design:**

The pilot program was a series of five meetings with the PA students consisting of formal presentations on the major areas of cytogenetics, molecular and biochemical genetics, and dysmorphology, through the use of patient cases. Through grant support to expand PA education in biochemical genetics, students were also introduced to lysosomal storage diseases, a group of rare conditions that are thought to be under diagnosed, but for which many medical management options exist.

### **Results and Conclusion:**

Formal student evaluations and pre and post assessment scores indicated an improved understanding of the genetic content. This is a positive result but one on which to continue building. Genetics content should be more fully integrated throughout the training curriculum to reinforce the role and application of genetics and genomics in family medicine, oncology, cardiology, and other areas of practice. Genomic medicine is here and PAs must be prepared to be the first medical encounter for the patient and to provide professional, long-term care for the patient.



## **PgmNr 731: Genetically modified education: Exploring new paradigms for public engagement with genomics at community biology (DIYbio) labs.**

**Authors:**

E. Tuck

[View Session](#) [Add to Schedule](#)**Affiliation:** Genspace, Brooklyn, New York.

---

As genomic information and technologies become increasingly prevalent and integrated into our daily lives in sectors such as medicine, material design, agriculture, and environmental conservation, there is an increasing need for robust public engagement and community-focused dialogs about these myriad applications. With respect to human health specifically, many people express curiosity to learn more about their personal genomic information, as evidenced by the growing popularity of consumer genetic testing services, but they are often left with significant gaps in the ability to dive more deeply into this information with trusted experts, and there are very few places for dialog to explore what this information might mean for themselves, their families, and their communities. Community Biology Labs (sometimes referred to as DIYbio labs or biohackerspaces) have emerged over the last ten years as community-centered spaces where anyone, regardless of educational background or prior experience, can explore various aspects of emerging biotechnology in a hands-on, experiential, and personally relevant way. Popular media representations of these spaces implies they are full of reckless, radical self-experimentalists, but in reality this growing global movement provides a space for people to learn about new technologies and participate in the growing bioeconomy in a safe, environmentally responsible, and inclusive setting.

In this study, we investigate the role that Genspace, a community biology lab based in Brooklyn, NY and the first community lab world-wide, play in public engagement with genomics. Over the past 6 months, we have collected demographic information and conducted focus groups to discuss the types of information people are seeking to investigate about their genomes and their role in understanding their identity, their health risks, and the ethical, legal, and social implications of exploring this information in a Community Biology Lab setting. Emerging themes include a strong emphasis on data access and sharing, as well as a larger interest in developing the computational and data literacy skills needed to investigate genomic data themselves. Community Lab participants express strong views about the autonomy of personal genomic data, concerns over the role of industry in monetizing genomic data, and general optimism about the role genomic information can play in making healthcare and other personal decisions.



## PgmNr 732: Improving undergraduate medical genomics education with an elective intercession course.

### Authors:

H.C. Wick <sup>1</sup>; V. Kulo <sup>2</sup>; C.J. Bowen <sup>1,2</sup>; J. Prudent <sup>2</sup>; D. Valle <sup>1</sup>; J. Bodurtha <sup>1</sup>; N. Sobreira <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Medical Institutions, Baltimore, MD;  
2) Johns Hopkins University School of Medicine, Baltimore, MD

---

Next generation sequencing (NGS) technology is increasingly used as a diagnostic tool and for aiding in complex trait, genetic risk assessment and pharmacogenetic testing. However, physician education on how to use and interpret NGS technology, particularly whole exome sequencing (WES) lags behind. Undergraduate medical education needs to encompass NGS technology, data analysis, interpretation, counseling, and ethical consideration.

We developed an elective Topics in Interdisciplinary Medicine (TIME) course available to Johns Hopkins University (JHU) School of Medicine students as a pilot in 2017, then as an integrated course in the Genes to Society curriculum in 2018. The course objectives were: 1) Understand differences, applications and consequences of single gene, disease gene panels, and WES/WGS in the clinic; 2) Understand laboratory/analysis techniques used to generate NGS data; 3) Identify key databases and programs to evaluate sequence data and interpret variants; 4) Understand variant classification criteria, causality, variants of unknown significance, incidental findings, and how these results are reported; and, 5) Appreciate the necessity for pre-/post-genetic counseling for sequence-based tests, especially for WES/WGS, including review of consent, test reports, and related ethical issues. The 3 day course included lectures and small group sessions with literature discussion, role-playing of patient consenting, and identifying pathogenic variants with PhenoDB (Sobreira et al, Hum Mut, 2015), an online, problem-based teaching tool developed by as part of the Baylor-Hopkins Center for Mendelian Genomics. A student survey was used to evaluate perception of course quality/effectiveness and identify areas of improvement.

The course was positively received by the majority of students. 61.5% of students rated the quality of the course as Very Good or Excellent. 84% Agree/Strongly Agree that they received clear learning objectives for the course. 84% Agree/Strongly Agree that the course content reflected those learning objectives. And, 80.5% of students Agree/Strongly Agree that they feel confident in their ability to share information on genetic conditions and other supportive resources with patients and families. The course will be offered annually for third year medical students at JHU School of Medicine, with the long-term goal of increasing future physicians' familiarity with genetic testing and comprehension of clinical WES/WGS.

## **PgmNr 733: “Genetics for Clinicians,” an online-only graduate course: Design, methods, and experience.**

### **Authors:**

R.S. Wildin <sup>1</sup>; C.A. Giummo <sup>1</sup>; T.F. Williams <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Pathology and Laboratory Medicine, University of Vermont Health Network, Burlington, Vermont.; 2) Equitas Life Sciences, Essex, Vermont.

---

Genetics education in graduate nursing programs is essential to prepare nurses and other health providers for advanced clinical practice. A 3-credit, 14-week course in healthcare genetics has been offered annually at the University of Vermont since 2014 as part of the Ph.D. Nursing and Masters-level Medical Laboratory Science programs. Most students work in healthcare settings across a rural state with limited availability to attend on-site classes. The course is designed with modular content including video lectures, reading assignments, quizzes, and facilitated discussions. Cross-functional Genetics professionals have expanded and updated the clinical and case-based content, and created or modified modules. An online teaching platform is used with robust capabilities for multimodal teaching and grading, and for facilitating interactions via asynchronous discussion boards. Despite the limitations of online learning, ratings from the over 100 students completing the course are consistently high. Mentored clinical experiences are replaced by reflections placing their learning in their own clinical context by retelling real-life nursing or personal experiences on a guided discussion board.

Advantages of online teaching of genetics include reusability of modules, automated module release, integration with grade management platforms, accessibility for geographically distributed students and those with diverse work schedules, and instructor-facilitated discussions involving >30 students. Asynchronous teaching alleviates scheduling conflicts with patient care responsibilities for the instructors as well. Incorporating simulated clinical encounters and other web-based tools enriches the learning experience. Lack of face-to-face interaction with instructors, and the associated barriers to mutual insight and respect, is a potential disadvantage. Overall, the utilization of multiple interactive learning modalities, schedule and teaching flexibility, and integration into technology platforms has empowered a highly successful learning experience. Expansion to other clinician learners is contemplated.

## PgmNr 734: An analysis of genetics-related discussions on social media over the last decade.

**Authors:**

N.M. Ferraro; A. Lavertu; R.B. Altman; S.B. Montgomery

[View Session](#) [Add to Schedule](#)

**Affiliation:** Stanford University, Stanford, California.

---

Increased accessibility of genetic testing is shaping public discussions around the impact of genetics. The coinciding rise in social media usage provides an opportunity to assess how genetics is discussed on these platforms. We focus on the social media site Reddit, the 5th most trafficked site in the US. Reddit is organized into topic-specific communities that indicate at a high-level the focus of that set of users' interactions and comments.

To identify a set of relevant terms, we created a word embedding model and conducted a beam-search within the embedded space starting with the term "genetic". This resulted in a set of 8,580 genetics-related terms and hundreds of millions of comments containing those terms across 318,105 topic-specific boards.

The main variation in this set of terms in the embedded space occurs on an axis of formal usage more typical of academic fields like biology or geography to informal usage, with references to racial identifiers, stereotypes, and relationships. Clustering terms reveals 7 primary topic spheres, including discussions of (1) racial/ethnic identity, (2) sociodemographics, (3) geography and history, (4) populations and communities, (5) ancestry and population genetics, (6) molecular biology, and (7) trait associations.

We track comments containing terms in each cluster from 2008-2018. The only cluster to increase in proportional frequency relative to 2008 is (1) racial/ethnic identity. Based on topic board usage, clusters 1 and 5 are the most dissimilar, indicating little overlap in discussions around genetics between these communities. Specifically, for a forum discussing DTC genetic testing, 23andme, we find that the predominant cluster switches in 2017 from discussions centered around (6) biology to (3) geography and (4) populations.

As an example, we highlight responses to high-profile coverage of genetics in the media, following presidential candidate Elizabeth Warren's release of genetic test results. Comments mentioning Warren in all clusters increased substantially in the month of the release, before returning to baseline in the next month. Cluster 4 increased most dramatically. This may indicate a short window of opportunity for researchers to inform public perception surrounding high-profile genetics work. Overall, we aim to inform genetics researchers of varied public interpretations of the field, and leverage these insights to understand how researchers could better communicate with the public.

## PgmNr 735: A mixed recruitment approach to enrolling participants in monogenic diabetes genetic research.

**Authors:**

J. Goehring; M. Snyder; K. Blessing; P. Gingrich; S. Thomas; Y. Hu; D. Carey

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genomics, Geisinger Med Ctr, Danville, Pennsylvania.

---

Diagnosis of monogenic diabetes (MD) is complicated by phenotypic heterogeneity and lack of adequate screening methods. Via a multicenter study, Geisinger and 3 other sites sought to develop an approach to identify MD using varied recruitment, systematic screening, and genetic testing. Unlike the 3 urban recruitment sites, Geisinger is a rural healthcare system in central Pennsylvania and southern New Jersey. To identify the optimal recruitment method in a rural-based endocrinology population, we compared the success of various recruitment strategies used in this study.

Patients were screened for enrollment through an electronic health record (EHR) data pull and patient intake questionnaire (IQ), administered in one of three ways: regular mail, via a secure EHR patient portal (MyG), or in person during a routine endocrinology outpatient visit. The research team reviewed all IQs for study eligibility. Eligible participants were invited to complete a study visit, during which study data forms and venipuncture for laboratory testing were completed. The data were then used to select participants for genetic testing. Those who had genetic testing were notified of a positive or negative result. Cascade testing was offered to family members.

IQs were mailed to 2,240 individuals (1,970 adult, 270 pediatric); 67 (55 adult, 12 pediatric) enrolled in the study (3%). IQs were sent to 1,541 individuals (1,266 adult, 275 pediatric) via MyG; 50% of adults and 33% of guardians of pediatric patients read the recruitment message without responding; 34% of adults and 49% of guardians did not open the message; 14% of adults and 17% of guardians responded to the message; 26 enrolled in the study (2%). IQs were administered to 580 adult patients at the time of an outpatient endocrinology visit; 58 enrolled in the study (10% enrolled). These results suggest that in this rural health care system in-person encounters are more effective enrollment strategies than either mail or electronic approaches. Further research is needed to determine the factors that account for this difference.

## PgmNr 736: A pilot study integrating clinical genome sequencing into the US Air Force: Airmen's perceptions and attitudes after receiving results in the MilSeq Study.

### Authors:

R.L. Hsu <sup>1</sup>; S. Pereira <sup>1</sup>; J.O. Robinson <sup>1</sup>; R. Islam <sup>1</sup>; M.A. Majumder <sup>1</sup>; E. Parasidis <sup>2</sup>; M.J. Mehlman <sup>3</sup>; K. Christensen <sup>4</sup>; M.D. Maxwell <sup>4</sup>; C. Blout <sup>4</sup>; B. Zettler <sup>4</sup>; M. Lebo <sup>5</sup>; J.M. Killian <sup>6</sup>; M. De Castro <sup>7</sup>; R.C. Green <sup>4</sup>; A.L. McGuire <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, Texas.; 2) Ohio State University, Columbus, OH; 3) Case Western Reserve University School of Law, Cleveland, OH; 4) Brigham and Women's Hospital, Boston, MA; 5) Partner's Personalized Medicine, Cambridge, MA; 6) Joint Base San Antonio - Lackland AFB, JBSA-Lackland AFB, TX; 7) Keesler AFB, Keesler Air Force Base, MS

---

Clinical integration of exome sequencing (ES) in active-duty military settings raises unique considerations, including how ES results may influence service members' careers.

The MilSeq Project is a pilot study that enrolled ostensibly healthy, active-duty Airmen in the United States Air Force (USAF) to undergo ES (n=75). Airmen completed surveys to assess their perceptions of ES prior to (pre) and 6 weeks after (post) receiving results (n=69). Airmen's ES results were incorporated into their medical records and included monogenic findings, carrier status variants, disease-associated risk alleles, and pharmacogenomic information.

The majority of Airmen who responded at post (66%) agreed that they trust the USAF with their genetic information. When asked whether ES should be used for duty assignments, 32% disagreed and 56% were neutral; when asked whether ES should be used for deployment decisions, 34% disagreed and 50% were neutral. Most (76%) did not support the USAF requiring Airmen to undergo ES.

Despite this, before undergoing ES, 87% of Airmen thought that their results were unlikely to influence their career in the Air Force, and at post, 96% reported that the ES results had in fact no impact on their career (regardless of results type(s) received). The majority of those respondents (99%) also did not think there would be an impact to their career in the future.

At post, 85% of Airmen agreed that ES has health benefits and 83% agreed that the benefits outweigh the risks. Airmen on average scored the future utility of ES at post an 8 ( $\pm 1.6$ ) out of 10 on a scale anchored by (1) not at all useful to (10) extremely useful. Participants concerns about future insurability and discrimination decreased significantly ( $p < 0.05$ ) at follow up (38% to 19% and 18% to 10%, respectively).

We found that Airmen perceived benefits to receiving ES results. At post, Airmen were less worried about discrimination than at pre, and did not think the information would impact their career. Yet,

they remained wary about the potential influence of ES in duty and deployment decisions. Our findings suggest an openness among participant Airmen to incorporation of ES in the Military Health System with some use limitations.

## PgmNr 737: Political slant, gaps, and misinformation in public discussions of DNA testing of migrant families at the US-Mexico border.

### Authors:

D. Madden <sup>1,3,4</sup>; J.K. Wagner <sup>1</sup>; V. Oray <sup>2</sup>; S.H. Katsanis <sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Translational Bioethics & Health Care Policy, Geisinger, Danville, PA; 2) Duke Initiative for Science & Society, Duke University, Durham, NC; 3) Mary Ann & J. Milburn Smith Child Health Research, Outreach, and Advocacy Center, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 4) Department of Pediatrics, Northwestern University, Chicago, IL

---

The 2018 separation of migrant families at the US-Mexico border and subsequent use of DNA testing for family reunification sparked controversy in news and social media. Public debate continued into 2019, fueled by a rapid DNA program at the border by US Immigration and Customs Enforcement to perform kinship analysis to verify migrant family claims. We sought to evaluate how the use of DNA in a politically-charged, non-medical context was covered by news and social media. We conducted systematic searches of newspaper databases and the social media platform Twitter for coverage of DNA testing and migrants for the 2-month period of June and July 2018. After refining for relevance, we identified 183 articles and 153 Tweets. Of the articles, only 70 contained discussion of DNA testing beyond reporting that it was being used, and only 27 had substantive DNA testing-related content. For comparison, our search parameters were repeated in May 2019 to capture articles and Tweets about the rapid DNA program. Using qualitative approaches, we evaluated the data for biases: (1) political, (2) immigration (i.e., "zero-tolerance") policy, and (3) DNA testing (i.e., for/against testing). We then coded content to identify coverage of the science, process of DNA testing, and major ethical considerations. We identified trends, including common topics and miscommunications, as well as significant gaps in coverage of DNA testing and ELSI topics. In both news articles and Tweets, we found that a politically conservative or pro-zero tolerance policy slant strongly corresponded to a pro-DNA testing stance, while a liberal or anti-zero tolerance policy slant corresponded diversely to pro-, anti-, and neutral DNA testing stances. Public discussion of recent events at the US-Mexico border demonstrates both the complexity of the arena in which DNA testing for immigration is considered as well as key communication issues facing the genetics community. These issues include: (a) lack of transparency on the use of genetic information (causing speculation); (b) ineffective inclusion of genetics experts in public conversations (causing misinformation to spread); and (c) ill-preparedness of the genetics community to react to public confusion and outcry (causing scientific misunderstandings to go uncorrected). Genetic experts' contributions to such discussions shape public opinion and are an opportunity to accurately communicate the risks and benefits of DNA testing in non-medical contexts.

## **PgmNr 738: The ethical considerations of using ancestry DNA testing in statelessness and tribal citizenship cases to determine heritage and attribute nationality.**

### **Authors:**

V. Oray <sup>1</sup>; A. Porter <sup>1</sup>; S. Katsanis <sup>1, 2, 3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Duke Initiative for Science & Society, Duke University, Durham, NC; 2) Mary Ann & J. Milburn Smith Child Health Research, Outreach, and Advocacy Center, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 3) Department of Pediatrics, Northwestern University, Chicago, IL

---

As national security interests lead to the strengthening of borders, there are heightened efforts to detect fraud in nationality claims. One tool to evaluate these claims is through ancestry DNA testing. However, results from commercial DNA testing for ancestral origins can be erroneously confounded with nationality. While DNA can reveal clues regarding ancestral geographic migration patterns, it cannot determine countries in which past generations might have resided or held citizenship. Nevertheless, such DNA tools might be informative, or even the only recourse to provide evidence of ancestral heritage, such as in the case of statelessness or tribal membership. The UNHCR estimates that 10 million people are stateless, that is lacking documentation of citizenship to any country. Stateless people might not be permitted to travel, marry, and/or work either in their country of origin or their destination country. Similarly, other sovereign states, such as indigenous tribes in the United States and Canada, have begun using DNA testing for citizenship. The ability to provide scientific evidence of native ancestry and receive an authentic tribal membership can allow one to have access to material goods and emotional support. These two examples provide an opportunity to examine how ancestry DNA results are used as scientific evidence of nationality and group membership. As part of our research into the applications of genetic information in immigration, we evaluated examples of contexts where DNA testing was applied to determine ancestral origin as a proxy for nationality, considering the personal, sociopolitical, and cultural risks and benefits. In one of these contexts, an ancestral DNA test was allegedly used to screen for claims of nationality for a stateless person to support a refugee claim. We also examined the use of ancestry DNA results to support claims to American indigenous tribal citizenship. We compared these examples to identify overlapping themes with regard to autonomy, identity, human right to migration, documentation of statehood, informed consent and risk of stigmatization. Our result is an examination of the ethical considerations for use of ancestry DNA testing, alternatively, to screen out nationalities, versus as an individual tool for documentation of nationality. Ultimately, the use of ancestry DNA testing as a proxy for nationality leads to the potential conflation of biological origins with political borders and cultural narratives.



## PgmNr 739: Can forensic genetic genealogy and SNP typing be used to investigate missing migrants' cases? Should it?

### Authors:

A. Porter <sup>1</sup>; E. Sweeney <sup>2</sup>; S. Katsanis <sup>1,3,4</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Duke University, Durham, LA.; 2) Bode Technology Group, Lorton, VA; 3) Mary Ann & J. Milburn Smith Child Health Research, Outreach, and Advocacy Center, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 4) Department of Pediatrics, Northwestern University, Chicago, IL

---

The reported successes of forensic genetic genealogy (FGG) in criminal cases and the ensuing debates regarding the privacy implications of this investigative approach prompted us to question whether SNP typing could be useful-or should be considered-to investigate missing migrants' cases. Over the last decade, thousands of human remains have been found along the southern US border. Due to a lack of DNA data-sharing, deceased migrants remain unidentified for years. Similar to the cold cases comprising most of the criminal FGG cases to date, missing migrants' cases have few options for resolution. But the privacy implications of law enforcement investigations resulting from genetic leads to biological relatives are unresolved, and not yet considered for missing migrants' cases. The data in public databases (GEDmatch) are thought to not be diverse enough to provide any identity by descent (IBD) similarities useful for non-European origin cases. Yet, if a database were diverse enough to benefit Latinx cases, what would be the ethical and privacy implications of such an approach? We sought to test the feasibility of using SNP typing to develop leads via FGG and phenotypic markers to inform unidentified migrants' cases. We typed DNA specimens from three sets of unidentified migrant remains for genome-wide SNPs. SNP data with greater than 60% coverage were compared to both public and private genealogical databases to determine IBD overlap that could generate leads. We also estimated biological origin of human remains through analysis of ancestry informative markers (AIMs). These data were used as leads for nationality, which is useful for verifying post-mortem data, such as ID cards. We chronicled the successes and failures of (a) gaining leads to biological relatives of deceased migrants and (b) predicting country of origin for the remains. We documented ethical considerations as each case was investigated, outlining concerns related to (a) data privacy, (b) personal privacy of living relatives, (c) the effect of stringent border policies on investigations, and (d) the representation of Latinx genomic data in SNP databases. Examining the ethical challenges in these cases can inform broader policy options for use of FGG in criminal cases. With this study, we highlighted the key considerations of FGG and SNP typing in missing migrants' cases with an aim to ensure that the rights of those involved in this process (biological relatives, the deceased) are respected.

## PgmNr 740: Cancer patients' preferences for incidental genomic sequencing results.

### Authors:

S. Shickh <sup>1,2</sup>; A. Sebastian <sup>1,2</sup>; M. Clausen <sup>1</sup>; C. Mighton <sup>1,2</sup>; C. Elser <sup>2,3</sup>; A. Eisen <sup>4</sup>; R. Kim <sup>2,5</sup>; M. Aronson <sup>3</sup>; S. Panchal <sup>3</sup>; T. Graham <sup>4</sup>; J. Carroll <sup>2,3</sup>; J. Lerner-Ellis <sup>2,3</sup>; K. Thorpe <sup>1,2</sup>; A. Laupacis <sup>1,2</sup>; Y. Bombard <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) St. Michael's Hospital, Toronto, ON; 2) University of Toronto, Toronto, ON; 3) Mount Sinai Hospital, Toronto, ON; 4) Sunnybrook Health Sciences Centre, Toronto, ON; 5) University Health Network, Toronto, ON

---

**Background:** Genomic sequencing (GS) has the potential to identify a large volume of secondary or incidental results (IR), which can be associated with various types of diseases and levels of clinical actionability. Guidelines recommend that patients undergoing GS have the option to receive medically actionable IR. However, the literature suggests that patients are interested in learning results beyond those classified as medically actionable. Most studies evaluating patient preferences use a limited range of IR; there are few studies assessing patients' preferences for the broad range of possible results that can be revealed by GS.

**Aim:** We aimed to describe patient preferences for receiving a broad range of IR.

**Methods:** We conducted an analysis of preference data generated within a randomized controlled trial of a decision aid (DA) ([www.GenomicsADvISER.com](http://www.GenomicsADvISER.com)) for the selection of IR from GS. Participants were adult cancer patients who had previously received an uninformative result from genetic testing related to their cancer. Participants enrolled in the trial used the DA or spoke with a genetic counselor to choose between five categories of IR: 1) medically actionable and pharmacogenomics results 2) common disease SNPs, 3) rare Mendelian disorders, 4) early-onset neurological diseases and 5) carrier status.

**Results:** One hundred thirty-three patients participated in the trial. Participants were predominantly White/European (74%), female (90%), and  $\geq 50$  years old (60%). The majority of participants chose to receive incidental results. About 97% of participants (129/133) elected to receive medically actionable and pharmacogenomics results. Despite the lack of medical actionability of the other categories, participants were overall very interested in receiving them; 74% selected common disease SNPs, 75% chose carrier results, while 53% chose to receive results associated with rare Mendelian diseases and 59% early-onset neurological disorders.

**Conclusions:** Our results suggest that patients are interested in receiving a broad range of incidental results, including those without evidence of medical actionability such as early-onset neurological diseases. This is in contrast to current guidelines that recommend only the return of medically actionable results. In addition to consideration of the clinical evidence, future guidelines should take patient preferences into account.

## PgmNr 741: Balance of requests for molecular studies between different medical specialties in 2018.

### Authors:

I. Bernal <sup>1,2</sup>; Y. Gomez <sup>1</sup>; A. Muñoz <sup>1,2</sup>; S. Maradei <sup>1,2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Genetics, BIOTECGEN S.A, Bogota, Colombia; 2) National University of Colombia

---

### Introduction

In recent years, there has been a marked increase in the technical capacity for detecting variants in DNA that allow a diagnostic approach to multiple pathologies that previously could not be determined by aetiology.

With the advent of these new technologies and knowledge about the human genome there has been a dramatic increase in the request for molecular studies by multiple medical specialties, but without increasing knowledge in the relevance of request for it or its proper analysis.

The studies that present greater difficulties for its adequate interpretation are the complete exomas and the genomes.

This makes it necessary to evaluate the capacity of the different medical specialties for the interpretation and classification of all the findings observed in molecular tests and their impact on the patient.

### Objectives

To determine which are the specialties that most requested molecular studies in 2018 and the pertinence of requesting the same.

### Methodology

A review of the BIOTECGEN S.A. molecular studies databases corresponding to the year 2018 is carried out, grouping them by applicant speciality, type of study and result.

### Results

During 2018, a total of 1450 orders of molecular studies were received from 18 different medical specialties, with a total of 363 positive results (25.03%) and 209 variants of uncertain significance (14.41%). The specialties that most requested molecular studies were clinical genetics and oncohematology. Breast surgery presented a percentage equal to clinical genetics (30%) while positive results and 44.4% of the requesting specialties obtained only negative results or variants of uncertain significance.

### Discussion

It is important to analyze the clinical context of each patient together with management guidelines aimed at the molecular study of each pathology, seeking of pertinent molecular tests, taking into account that only 30% of all requested studies are positive in the best of cases.

Given the percentage of studies with variants of uncertain significance, it is necessary to instruct the different medical specialties in order to be more assertive

It is striking to note that approximately half of the specialties requesting molecular tests do not obtain any positive results, which translates into misdiagnosis

It should be assessed whether all medical professionals should be allowed to request molecular studies or whether this activity should be restricted to a select group of specialties with the appropriate training

## PgmNr 742: Models of shared decision making for counseling of genetic tests.

**Authors:**

S. Lin

[View Session](#) [Add to Schedule](#)**Affiliation:** Genetic Counseling Center, Chi Mei Hospital, Tainan, Taiwan, Taiwan

---

Model of shared decision making (SDM) offers a structured way to help patients to reach evidence-informed and value-based decisions. The SDM process is relevant in genetic counseling especially for conditions which need a close discussion and trade-off between harms and benefits. But there is a lack of examples about how to accomplish the approach using SDM in genetic counseling. Our aim here is to share our experience of how to do SDM in genetic testing. We developed two models of SDM, one focus on non-invasive prenatal testing and another on breast cancer gene testing. In our project, we constructed tools to facilitate SDM communication and measurement of the effectiveness of the process. For the convenience of genetic counselors and genetic health educators, in addition to SDM patient aid tools, we also develop visual aid APP with local language to assist communication in related genetic concepts and knowledge. The SDM models were conducted in prenatal care clinic and familial cancer prevention program in a genetic counseling center located in southern Taiwan. Participants included pregnant women who ask for prenatal genetic test and family members of breast cancers who came for cancer preventive screening. An overall rating of satisfaction from participants was 95% agree that good and above. Comparison of pre- and post-test for the genetic knowledge content after the SDM process showed that genetic knowledge score increased and most significantly in the concept of risk and detection rate of gene tests. In summary, we presented facilitating strategies, patient decision aids and measurement tools for conducting and evaluation of SDM in genetic counseling. The SDM models provide structured communication, increase patient engagement and autonomy, and facilitate patient-centered decision-making.

## PgmNr 743: Alabama Genomic Health Initiative: Outcomes from the first two years.

### Authors:

K.M. East <sup>1</sup>; A. Cannon <sup>2</sup>; W.V. Kelley <sup>1</sup>; I.P. Moss <sup>2</sup>; E. Miskell <sup>2</sup>; K.M. Bowling <sup>1</sup>; M.L. Thompson <sup>1</sup>; T. May <sup>1</sup>; M. Nakano-Okuno <sup>2</sup>; S. Sodeke <sup>2</sup>; J. Edberg <sup>2</sup>; J.J. Cimino <sup>2</sup>; M. Fouad <sup>2</sup>; W. Curry <sup>2</sup>; A. Hurst <sup>2</sup>; G.M. Cooper <sup>1</sup>; G.S. Barsh <sup>1</sup>; M. Might <sup>2</sup>; B.R. Korf <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama.; 2) University of Alabama at Birmingham, Birmingham, Alabama

---

The Alabama Genomic Health Initiative (AGHI) is one of the first statewide, state-funded efforts to provide genomic testing on a broad basis. AGHI engages two distinct cohorts who enroll into a research protocol that provides access to genetic testing and offers biobank participation. The 'affected' cohort consists of children and adults with undiagnosed disease who receive whole genome sequencing (WGS) to identify primary and secondary findings. The 'population cohort' consists of interested adults who receive a genotyping assay to identify risk factors in medically-actionable genes. In the first two years of AGHI, 191 affected and 4527 population participants have been enrolled. Consent and result disclosure in the affected cohort is facilitated by staff at clinical research sites. Among 142 affected individuals analyzed to date, 60 (42%) have received a primary finding (15 pathogenic, 13 likely pathogenic, 32 variants of uncertain significance); seven individuals (2.8%) have received a secondary finding. Among 4357 analyzed population participants, 66 pathogenic or likely pathogenic variants have been returned to 65 participants (1.5%). Genetic counselors disclose positive results via phone, and written result reports are mailed to all participants. An online survey of population participants assessed their motivation, satisfaction, and outcomes. Of 3874 participants that received a result and agreed to be re-contacted, 858 participants completed the survey (22% overall response rate), including 21 individuals with a medically-actionable result (40% positive result response rate). The largest motivators for participating were concern for future health (29%), contribution to genetic research (23%), and general curiosity (18%). Twelve incorrectly recalled receiving a medically-actionable result and 45 reported they were unsure of their result type. Seven percent of respondents have or plan to talk with a medical provider, 2% had a test/procedure, and 12% made a health or wellness change because of their AGHI result. The majority of respondents (78%) shared results with family and 4% reported that a family member took an action based on this information. Most respondents (89%) were very or extremely satisfied with their decision to participate in AGHI. Results from this study elucidate the impact of genomic screening for AGHI participants, as well as inform changes to improve communication with participants and increase understanding

## **PgmNr 744: Parent interpretations of their child's genomic test results: Informing development of a return of results communications tool for NYCKidSeq.**

### **Authors:**

D. Watnick <sup>1</sup>; S.A. Suckiel <sup>3</sup>; K. Gallagher <sup>2</sup>; J.A. Odgis <sup>3</sup>; N. Teitelman <sup>1</sup>; S.M. Dolan <sup>4</sup>; N. Kelly <sup>2</sup>; M.A. Ramos <sup>5</sup>; R. Zinberg <sup>7</sup>; J. Lopez <sup>2</sup>; J.E. Rodriguez <sup>3</sup>; B. Gelb <sup>6</sup>; C. Horowitz <sup>5</sup>; E.E. Kenney <sup>3</sup>; M. Wasserstein <sup>2</sup>; L.J. Bauman <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Pediatrics, Montefiore Medical Center / Albert Einstein College of Medicine, Bronx, NY; 2) Pediatric Genetic Medicine, Montefiore Medical Center, Bronx, NY; 3) Center for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Department of Obstetrics & Gynecology and Women's Health, Montefiore Medical Center / Albert Einstein College of Medicine; 5) Center for Health Equity and Community Engaged Research, Icahn School of Medicine at Mount Sinai, New York, NY; 6) Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

---

### Background

Returning the results of genetic and genomic testing to families is a complex process that combines education about the testing itself, interpretation of often ambiguous results, and sensitivity to the emotional roller coaster families experience.

### Objective

The NYCKidSeq Study is developing a novel web-based tool to improve return of results from whole genome sequencing in a racially and ethnically diverse population. To guide tool development, we conducted formative research to understand how parents interpret their child's genomic test result.

### Methods

We recruited and screened parents from physicians' patient lists at two NYC hospitals where their children had undergone whole exome sequencing, target gene panel or microarray within the previous 12 months. We conducted in-depth interviews with 22 parents of diverse children (5 Black; 10 Latinx, 5 white; 2 multi-racial) who had received results classified as 'negative' (N=10), 'uncertain' (N=7), or 'positive' (N=5). Analysis was conducted using grounded theory's constant comparative method across data points, cases and themes.

### Results

Parents described the diagnostic odyssey they undertook to identify and treat their child's illness and symptoms, with genomic testing being only one of many diagnostic experiences. Parents' understanding of 'negative' test results was variable: minimal, but un concerning because results were negative; simple, that "everything was normal"; or self-censored, that further inquiry was unwarranted. 'Positive' results were clearer, potentially overwhelming but also satisfying to "know the name" of the disease and a path to treatment. 'Uncertain' results left parents concerned and confused, with their diagnostic odyssey unresolved. Regardless of classification, parents conflated

test results with the presence of their child's illness, causing misunderstandings about the actionable meaning of test results. Parents generally accepted not understanding results because they assigned responsibility for that understanding to trusted providers.

### Conclusion

Parents struggled with the meaning of genomic test results delivered as being 'negative' or 'uncertain' but understood 'positive' results clearly. Clinicians should be mindful that patients usually don't understand results and may not ask questions, trusting their providers to manage results for them. For parents, the treatment implications of test results rather than complete understanding of genetic findings is the priority.



## PgmNr 745: The NHGRI Genome Sequencing Program: Products & deliverables.

**Authors:**

T. Stephan; NHGRI Genome Sequencing Program

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Human Genome Research Institute, Bethesda, Maryland.

---

The Genome Sequencing Program, funded by the National Human Genome Research Institute, uses genome sequencing to identify genes and genomic variants involved in human disease. The program supports gene and variant discovery in rare, inherited (Mendelian) cases as well as common, genetically complex diseases. The program also supports analysis and tool development.

The Centers for Mendelian Genomics (CMG) use exome sequencing and complementary genomic approaches to discover as many genes and variants underlying the full spectrum of Mendelian inheritance patterns. The four centers coordinate with sample providers, clinicians, clinical sequencing laboratories, patients, and model organism researchers to disseminate knowledge and discoveries. To date, the CMGs have sequenced over 61K individuals and made over 1,700 novel disease associations - a significant contribution to the 65% of Mendelian phenotypes with a known underlying gene (OMIN).

The Centers for Common Disease Genomics (CCDG) have sequenced over 80K WGS and 81K WES across three main disease areas: neuropsychiatric (Alzheimer's, autism, and epilepsy), immune-mediated (asthma, inflammatory bowel disease, T1D), and cardiovascular (early onset coronary disease, atrial fibrillation, and hemorrhagic stroke). In their quest to identify rare risk and protective variants in common disease, the centers aim further the understanding of genomic architecture; design effective rare variant association studies (RVAS); develop new sequencing technologies and pipelines; and informatics tools.

The GSP Analysis Centers (GSPAC) undertake novel analyses on CMG and CCDG data, develop new methods and tools, and collaborate with researchers within and outside the GSP. The GSP Coordinating Center (GSPCC) provides scientific leadership and expertise for GSP-wide initiatives and coordinates administrative and outreach activities.

In the final year of the program, the GSP will: build an imputation server; map genome variation; develop a standard for sharing RVAS summary statistics; create a common controls resource; train new investigators; collaborate with biobanks and model organism communities; build an annotation server; calculate and validate polygenic risk scores; and develop ethnicity-specific linkage disequilibrium and recombination maps.

GSP data, analyses, methods, tools, and the functional equivalence WGS pipeline are or will be available via: dbGaP, AnVIL, ClinVar, Matchmaker Exchange Network, gnomAD, BRAVO, and GitHub.

## **PgmNr 746: Systematic review of online media coverage of GWAS publications from 2005 to 2018: Insights on trends in media coverage, readability of websites, and mentions to key social issues using text mining.**

### **Authors:**

J.J. Morosoli <sup>1,2</sup>; L. Colodro-Conde <sup>1</sup>; F.K. Barlow <sup>2</sup>; S.E. Medland <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia;  
2) School of Psychology, University of Queensland, Brisbane, Queensland, Australia

---

We present a systematic review of online media coverage of GWAS since its inception, using statistical and text mining techniques. GWAS publications were identified via the NHGRI-EBI GWAS Catalogue and classified into phenotype categories using the ICD-10. Blog and news articles about these studies were identified using the Altmetric database. As of 19/09/18, there were 3557 GWAS studies on 1945 different traits. We were able to retrieve 5349 different English language websites that mentioned these publications. Only 41.4% of those websites offered original content (not identical to another website). The most researched areas in the GWAS literature have been non-disease traits (e.g., body mass index, educational attainment; N=1197), neoplasms (N=462), and behavioral disorders (N=371). Only 20.1%, 21.4%, and 15.9% of the works in these areas were mentioned in two or more websites, respectively. Both publications and online mentions have increased every year. Partial correlation between year and number of online mentions (given GWAS publications per year) was  $r = 0.80$ . A regression analysis on number of online mentions including year of publication, number of hits, impact factor, and sample size showed that all predictors were positively and significantly associated with the outcome. However, residuals were not constant across phenotype categories, suggesting different patterns in public interest. In particular, GWAS on non-disease traits, and behavioral and nervous system disorders, generated 1.5 times more attention than predicted. We conducted a topic modeling analysis of the websites' content, and identified to which topic each website most probably belonged. Then we searched for key terms associated with public understanding of genetics. The five most frequent topics were Alzheimer's, depression, breast and prostate cancer, intelligence, and asthma. We found that only a low percentage of websites mentioned 'insurance' (0.9%), 'ethic' (2.2%), 'discrimination' (0.9%), or 'eugenic' (1.0%). However, 'environment' was frequently mentioned (28.6%). These keywords were most frequently found in articles about intelligence, and about 23andMe. Finally, we estimated the complexity of the language used in the websites using traditional readability formulas. More than 95% of all online articles would require at least university level reading skills to understand them. Implications for science communication and genetic literacy campaigns will be discussed.

## PgmNr 747: Physician attitudes, confidence, and perceived utility following experience with newborn genomic sequencing in the BabySeq Project.

### Authors:

H. Smith<sup>1</sup>; S. Pereira<sup>1</sup>; J. Robinson<sup>1</sup>; L. Frankel<sup>2</sup>; K. Christensen<sup>3,4</sup>; R. Parad<sup>3,4,5</sup>; P. Agrawal<sup>3,5</sup>; T. Yu<sup>3,5</sup>; I. Holm<sup>3,5</sup>; A. Beggs<sup>3,5</sup>; R. Green<sup>3,4,6</sup>; A. McGuire<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor College of Medicine, Houston, TX; 2) University of Houston, Houston, TX; 3) Harvard Medical School, Boston, MA; 4) Brigham and Women's Hospital, Boston, MA; 5) Boston Children's Hospital, Boston, MA; 6) The Broad Institute of MIT and Harvard, Cambridge, MA

Physicians' perceived utility and confidence in managing results of newborn genomic sequencing (nGS) compared to conventional newborn screening (NBS) are unknown. We studied physician perceptions as part of the BabySeq Project, a trial in which newborn participants were randomly assigned to receive either NBS and a family history report alone (control arm) or with a nGS report (sequencing arm). We enrolled physician participants from the same hospital units as newborns, some of whom had patients who were BabySeq participants. At the end of the study, we surveyed physicians (n=95; 27% response rate) to study whether receipt of nGS report (n=31) was associated with differences in perceived utility.

Physicians perceived NBS to currently be more useful than nGS to identify patients' conditions (mean 9.0 vs. 6.6 on a 1-10 scale;  $p<0.001$ ) and manage care (7.9 vs. 5.9;  $p<0.001$ ) but similarly useful to predict future risk of disease (6.3 vs. 6.1;  $p=0.51$ ). Physicians' perceived nGS utility did not differ by exposure to BabySeq nGS reports (all  $p>0.11$ ). However, respondents expected the utility of nGS to increase for each purpose over the next decade (all  $p<0.001$ ). Respondents predicted that 10 years in the future the utility of NBS and nGS will be similar to identify conditions (8.7 vs. 8.7;  $p=0.75$ ) and manage care (8.2 vs. 8.1;  $p=0.84$ ), while NBS is anticipated to be less useful than nGS to predict future disease risk (7.4 vs. 8.5;  $p<0.001$ ).

Only 27% of physicians reported being somewhat or very confident in their ability to make recommendations based on a positive genomic sequencing result, which was in contrast to 92% ( $p<0.001$ ) for NBS and not associated with receipt of a nGS report ( $p=0.22$ ). While all respondents agreed that every newborn should receive NBS, only 12% agreed that every newborn should receive nGS ( $p<0.001$ ), although a majority (59%) of physicians agreed there are health benefits associated with nGS.

Among physicians studied as part of the BabySeq project, experience with an nGS report is not associated with differences in perceived nGS utility at the end of the study. Physicians perceive NBS to have higher utility than nGS and are more confident in their ability to interpret and make recommendations based on results from NBS than nGS. Physicians expect the utility of nGS to increase in the next 10 years and feel that there are potential health benefits of nGS, attitudes which did not differ based on nGS report receipt.

## **PgmNr 748: The use of public genealogy databases in scientific research: A proposed approach to terms of use (TOU) authorization.**

### **Authors:**

J. Contreras <sup>1,2</sup>; C. Teerlink <sup>2</sup>; L. Cannon-Albright <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) College of Law, University of Utah, Salt Lake City, UT.; 2) University of Utah, School of Medicine, Salt Lake City, UT

---

In the wake of the recent Golden State Killer case, substantial public attention has focused on the use of public genealogy databases (PGD) such as GEDMatch by law enforcement authorities. In this paper, we address a different, but even more prevalent, use of PGD: scientific research. In recent years, PGDs have been used in dozens of epidemiological and genetic studies. But despite the success of these studies, it is not clear that the contractual Terms of Use (TOUs) associated with the PGDs from which data was obtained permit the conduct of this research.

We analyzed the TOUs associated with 38 major PGDs and assessed the degree to which they permit or prohibit the use of publicly-accessible data for scientific research. We found that many TOUs contain prohibitions on use, ranging from broad blanket prohibitions on all uses other than personal (non-commercial) genealogy research to more narrow prohibitions on commercial marketing, insurance underwriting and law enforcement uses. Moreover, no TOU that we examined expressly permits the use of available data for scientific research, making its use questionable under applicable data privacy laws. In all cases, the use of PDG data for non-permitted purposes could expose researchers to both civil and criminal penalties. And while we are not aware of any such legal action, rising public awareness of issues associated with PGDs increases the likelihood that such liability may arise in the future.

Accordingly, we propose a solution to enable scientific research using publicly-available PGD data without the threat of legal liability. It involves the development of a standardized rider that could be added to the TOU for all PGDs (PGD Rider). This approach is modeled after the SPARC Author Addendum developed by the Scholarly Publishing and Academic Resources Coalition to allow self-archiving of scientific journal articles. The proposed PGD Rider would expressly authorize scientific research using PGD data without addressing the more controversial issues associated with law enforcement and commercial uses. In addition, it would require researchers to observe data privacy and security measures with respect to personal data obtained from PGD sources.

We believe that the broad adoption of a PGD Rider would contribute to the continued use of these valuable data resources in scientific research without attempting to resolve the more challenging questions arising from law enforcement and commercial uses of this data.

## PgmNr 749: Increased demand for genomic medicine in the VA: Community care may not be the solution.

### Authors:

M. Scheuner<sup>1,2</sup>; A. Huynh<sup>3</sup>; C. Chanfreau-Coffinier<sup>4</sup>; B. Lerner<sup>5</sup>; O. Patterson<sup>4</sup>; L. Lewis<sup>4</sup>; M.M. Russell<sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) San Francisco VA Healthcare System, San Francisco, California.; 2) Departments of Pediatrics and Medicine, University of California, San Francisco; 3) VA Greater Los Angeles Healthcare System; 4) VA Salt Lake City Healthcare System; 5) VA Boston Healthcare System; 6) David Geffen School of Medicine at UCLA

**Objective:** As part of a larger initiative to examine coordination of genetic healthcare in the Department of Veterans Affairs (VA), we examined completion of genetic consults for VA patients.

**Methods:** We identified all genetic consults from 2010-17. Data extracted: consult status, patient characteristics, referral reasons, submitting and receiving sites, and attempts to schedule. Multiple logistic regression evaluated associations of variables with consult status.

**Results:** We identified 24,956 patients with genetic consults. Mean age was 50.6 (SD,14.9), 51% were female, and 63% white. Cancer was the most common (53%) referral reason. Most consults (58%) were received by the national VA program that provides telehealth consults; 28% were received by other VAs; and 14% were referred to the community (non-VA). Consults to VA were twice as likely to be completed than non-VA (national: OR=1.78;1.48-2.14; other: OR=1.90;1.54-2.33), with longer time to complete (mean days 162.3[154.8] vs 64.7[77.1],  $p<0.001$ ) and fewer scheduling attempts (mean attempts 1.11[0.39] vs 1.49[0.91],  $p<0.001$ ) in non-VA vs VA. Compared to non-VA, women and African Americans referred to the national VA program were less likely to complete consults (OR=0.73;0.61-0.87 and OR=0.69;0.57-0.83, respectively); this was not seen with other VAs. We did not find differences in completing consults for common referral reasons (cancer vs cardiovascular/connective tissue, gastrointestinal/polyps, or neuro-psychiatric). In contrast, less common referral reasons (e.g., immunology, endocrine, nephrology, reproductive) were 19% less likely to be completed (OR=0.81;0.70-0.95) compared to cancer. However, consults for less common referral reasons were more likely completed in the VA than non-VA (national: OR=1.70;1.41-2.06, other: OR=1.51;1.23-1.86).

**Conclusions:** Veterans are more likely to complete genetic consults if referred to the VA than the community. Completing a consult for less common referral reasons was greater in the VA, suggesting better access to a breadth of genetics expertise in the VA. Gender and race/ethnicity differences in completing consults with the national VA program might be due to patient preferences regarding telehealth. Future studies should address disparities that may increase with the growing number of women Veterans and community referrals. Implementation strategies should leverage the genetics expertise within the VA to improve care coordination and reduce disparities.

## **PgmNr 750: Clustering and classification analysis of Korean government-funded R&D projects and PubMed literature using machine learning techniques to support R&D investment decision making on biomedical fields.**

### **Authors:**

G. Yoo; H. Kim

[View Session](#) [Add to Schedule](#)

**Affiliation:** Korea Institute of Science & Technology Evaluation and Planning, 60 Mabang-ro, Seocho-gu, Seoul, Korea, Republic of

---

South Korean government's budget for R&D on biomedical technology exceeds 1 trillion won (850 million dollars), and deeper analysis of R&D investment is needed more than ever to support investment decision making on advanced biotechnology. However, the current practice of R&D investment portfolio management cannot reflect the rapidly changing biomedical technology trends as the analysis of R&D investment is merely based on simple statistics that utilizes fixed classification hierarchies. To solve this problem, we implemented a context search engine, a clustering analytic tool, and an automatic classifier for Korean national R&D projects and worldwide literature from PubMed using machine learning techniques. The purpose of this system is to provide more flexible and rapid tools for analyzing the status of national R&D investment and the global R&D trend (PubMed literature), and ultimately to help policy makers in investment decision making. First, all abstracts of the national R&D projects and the PubMed literature were embedded into 300-dimensional vectors through the doc2vec algorithm. Using the context search and the clustering analysis, which operate based on cosine similarities of the embedded vectors, a user can grasp the R&D investment status and knowledge structure of the desired technology field. We applied the system to the research projects and PubMed literature in the field of genetics between 2013 and 2017, and the results were consistent with common sense. For example, the amount of literature on whole-genome sequencing has been reduced, and clusters of literature such as genome-phenotype association studies and CRISPR gene editing expanded. Secondly, we implemented an automatic classifier for the national R&D projects using machine learning algorithms such as gradient boosting machine, multilayer perception, and support vector machine. This tool provides two most likely predictions according to the classification model trained with existing classified data. As a result, we achieved high accuracy rates (>80%) of classifying the R&D projects on drug development by the stage of development, the target disease, and the type of drug. We are planning to launch a Web service based on this system in the future for R&D policy makers.

## PgmNr 751: The yield of full *BRCA1/2* genotyping in Israeli Arab high-risk breast /ovarian cancer patients.

### Authors:

Y. Goldberg <sup>1,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Recanati Genetics Institute, Beilinson Hospital, Rabin Medical Center, Petach Tikva, Israel; 2) Breast Cancer Center, Oncology Institute, Chaim Sheba Medical Center, Tel-Hashomer, Israel; 3) Medical Genetics Institute, Meir Medical Center, Kfar Sava, Israel; 4) The Genetics Institute, Rambam Health Care Campus, Haifa, Israel; 5) The Ruth and Bruce Rappaport Faculty of Medicine, Technion – Israel Institute of Technology, Haifa, Israel; 6) Genetic Unit, Ziv Medical Center, Tzfat, Israel; 7) Faculty of Medicine, Bar Ilan University, Tzfat, Israel; 8) Center for Clinical Genetics, Hebrew University Hadassah Medical Center, Jerusalem, Israel; 9) Sharett Institute of Oncology, Hebrew University-Hadassah Medical Center, Jerusalem, Israel; 10) The Genetics Institute, The Emek Medical Center, Afula, Israel; 11) Maccabi Health Services, Rehovot, Israel.; 12) Davidoff Cancer Center, Rabin Medical Center, Beilinson Hospital, Petah Tikva; 13) Institute of Medical Genetics, Wolfson Medical Center, Holon, Israel; 14) Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel; 15) Susanne Levy Gertner Oncogenetics Unit, The Danek Gertner Institute of Human Genetics, Chaim Sheba Medical Center, Tel-Hashomer, Israel; 16) The Sackler School of Medicine, Tel-Aviv University, Tel-Aviv Israel

---

**Purpose:** While the spectrum of germline mutations in *BRCA1/2* genes in the Israeli Jewish population has been extensively studied, there is a paucity of data pertaining to Israeli Arab high-risk cases.

**Methods:** Consecutive Israeli Arab breast cancer (BC) or ovarian cancer (OC) patients were recruited using an ethically approved protocol from January 2012 to February 2019. All OC cases were referred for *BRCA* genotyping. BC patients were offered *BRCA* sequencing and deletion/duplication analysis after genetic counseling, if the calculated risk for carrying a *BRCA* mutation by risk prediction algorithms was  $\geq 10\%$ .

**Results:** Overall, 188 patients participated; 150 BC cases (median age at diagnosis: 40 years, range: 22-67) and 38 had OC (median age at diagnosis: 52.5 years, range: 26-79). Of genotyped cases, 18 (10%) carried one of 12 pathogenic or likely pathogenic variants, 12 in *BRCA1*, 6 in *BRCA2*. Only one was a rearrangement. Three variants recurred in more than one case; one was detected in five seemingly unrelated families. The detection rate for all BC cases was 4%, 5% in bilateral BC cases and 3% if BC was diagnosed < 40 years. Of patients with OC, 12/38 (32%) were carriers; the detection rate reached 75% (3/4) among patients diagnosed with both BC and OC.

**Conclusions:** The overall yield of comprehensive *BRCA1/2* testing in high-risk Israeli Arab individuals is low in BC patients, and much higher in OC patients. These results may guide optimal cancer susceptibility testing strategy in the Arab Israeli population.



## PgmNr 752: Hemochromatosis genotype does not increase colorectal cancer risk or age of onset.

### Authors:

G.P. Jarvik<sup>1</sup>; W. Wang<sup>2</sup>; P. Fontanillas<sup>3</sup>; A. Gordon<sup>1</sup>; L. Bastarache<sup>4</sup>; T. Harrison<sup>2</sup>; E. Rosenthal<sup>1</sup>; M. Palmer<sup>1</sup>; I. Stanaway<sup>1</sup>; E. Kim<sup>3</sup>; S. Gruber<sup>5</sup>; M. Jenkins<sup>6</sup>; U. Peters<sup>2</sup>; D. Crosslin<sup>1</sup>; 23andMe Research Team

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Washington, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA; 3) 23andMe, Mountain View, San Francisco, CA; 4) Vanderbilt Medical Center, Atlanta, GA; 5) University of Southern California, Los Angeles, CA; 6) University of Melbourne, Melbourne, Victoria, Australia

---

**Objective:** Homozygotes for the higher penetrance hemochromatosis risk allele, HFE C282Y (rs1800562), have been reported to be at a 2-3 fold increased risk for colorectal cancer (CRC; see PMIDS 26893171 and 23281741). These results have been reported in generally small sample sizes with no information about age of onset for such cancer. We evaluated a larger sample to assess HFE genotype effects on CRC risk and age of onset.

**Methods:** Using data from the Colorectal Cancer Translational Study and Genetics and Epidemiology of Colorectal Cancer Consortium, we evaluated distribution of CRC cases and age of onset in people of European ancestry with the HFE C282Y/C282Y genotype vs. those with zero C282Y or H63D alleles. These data included 52,151 CRC cases and 62,484 controls. We used logistic regression to estimate odds ratios (ORs) adjusted for age, sex, genotyping platform, and 3 principle components. As hemochromatosis penetrance is higher in males, analyses were repeated stratified by sex. Similar analyses were undertaken in 8606 self-reported CRC cases and 984,723 controls in the 23andMe cohort of consented research participants.

**Results:** The frequency of the HFE C282Y/C282Y genotype was 0.46% in CRC cases (240 of 52,151) and 0.54% in controls (340 of 62,484). The OR for CRC risk was 1.08 (95% CI: 0.91-1.29; p=0.39). Stratified analyses by sex identified no increased risk in males (OR: 0.94; p=0.64) or females (OR: 1.22; p=0.12). Age of CRC onset also did not differ by genotype; 64.2 years for the 199 C282Y/C282Y genotype individuals and 64.7 years for 34,832 non-carriers (p=0.6). Secondary analyses considering N=2,307 C282Y/H63D compound heterozygotes (967 CRC cases and 1340 controls) both separately or combined with the C282Y homozygotes vs. controls also detected no association with CRC risk or CRC onset age. Preliminary results in the 23andMe cohort suggest no increased CRC risk in C282Y homozygotes. Ongoing analyses considering hemochromatosis diagnosis suggest that penetrant disease has a modest increased CRC risk and no change in age of onset.

**Conclusion:** These are the largest analyses of the relationship between CRC and HFE risk and the only analyses of age of onset to date. We find no increased risk with genotype, but increased CRC risk with no age of onset change for those with penetrant disease. These results should be reassuring for those with HFE risk and suggest that these persons can follow population screening recommendations for CRC.



## PgmNr 753: Frequency of pathogenic variation in germline cancer susceptibility genes in rhabdomyosarcoma.

### Authors:

J. Kim<sup>1</sup>; N. Light<sup>2,3</sup>; V. Subasri<sup>2,4</sup>; E.L. Young<sup>5</sup>; T. Wegman-Ostrosky<sup>6</sup>; D. Barkauskas<sup>7</sup>; D. Hall<sup>7</sup>; P.J. Lupo<sup>8</sup>; R. Patidar<sup>9</sup>; L.D. Maese<sup>10</sup>; A. Shlien<sup>11,12</sup>; S.X. Skapek<sup>13</sup>; J.S. Wei<sup>9</sup>; X. Wen<sup>9</sup>; D.S. Hawkins<sup>14</sup>; J.D. Schiffman<sup>5,9</sup>; J. Khan<sup>9</sup>; D. Malkin<sup>2,3,4,15</sup>; D.R. Stewart<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, USA; 2) Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, ON, Canada; 3) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 4) Department of Medical Biophysics, University of Toronto, ON, Canada; 5) Department of Pediatrics, University of Utah, Salt Lake City, UT, USA; 6) Hereditary Cancer Clinic, Instituto Nacional de Carcerologia, Mexico; 7) QuadW-COG Childhood Sarcoma Biostatistics and Annotation Office, Children's Oncology Group, Monrovia, CA, USA; 8) Department of Pediatrics, Hematology-Oncology Section, Baylor College of Medicine, Houston, TX, USA; 9) Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, MD, USA; 10) Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT, USA; 11) Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 12) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON Canada; 13) University of Texas Southwestern, Dallas, TX, USA; 14) Division of Hematology and Oncology, Seattle Children's Hospital, University of Washington, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 15) Division of Hematology-Oncology, The Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, ON, Canada

**Introduction.** Rhabdomyosarcoma (RMS) is the most common childhood soft-tissue sarcoma. Genetic disorders associated with an increased risk of RMS include Li-Fraumeni, *DICER1*, NF1, Costello, Beckwith-Wiedemann, and Noonan syndromes. We investigated the frequency of pathogenic germline variation in discovery and secondary cohorts of individuals with RMS.

**Methods.** Patient blood-derived DNA samples were collected by Children's Oncology Group clinical trials and other institutional cohorts. Genome sequencing was conducted at The Hospital for Sick Children (discovery; n=273); exome sequencing was conducted at the National Cancer Institute (secondary; n=121). Variant pathogenicity in 310 candidate (syndrome, pathway, RMS-somatic, RMS-GWAS) genes was determined using ClinVar "badged labs," HGMD DM manual review and InterVar. The frequency of pathogenic/likely pathogenic (P/LP) variation was compared with non-TCGA gnomAD controls (n=118,479) using the Fisher exact test; multiple testing correction was applied using false discovery rate.

**Results.** In the discovery cohort, 8% harbored a statistically significant excess of P/LP variants compared with 0.5% in non-TCGA gnomAD controls in 11 genes (n=21): *CTC1*(3), *TRIM37*(2), *MSH2*(1), *L2HGDH*(2), *NF1*(4), *DICER1*(1), *TP53*(4), *SMARCA4*(1), *PTCH1*(1), *MLLT1*(1), *MYCN*(1). P/LP variants in *TP53*, *NF1*, *L2HGDH*, *MYCN* were significant after correction for multiple testing (q<0.05). In the secondary cohort, 10% harbored a statistically significant excess of P/LP variants compared with 0.3% in non-TCGA gnomAD controls in 11 genes (n=12): *RET*(1), *JAK3*(2), *TP53*(1), *L2HGDH*(1), *NHP2*(1), *MSH6*(1), *DIS3L2*(1), *TSC2*(1), *DICER1*(1), *IL1RN*(1), *BCOR*(1); none were significant after

correction for multiple testing. Excess P/LP variants in *TP53*, *DICER1*, *L2HGDH* were observed in both cohorts. Homozygous and compound heterozygous *L2HGDH* P/LP variation is associated with an excess of L-2-hydroxyglutarate, an epigenetic modifier and putative oncometabolite. Notably, there was an excess of P/LP variants in Lynch syndrome (*MSH2*, *MSH6*) and telomere biology genes (*CTC1*, *NHP2*) in both cohorts.

**Conclusion.** In the largest RMS exome/genome dataset to date, we observed in two independent cohorts an excess (8-10%) of P/LP variants, most commonly in known RMS-associated disorders (*NF1*, *TP53* and *DICER1*). Our findings suggest that Lynch and telomere biology genes are contributors to RMS germline risk. Risk associated with *L2HGDH* variation merits follow-up.

## PgmNr 754: Presymptomatic knowledge of BRCA carrier status: Impact on subsequent breast cancer stage, therapy and survival.

### Authors:

E. Levy-Lahad<sup>1,2</sup>; S. Lieberman<sup>1</sup>; P. Mor<sup>1</sup>; G. Amit<sup>2,3</sup>; E. Tahover<sup>4</sup>; O. Rosengarten<sup>4</sup>; M. Carmon<sup>3</sup>; O. Olsha<sup>3</sup>; R. Abu-Dalo<sup>3</sup>; R. Michaelson-Cohen<sup>1,2</sup>; E. Golomb<sup>5</sup>; S. Paluch-Shimon<sup>4</sup>; R. Rabinovitch<sup>6</sup>; T. Hadar<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel.; 2) Faculty of Medicine, Hebrew University Medical School, Jerusalem, Israel.; 3) Breast Surgery Unit, Department of General Surgery, Shaare Zedek Medical Center, Jerusalem, Israel.; 4) Institute of Oncology, Shaare Zedek Medical Center, Jerusalem, Israel.; 5) Department of Pathology, Shaare Zedek Medical Center, Jerusalem, Israel.; 6) Department of Radiation Oncology; University of Colorado Comprehensive Cancer Center, Denver CO, USA

**Background:** Screening healthy Ashkenazi Jews for germline BRCA1/BRCA2 mutations is not standard policy, despite high (2.5%) carrier rates. Most carriers are identified only after diagnosis of a BRCA-associated cancer.

**Aim:** To determine if pre-symptomatic knowledge of carrier status affects breast cancer stage, management and survival.

**Methods:** We reviewed medical records of BRCA1/BRCA2 carriers diagnosed with breast cancer between 1/2005-4/2016 at a single institution. We compared outcomes between women identified as BRCA carriers prior to their breast cancer diagnosis (Pre-Dx carriers) vs. those identified only after breast cancer diagnosis (Post-Dx carriers).

**Results:** Of 105 BRCA carriers with breast cancer, 42 (40%) were Pre-Dx and 63 (60%) were Post-Dx. None had had risk-reduction mastectomies. Mean age at diagnosis (50.4y) and BRCA1:BRCA2 distribution (64%:36%) were similar in both groups. Pre-Dx carriers were significantly more likely to have suggestive family history (93% vs. 63%). Pre-Dx carriers had lower clinical and pathological stage at diagnosis ( $p < 0.001$ ), with no differences in tumor grade, hormonal receptor or Her2 status. In Pre-Dx carriers, 38% of tumors were DCIS vs. 3.2% in Post-Dx carriers ( $< 0.001$ ). Pre-Dx carriers were less likely to receive chemotherapy (34% vs. 94%), and more likely to elect bilateral mastectomies (66% vs. 17%) than Post-Dx carriers ( $p \leq .001$ , all comparisons). In multivariate analysis, identification of carrier status prior to breast cancer diagnosis was the only significant predictor of early stage (0-I) breast cancer diagnosis (OR 10.6,  $p < 0.001$ ; age, calendar year of diagnosis, family history, and gene were all NS). 5-year overall survival in Pre-Dx carriers and Post-Dx carriers was 94% and 78%, respectively (HR=0.2,  $p = 0.04$ ).

**Conclusions:** Presymptomatic identification of BRCA carrier status is associated with significantly earlier stage at breast cancer diagnosis, less extensive treatment, and possible improvement in overall survival. These results support routine BRCA testing in the Ashkenazi Jewish population.

## PgmNr 755: Genetic determinants of recurrence in *BRCA1/2* germline mutation-associated breast and ovarian cancers.

### Authors:

J.B. Shah; B. Wubbenhorst; J. Pluta; C. Feltcher; K. D'Andrea; H. Symecko; C. Ruan; A. Nayak; K.N. Maxwell; S. Domchek; K.L. Nathanson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

*BRCA1/2* mutation-associated breast and ovarian tumors are sensitive to platinum-based chemotherapies and poly (ADP-ribose) polymerase inhibitors (PARPi). However, these tumors frequently return as therapy-resistant recurrences. To identify mechanisms of therapeutic resistance, we are evaluating genetic, transcriptomic, and immune features between matched primary and recurrent *BRCA1/2* tumors. We have compared whole exome sequencing from five pairs of *BRCA1* breast and nine pairs of *BRCA1/2* ovarian tumors, which is the largest such sample set collected to date. We analyzed multiple recurrences for five patients in this group. Ten patients received platinum-based chemotherapy, four patients received PARPi, and all patients received other forms of chemotherapy. We first compared homologous recombination deficiency (HRD) scores in matched tumors, finding that recurrences had significantly fewer large-scale state transitions (LSTs) compared to matched primary tumors ( $p=0.026$ ). We also identified a *BRCA1* reversion mutation, which was absent in the patient's first recurrence (after platinum treatment) but present in the second recurrence (after PARPi). Next, we assessed chromosomal copy number changes in matched tumors, after normalizing for cellularity and ploidy. We identified segments of chromosomal losses ( $CN=1$ ), gains ( $CN=4-5$ ), and amplifications ( $CN\geq 6$ ) that were present exclusively in recurrences. Recurrence-exclusive chromosomal copy number profiles showed low concordance between multiple recurrences from the same patient, but many large ( $>30\text{Mb}$ ) chromosomal losses were shared across several patients' recurrences, irrespective of cancer type or treatment. We also identified heterozygous, recurrence-exclusive losses in *MAD2L2* and *TP53BP1*. We found that shared gains and amplifications encompassed genes that could be pro-tumorigenic in the metastatic setting. Specifically, we noted gains and amplifications in lineage-related genes in  $\geq 4/14$  patients (*PAX1*, *PAX4*, *PAX6*, *EZH1*, *ELF5*, and two *KRT* gene clusters). We also found gains and amplifications in genes for tumor microenvironment factors (*FGFR2*, *ITGB4*, *CD44*), TNF $\alpha$  signaling (*TRAF6*, *MADD*), and DNA repair (*PARP1*, *PARP2*, *CHD2*, *EMSY*) in  $\geq 5/14$  patients. Next, we will use RNA-sequencing and immunohistochemistry to compare transcriptomes and immune milieus of matched *BRCA1/2* tumors. Ultimately, we seek to identify drivers of therapeutic resistance in *BRCA1/2* cancers, thereby informing novel treatment strategies.

## PgmNr 756: The genomic landscape of *de novo* mutations in first-generation carriers of TP53 germline mutations.

### Authors:

C. Vera Recio <sup>1</sup>; X. Pan <sup>1</sup>; J. Bojadzieva <sup>2</sup>; E. Dodd <sup>1</sup>; B. Li <sup>4</sup>; V. Johnson <sup>3</sup>; L. Strong <sup>1</sup>; W. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics and Computational Biologist, MD Anderson Cancer Center, Houston, Texas.; 2) Genetics, MD Anderson Cancer Center; 3) Statistics, Texas A&M; 4) Molecular and Biophysics, Vanderbilt University

---

Germline mutations are the driving force behind genome evolution and etiology of diseases. It has been well established that the number of *de novo* germline mutations is positively correlated with the age of the father and that of the mother. However, it remains largely unknown what factors other than perhaps parental ages would affect the initiation and accumulation of germline mutations. This question is particularly important for Li-Fraumeni syndrome (LFS), a cancer predisposition syndrome mostly caused by germline *TP53* mutations, in which there is an unusually high prevalence of *de novo* mutations (7-35%). To investigate this issue, we generated whole-genome sequencing data (at a 30x read depth) from five quartets of LFS probands who carry DNMs in *TP53*, healthy siblings and their parents, and compared the whole genomes of the probands with the siblings. Our cohort includes over ~8,600,000 single-nucleotide variants (SNVs) of which 702 are *de novo* mutations. We estimate a mutation rate of  $\sim 1.26 \times 10^{-8}$  SNVs per site per generation for *de novo TP53* mutation carriers and  $\sim 1.07 \times 10^{-8}$  SNVs per site per generation for the healthy siblings. Comparing probands and siblings we found a statistically significant increase in *de novo* SNV counts ( $p = 0.01$ , paired t-test). With the exception of *TP53*, probands do not carry more gene-disruptive mutations than their siblings, have a similar *de novo* mutation spectrum and were matched by age (within 4 years). Our findings suggest *de novo TP53* mutation carriers have an increase in the quantity but not quality of mutations in probands and differences in *de novo* mutation acquisition are unrelated to parental ages. Our study will advise on how disease causing and rare germline mutations are acquired and established in the genome.

## PgmNr 757: The contribution of MITF to melanoma and renal cancer.

### Authors:

E.A. Wiley <sup>1</sup>; N.J. Carter <sup>1</sup>; L. Susswein <sup>1</sup>; A.K. McGill <sup>1</sup>; B.A. Thompson <sup>1</sup>; R.T. Klein <sup>2</sup>; K.S. Hruska <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) GeneDx, Gaithersburg, Maryland.; 2) My Gene Team, Miami, Florida

---

**Background:** A single missense variant (c.952G>A, p.Glu318Lys) in microphthalmia-associated transcription factor (*MITF*) has been implicated in melanoma and renal cancer. Studies suggest that up to 2.8% of patients with melanoma and up to 3.0% of those with renal cancer harbor the *MITF* Glu318Lys variant. The aim of this work was to evaluate the yield of *MITF* Glu318Lys in individuals undergoing multi-gene hereditary cancer panel testing, paying particular attention to those with cutaneous melanoma and renal cancer.

**Methods:** We performed a retrospective review of all individuals undergoing panel testing of up to 64 genes, including *MITF* Glu318Lys. The personal histories of Glu318Lys carriers were reviewed.

**Results:** In total, 9,431 individuals underwent panel testing that included *MITF* Glu318Lys testing and of those, 43 (0.5%) were found to have *MITF* Glu318Lys. Seven of 544 probands (1.3%) with cutaneous melanoma were found to carry *MITF* Glu318Lys, with a mean age of diagnosis of 60 (51-71). Over half of those individuals also reported a family history of melanoma and one individual reported multiple primary melanomas. Those positive for *MITF* Glu318Lys did not carry another pathogenic variant in another melanoma-associated gene, including *CDKN2A* and *CDK4*. One thousand twenty-eight probands with a renal tumor underwent panel testing including *MITF* Glu318Lys and 5 (0.5%) were found to carry the pathogenic variant. Reported renal pathologies included clear cell, papillary type 1, and oncocytoma. Aside from melanoma and renal tumors, other frequent findings in the 36 affected individuals included breast cancer (n=15) and colon polyps (n=6). Breast cancer was observed in 15/33 (45%) female carriers with only one of these individuals carrying a pathogenic variant in another gene that may explain this diagnosis.

**Conclusions:** *MITF* Glu318Lys was observed in 1.3% of patients with melanoma and 0.5% of patients with a renal tumor in our cohort. Though our yields are lower than published cohorts, literature reports included either smaller cohorts or those enriched for probands with previously negative genetic testing for known melanoma/renal cancer susceptibility genes. Our findings suggest that there may be additional phenotypes associated with *MITF* Glu318Lys, but further evidence is needed before definitive conclusions can be drawn.

## PgmNr 758: Assessment of the clinical phenotype of *BAP1* germline whole gene and large deletions.

### Authors:

M.H. Abdel-Rahman <sup>1,2</sup>; R. Pilarski <sup>2</sup>; J.B. Massengill <sup>1</sup>; F.H. Davidorf <sup>1</sup>; C.M. Cebulla <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Ophthalmology and Visual Science, The Ohio State University, Columbus, Ohio.; 2) Division of Human Genetics, The Ohio State University, Columbus, Ohio

---

**Purpose:** Germline pathogenic mutations in *BAP1* are associated with a hereditary cancer predisposition syndrome with four main cancers: uveal melanoma, cutaneous melanoma, mesothelioma and renal cell carcinoma. Most of the reported cases were for patients with single nucleotide polymorphisms or < 3 base pair alterations. The aim of this study was to assess the clinical phenotype of germline whole gene and large deletions of the *BAP1* gene.

**Methods:** Whole gene deletions were diagnosed in clinical laboratories or in our laboratory using multiplex ligation probe amplification (MLPA) or Sanger sequencing. Segregation in family members was carried out by quantitative PCR using at least 2 different probes or by sequencing.

**Results:** We identified three probands and one relative with germline whole gene deletions and one proband and a relative with multi-exon deletion of *BAP1*. Five of them were female and one male. Average age was 39 years (range 16-62 years). Two subjects presented with uveal melanoma (ages 16 and 30 years), one with a uveal nevus (age (29 years), two with *BAP1*-inactivated melanocytic nevus/melanocytoma (ages 39 and 42 years) and one with two primary cancers: colon (54 years) and basal cell (60 years). None of the cases with whole gene deletion were identified by Sanger sequencing.

**Conclusion:** Whole gene deletion is an important mechanism for germline alteration in *BAP1*. Assessment for large deletions should be included in clinical testing. The clinical phenotype of patients with germline large deletion in *BAP1* is similar to the range of phenotypes in patients with other pathogenic *BAP1* mutations. Further assessments of the phenotype in additional subjects, as well as, study of environmental and genetic modifier of the phenotype are warranted.

## PgmNr 759: Mutation spectrum and genotype-phenotype analyses of pheochromocytomas/paragangliomas in Korea.

### Authors:

S. Cho <sup>1</sup>; M. Kim <sup>1</sup>; J. Kim <sup>2</sup>; K. Lee <sup>3</sup>; S. Kim <sup>3</sup>; S. Park <sup>1</sup>; M. Seong <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea; 2) Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea; 3) Department of Surgery, Seoul National University Hospital, Seoul, Korea

---

Pheochromocytomas (PHEOs) and paragangliomas (PGLs) are rare neuroendocrine tumors that produce sporadic catecholamine. More than 30% of these tumors were found to have inherited origin due to germ-line mutations in at least 10 genes including *RET*, *VHL*, *SDHB*, *SDHD*. New reports of phenotype-genotype correlation are emerging within PHEOs and PGLs depending on the gene and mutation location. We aimed to characterize the mutation spectrum and characteristics in Korean PHEOs/PGLs patients through analyzing causative genes. In this study, we included 171 patients with clinically confirmed PHEOs/PGLs who were treated between March 2009 and January 2019 at the Seoul National University Hospital in Korea. Targeted next-generation sequencing of 10 genes *MAX*, *NF1*, *RET*, *SDHA*, *SDHAF2*, *SDHB*, *SDHC*, *SDHD*, *TMEM127*, *VHL* were performed on 56 (32.8%) patients. 107 patients were tested only some of these 10 genes using Sanger sequencing and gene dosage analyses using multiplex ligation-dependent probe amplification. This study revealed that 20.5% of all cases (35 patients) harbored causative mutations in one of the PHEOs/PGLs genes: *RET* (34.3%, 12 patients), *VHL* (25.7%, 9 patients), *SDHB* (20%, 7 patients), *SDHD* (14.3%, 5 patients), *NF1* (2.9%, 1 patient), *NF2* (2.9%, 1 patient). *SDHB* mutations were usually found in high-risk PHEOs/PGLs. *SDHD* mutations were associated with extra-adrenal paragangliomas. *RET* D631Y variants were frequently observed compared with previous reports. This study indicated that the 4 genes including *RET*, *VHL*, *SDHB*, *SDHD* is the most important causative genes among Koreans, and four genes showed a specific spectrum of mutations in Korea.



## PgmNr 760: Parotid gland desmoid tumor as clinical presentation of familial adenomatous polyposis.

### Authors:

A.M. Gomez <sup>1</sup>; J.A. Rojas <sup>1</sup>; R. Quero <sup>1,2</sup>; M.C. Manotas <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Hospital Universitario San Ignacio, Bogota, Colombia; 2) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogota, Colombia

### Introduction:

Familial adenomatous polyposis (FAP [MIM 175100]) is an autosomal dominant hereditary cancer syndrome characterized by the development of multiple adenomatous colorectal polyps in the second or third decade of life and variably extracolonic manifestations such as hepatoblastoma, congenital hypertrophy of the retinal pigment epithelium, desmoid tumors, soft tissue tumors, dental anomalies, osteomas, odontomas, epidermoid cysts, duodenal adenomas and associated cancer. It is caused by mutations in the *APC* gene which almost always cause premature truncation of the protein.

### Case report:

We present the case of a 20-year-old woman with a desmoid tumor in the left parotid gland diagnosed at 15 years of age with recurrence at 17 and 19 years of age. A total left parotidectomy was performed with preservation of the facial nerve, with pathology reporting fibromatosis of desmoid type. Five years after the parotid tumor diagnosis, her father was diagnosed with colonic polyposis at 55 years old, and at that time FAP was suspected. The genetic study revealed a heterozygous frameshift mutation c.1977\_1978insTTTCT (p.Asn660Phefs\*12) in the *APC* gene, inherited from her father, confirming the diagnosis. Gastroscopy and colonoscopy were performed, showing numerous colonic polyps. Total prophylactic colectomy will be performed, and both the patient and family members are under multidisciplinary clinical follow-up. It is of note, that this patient also has a kidney malformation and history of epilepsy, ataxia and coagulation disorder in study. However, until now no other molecular studies have been carried out.

### Discussion:

It is well known the association of desmoid tumors with FAP, but these are usually found within the abdomen or in the abdominal wall. There are few reported cases of parotid tumors in this syndrome, and, to the best of our knowledge, just one case of a fibromatous parotid tumor. We consider important to highlight different forms of clinical presentations of FAP, as the parotid gland desmoid tumor in our patient, in order to make a timely diagnosis, since it is essential to provide an early management to prevent colorectal cancer that occurs inevitably in untreated FAP patients.

## PgmNr 761: A spontaneous rhesus macaque model of heritable colorectal cancer associated with mutations in DNA Mismatch Repair (MMR) genes: An improved animal model of Lynch syndrome in humans.

### Authors:

S.B. Gray <sup>1</sup>; M. Raveendran <sup>2</sup>; B.K. Dray <sup>5</sup>; R.A. Harris <sup>2</sup>; F. Benavides <sup>3</sup>; N. Ozirmak <sup>4</sup>; C.R. Abee <sup>1</sup>; E. Vilar <sup>4</sup>; J. Rogers <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Michale E. Keeling Center for Comparative Medicine and Research (KCCMR), The University of Texas MD Anderson Cancer Center, Bastrop, TX.; 2) Human Genome Sequencing Center and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Smithville, TX; 4) Department of Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, Houston, TX; 5) Charles River Laboratories, Ashland, OH

---

Colorectal cancer (CRC) is the third most common cancer in the US. Most CRC cases arise on a sporadic background, but Lynch Syndrome (LS) is the most common hereditary condition predisposing to CRC. LS is secondary to germline mutations in one of four DNA mismatch repair (MMR) genes: *MLH1*, *MSH2*, *MSH6* or *PMS2*. Heterozygous carriers have high lifetime risk to develop cancers commonly located in the proximal colon or ileocecal valve, as well as extracolonic tumors (endometrium, ovaries, stomach, small intestine and others). Since 2001, 65 cases of spontaneous CRC have been diagnosed among rhesus macaques (*Macaca mulatta*) at the Keeling Center for Comparative Medicine and Research (KCCMR). Rhesus CRC commonly presents as anemia, weight loss, diarrhea and a palpable abdominal mass. Mean age at death due to CRC is 18.3 years, roughly equivalent to mid-60's in humans. Histopathology of these tumors closely resembles human Lynch Syndrome. We previously reported two potential causative germline mutations identified by whole genome sequencing (WGS) of 20 of the 65 total CRC cases: a premature stop codon in *MLH1* and a missense mutation in *MSH6*, predicted to be highly deleterious (Combined Annotation-Dependent Depletion score of 22.8). Subsequent genotyping of all adults in the KCCMR rhesus colony (n = 595) revealed allele frequencies for the *MLH1* premature stop codon and *MSH6* missense variants of 0.04 and 0.06 respectively, with *MLH1* and *MSH6* heterozygotes highly overrepresented in CRC cases ( $\chi^2$  test  $p < 0.01$ ). Human LS exhibits microsatellite instability (MSI), an increase in somatic indels in microsatellite repeats. Initial studies in rhesus CRC cases indicate high MSI in some but not all cases. Additional analyses of MSI are in progress. We draw three conclusions from these data: a) approximately half of KCCMR rhesus CRC cases can be described as LS based on the damaging mutations in *MLH1* or *MSH6* and clinicopathologic presentation; b) further analyses are needed to determine if other molecular changes (e.g. disruption of *MLH1* promoter methylation) may underlie CRC in cases not attributable to *MLH1* or *MSH6* variants; c) because attempts to model Lynch Syndrome in rodents by disrupting MMR genes do not generate pathology that parallels human LS, this spontaneous rhesus macaque model, due to natural mutations in MMR genes and displaying nearly identical pathology, will be uniquely valuable in efforts to develop improved immune therapies and prevention strategies.

## PgmNr 762: Whole exome sequencing identifies homozygous *MSH6* variants in patient presenting with multiple malignancies.

### Authors:

H. Holmes; A. Revah Politi; M. Ernst; V. Aggarwal; N. Stong; J. Cappell; S. Kernie; C. Aguilar Breton; K. Anyane-Yeboah; J. Schacht

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute for Genomic Medicine, Columbia University Irving Medical Center, New York, NY.

---

**Introduction:** Here we describe our experience using research whole exome sequencing (WES) in a pediatric patient with multiple malignancies admitted at the Columbia University Irving Medical Center/New York Presbyterian Pediatric Intensive Care Unit (PICU).

**Case description:** A 13-year-old male of Middle Eastern descent presented to the PICU with multiple brain masses and an abnormal brain MRI showing cortical tubers, as well as seizures and café-au-lait macules. A diagnosis of Tuberous Sclerosis Complex (TSC) was initially suspected. Research WES resulted in an unexpected diagnosis of Constitutional Mismatch Repair Deficiency Syndrome (CMMRD). CMMRD is a rare autosomal recessive condition caused by homozygous variants in the mismatch repair genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*), and is characterized by brain tumors, colon cancer, hematologic cancer, and café-au-lait macules. Interestingly, microsatellite instability (MSI) testing ordered on a glioma sample from the patient was MSI negative, a phenomenon only observed in 5% of tumors caused by pathogenic variants in a mismatch repair gene. Immunohistochemistry (IHC) testing following the WES results confirmed the lack of expression of the MSH6 protein. Major clinical implications following the diagnosis of CMMRD included a change in the patient's chemotherapy treatment, as well as the identification of an additional signet cell peritoneal/intestinal tumor in the patient's stomach. Social implications included the diagnosis of the patient's mother with Lynch Syndrome, as well as increased risk of Lynch Syndrome and CMMRD to the patient's father and two siblings respectively.

**Conclusion:** Although a rare condition, CMMRD should be considered in patients with multiple brain masses and other malignancies given the potential for targeted treatment as well as future screening for the patient and family members. Targeted molecular testing for TSC and IHC testing alone would have missed this diagnosis. Whole exome sequencing is a useful tool in identifying the etiology of multiple malignancies in the pediatric setting.

## **PgmNr 763: In-depth dissection of *APC* pathogenic variants: Spectrum of more than 400 pathogenic variants, challenges of variant interpretation, and new observations in a large clinical laboratory testing cohort.**

### **Authors:**

H.C. Kang; D. Pineda; C. Pardo; D. Iacoboni; E. Borrás; I.Y. Song; S. Michalski; S. Yang; S.H. Wang; S. Crowley; M. Manansala; M. Harrell; Y. Lu; M.J. Anderson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Invitae Corporation, San Francisco, CA.

---

*APC* as a classic tumor suppressor gene is involved in the initial tumorigenesis of colorectal cancer, and is responsible for inherited polyposis syndromes, familial adenomatous polyposis (FAP), attenuated FAP, and gastric adenocarcinoma and proximal polyposis of the stomach (GAPPS). Since *APC* was first discovered about three decades ago, numerous studies have reported germline *APC* variants and associated clinical findings in the literature. With recent technological advances in genomic sequencing, *APC* genetic testing is routinely included as part of hereditary cancer panel testing and easily performed in clinical laboratories. While *APC* has been extensively studied, we continue to encounter novel variants given the large number of individuals tested in our clinical laboratory.

Here we report an in-depth dissection of *APC* variants among over 208,000 individuals tested clinically, which includes analysis of over 400 unique pathogenic variants. *APC* comprises 16 exons, with a large last exon 16 encoding ~80% of the *APC* protein. *APC* also has well-characterized regulatory regions, promoter 1A and promoter 1B. Germline variants in promoter 1B that disrupt transcription factor binding have been reported as disease causing. Among the pathogenic variants, 88% of all variants were truncating variants, 4% were missense, silent and intronic variants affecting splicing, 6% were subgenic copy-number variants (CNVs), ~1% were regulatory variants in promoter 1B, and ~1% were exonic retrotransposon (Alu and LINE-1) insertions. Due to the high penetrance of disease, variant interpretation of novel variants is less complicated. However, there are some challenges: 1) interpretation of novel truncating variants escaping nonsense-mediated decay (NMD) in the large exon 16 requires additional evidence; 2) interpretation of variants affecting regulatory regions requires more rigorous evidence to prove pathogenicity; 3) interpretation of subgenic duplications that are assumed in-tandem is accounted for by clinical information and penetrance. Nonetheless, experience over the years has helped us to resolve some of these challenges, which we will cover in this study. Lastly, we report new observations from our clinical testing: 1) pathogenic germline variants affecting the YY1 motif of promoter 1B in GAPPS and FAP families; and 2) subgenic duplications of the 5' end of the gene involving promoter 1A and/or exon 2, but excluding promoter 1B, that are likely disease-causing variants.

## PgmNr 764: Novel presentation of jaw odontogenic cyst in *SUFU*-associated Gorlin syndrome.

### Authors:

J.A. Kennedy <sup>1</sup>; S. Lala <sup>1,2</sup>; M. Harlan Fleischut <sup>1</sup>; E. Fiala <sup>1</sup>; I. Dunkel <sup>2</sup>; E. Tonorezos <sup>1</sup>; C. Estilo <sup>3</sup>; M. Lacouture <sup>1</sup>; Z. Stadler <sup>1</sup>; Y. Khakoo <sup>2</sup>; S. Gilheeney <sup>2</sup>; M. Karajannis <sup>2</sup>; K. Offit <sup>1</sup>; M. Walsh <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY.; 2) Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY; 3) Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY

---

Gorlin syndrome (GS), also known as nevoid basal cell carcinoma syndrome, is an autosomal dominant benign and malignant tumor predisposition. GS is characterized by jaw keratocysts, ectopic calcification of the falx, skeletal anomalies, plantar/palmar pitting, cardiac and ovarian fibromas, and in some individuals, distinctive facial features. Tumor risk includes medulloblastoma (particularly desmoplastic type) and multiple basal cell carcinomas. Germline mutations in components of the Sonic hedgehog pathway, *Patched1* (*PTCH1*) and *Suppressor of fused* (*SUFU*) are responsible for GS. Jaw keratocysts are well described among *PTCH1* carriers but not described in *SUFU* mutation carriers. Given the morbidity associated with untreated cysts, digital surveillance of the jaw is recommended.

Historically, GS surveillance has been indifferent to genotype (*PTCH1* vs. *SUFU*). Foulkes et al. published the first protocol suggesting surveillance by genotype. This included elimination of odontogenic cyst screening in patients with *SUFU* mutations. Following this, Evans et al. reported genotype/phenotype data of 182 individuals meeting clinical GS diagnostic criteria, in which 9 individuals had *SUFU* mutations. None of the *SUFU* mutation carriers were reported to have jaw cysts.

To the best of our knowledge, we document the first case of a jaw keratocyst in an individual with *SUFU*-related GS. Our patient is a 33-year-old Hispanic male who presented at age four with medulloblastoma. He underwent subtotal resection, craniospinal radiation and chemotherapy. At age 18, an osteochondroma of the lumbar vertebrae was identified and resected. At age 20, he was found to have an infundibulocystic basal cell carcinoma of the left periapical scalp and seizures. At 23, his physical exam was significant for palmar pits and an odontogenic cyst that was excised. His history also includes bilateral cataracts, subsequent basal cell carcinomas, meningioma at age 30 and a transient ischemic attack, likely radiation-induced. Germline testing of *PTCH1* and *SUFU* revealed a novel heterozygous likely pathogenic *SUFU* mutation, c.595C>T, (p.Gln199\*).

Previous series of *SUFU*-related GS have limited numbers of patients given the rarity of the condition. Therefore, the finding of a jaw keratocyst in our patient is notable and suggests that jaw cysts may also be associated with germline *SUFU* mutations; potentially warranting digital jaw surveillance.

## PgmNr 765: A germline susceptibility variant in *VHL* in a patient with Maffucci syndrome.

### Authors:

C. Smith <sup>1</sup>; S. Robbins <sup>2,3</sup>; R. Martin <sup>2</sup>; D. Valle <sup>2</sup>; N. Sobreira <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) SOM Medical Genetics Clinic, Johns Hopkins Univ School of Medicine, Baltimore, Maryland.; 2) McKusick-Nathans Institute in the Department of Genetic Medicine, Johns Hopkins Univ School of Medicine, Baltimore, Maryland; 3) Predoctoral Training Program in Human Genetics, Johns Hopkins Univ School of Medicine, Baltimore, Maryland.

---

*VHL* downregulates HIF1A, the main regulator of adaptation to hypoxia, by targeting HIF1A for degradation. Loss-of-function variants in *VHL* are known to cause von-Hippel Lindau syndrome, an autosomal dominant familial cancer syndrome, and autosomal recessive familial erythrocytosis. *VHL* is also associated with other tumors, such as cerebellar hemangioblastoma, pheochromocytoma, and renal cell carcinoma. Maffucci syndrome (MS) is a rare disorder characterized by multiple enchondromas and vascular anomalies. Patients with MS are also at increased risk for cancer - mainly, chondrosarcomas and vascular malignancies. Here we describe a patient with MS and additional features not described before, such as a region of epidermal and dermal atrophy on the left chest, a lymphoepithelioma of the nasopharynx with parapharyngeal involvement, an exostosis in the left wrist, developmental delay and intellectual disability. We performed WES on multiple samples, including blood, an enchondroma and an exostosis. A germline *VHL* variant (c.C505T; p.R169W) was identified by WES. RNA-seq confirmed that the *VHL* variant is present at heterozygous levels (45-70%) in RNA from all tissues tested: an enchondroma, an exostosis, a vascular anomaly biopsy, and an unaffected skin biopsy. Western Blot showed that *VHL* in the patient's unaffected skin biopsy and vascular anomaly biopsy is within the range of the controls. WES also identified a somatic mosaic variant in *PDGFRB* (c.C1998A; p.N666K), known to be pathogenic and causative of somatic infantile myofibromatosis. This variant was identified in the enchondroma but not in the exostosis and in the blood. We performed digital droplet PCR and confirmed that the *PDGFRB*-p.N666K variant was not present in the blood, but was present in all 4 patient's enchondromas tested (21.8%, 27.5%, 50.5%, and 49.8%), in his in his exostosis (0.64%), in his unaffected skin biopsy (22%) and vascular anomaly biopsy (0.88%). Based on these findings we suggest that the *VHL*-p.R169W is likely pathogenic, and that, similarly to the *VHL*-R200W and the *VHL*-H191D variants that result in polycythemia, it causes overproduction of the HIF-targets by attenuating formation of the E3 ubiquitin ligase and attenuating binding to HIF1. We also suggest that this patient has a tumor predisposition syndrome characterized by the germline susceptibility *VHL*-p.R169W variant and additional tumor variants, such as the *PDGFRB*-p.N666, leading to tumor formation.



## PgmNr 766: Germline variants in predisposition genes in children with Down syndrome and acute lymphoblastic leukemia.

### Authors:

A.J. de Smith <sup>1</sup>; P. Winer <sup>1</sup>; I.S. Muskens <sup>1</sup>; K.M. Walsh <sup>2</sup>; A. Vora <sup>3</sup>; J.L. Wiemels <sup>1</sup>; A.V. Moorman <sup>4</sup>; I. Roberts <sup>5</sup>; A. Roy <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Epidemiology, Department of Preventive Medicine, University of Southern California, Los Angeles, California, USA; 2) Department of Neurosurgery, Duke University, Durham, North Carolina, USA; 3) Great Ormond Street Hospital for Children NHS trust, London, UK; 4) Wolfson Childhood Cancer Research Centre, Northern Institute for Cancer Research, Newcastle University, Newcastle-upon-Tyne, UK; 5) Department of Paediatrics and MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, Oxford University and BRC Blood Theme, NIHR Oxford Biomedical Centre, Oxford, UK

**Background:** Despite advances in treatment, acute lymphoblastic leukemia (ALL) remains a leading cause of childhood mortality. Children with Down syndrome (DS), caused by constitutive trisomy 21, have a 20-fold increased risk of ALL (DS-ALL) and significantly lower event-free survival than non-DS patients with ALL. There is a critical need to understand the increased risk of ALL in children with DS. We determined the frequency of rare predisposing germline variants in 77 DS-ALL patients in the International Study of Down Syndrome Acute Leukemia.

**Methods:** We identified 59 DS-ALL patients in the Bloodwise Childhood Leukemia Cell Bank (UK) and 18 patients in the New York State Department of Health Newborn Screening program, with available germline material. Whole exome sequencing of extracted DNA was performed at 60X coverage. FASTQ files were aligned to human reference genome hg38 using BWA, and variant calling performed following GATK “best practices” guidelines. For quality control, we used Variant Quality Score Recalibration for non-chromosome 21 variants, and hard filtering for chromosome 21 variants. After annotating variants using ANNOVAR, we retained those with allele frequency <0.01% in the Genome Aggregation Database, and limited to loss-of-function variants (nonsense, frameshift, splicing), and/or those classed as “Pathogenic” or “Likely pathogenic” in the ClinVar database.

**Results:** At least 5/77 (6.5%) DS-ALL patients harbored rare and predicted pathogenic germline variants; a similar frequency was reported in non-DS ALL. Three variants were previously described: a deleterious missense variant in *IKZF1*, associated with immunodeficiency and childhood ALL; a pathogenic frameshift mutation in the Nijmegen breakage syndrome gene *NBN*; and a missense variant in *RTEL1*, reported previously in patients with dyskeratosis congenita. Two additional variants were identified in genes without a known role in leukemia predisposition: a loss-of-function frameshift mutation in the B-cell transcription factor gene *FOXP1*, and a mutation in the leukemia fusion gene *MLLT1* listed as likely pathogenic in ClinVar in a patient with congenital abnormalities. No putative pathogenic mutations were found on the trisomic chromosome 21.

**Conclusions:** Children with DS-ALL may harbor pathogenic germline variants in predisposition genes, likely contributing to their leukemia development, which may inform future risk stratification and genetic counseling in this vulnerable population.

# PgmNr 767: Copy number burden characterization and prediction across human cancers.

**Authors:**

B. Lasseigne <sup>1</sup>; R. Darji <sup>2</sup>; J. Fisher <sup>3</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Cell, Developmental and Integrative Biology, The University of Alabama at Birmingham School of Medicine, Birmingham, Alabama.; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 3) Mississippi State University, Starkville, MS

---

**Introduction**

Chromosomal instability, or altered chromosome number and structure, is an enabling characteristic of the hallmarks of cancer and is associated with disease etiology, patient prognosis, and drug efficacy. However, copy number burden has been calculated by many different metrics and the relationship between them, or which are most biologically relevant with respect to context or biological question, is unknown.

**Methods**

Using The Cancer Genome Atlas cancer cohort (n=22,158 patients, 33 cancers), we calculated, compared, and characterized copy number burden by metrics including total aneuploidy index, copy number aberration count, the percent of bases with copy number variation, and the total number of breakpoints.

**Results**

We demonstrate these metrics are not equivalent, occur at different frequencies within and across cancer types, and likely represent distinct biological properties like homologous repair defects and selective pressure for copy number regions with driver genes. We also identified molecular features including DNA methylation, transcription factor occupancy profile overlap, gene ontology term enrichment, and proximity to tissue-specific eQTLs associated with each chromosomal instability metric both within and across cancer types, as well as clinical factors (e.g., patient age, tumor stage) and drug responses associated with each metric. Additionally, because chromosomal instability is linked to widespread DNA hypomethylation and characteristic gene expression changes even across species, we hypothesized measures of chromosomal instability can be inferred from these and other data types. Therefore, we predicted chromosomal instability metrics from each TCGA data type using machine learning approaches, consistently building models with favorable performance characteristics.

**Discussion**

The models we generated are being implemented as an R package useful for inferring different chromosomal instability metrics from genomics data sets where copy number variation data may not



be available, enabling more information to be extracted from available molecular profiles. This study improves chromosomal instability signature interpretation by providing insight into the role of chromosomal instability in human cancer etiology, progression, and potential treatment strategies, while providing models for predicting it in other data types like gene expression and DNA methylation.

## PgmNr 768: LD score regression analysis of lung cancer using multi-traits from UK Biobank.

### Authors:

J. Edelson<sup>1,2</sup>; J. Byun<sup>1,2</sup>; Y. Han<sup>1,2</sup>; Q. Ostrom<sup>2</sup>; K. Walsh<sup>3</sup>; M. Bondy<sup>2</sup>; C. Amos<sup>1,2</sup>; Integral-ILLCO group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX.; 2) Department of Medicine, Epidemiology and Population Sciences, Baylor College of Medicine, Houston, TX; 3) Department of Neurosurgery, Duke University, Durham, NC

Identifying genetic correlations between complex traits and diseases can provide valuable insights into epidemiological and etiological studies. The major challenge for estimating genetic correlation from genome-wide association study (GWAS) data is the lack of availability of individual-level genotype data and sample overlap among meta-analyses. Linkage disequilibrium (LD) score regression using GWAS summary statistics quantifies the genetic correlation and SNP heritability, which is proportion of phenotypic variance explained by all SNPs between pairs of traits. This technique could provide insights into the genetic contributions of different phenotypes potentially associated with lung cancer risk.

We harmonized 824 publicly-available GWAS summary-level datasets from 36 consortia including 82 diseases, 154 complex traits, 181 metabolites, and 12 immune-related disorders from the UK Biobank and OncoArray lung consortium. We calculated pairwise genetic correlation between lung cancer including histologic subtypes and other phenotypes.

Results demonstrated that smoking and socio-economic status related traits are the strongest factors associated with an increased risk of lung cancer. The correlation between smoking and overall lung cancer was 0.42 ( $P=1.70E-84$ ,  $h^2=0.10$ ) and a negative correlation was observed between lung cancer and education years, -0.34 ( $P=5.10E-63$ ,  $h^2=0.15$ ). Additionally, primary biliary cirrhosis (PBC), ulcerative colitis (UC), and multiple sclerosis (MS) all showed positive correlation with lung adenocarcinoma. To correct for the contribution of smoking to these associations, chromosomal regions previously associated with cigarettes per day, smoking cessation and age at initiation were excluded (500kb regions centered on SNPs identified by Liu et al. (Nature Genetics, 2019)). After removing these regions, the greatest genetic correlation across all lung cancer subtypes was a paternal history of lung cancer ( $r^2 = 0.77$ ,  $P= 3.63E-34$ ,  $h^2=0.01$ ), followed by PBC ( $r^2 =0.58$ ,  $P= 9.44E-31$ ,  $h^2=0.74$ ), UC ( $r^2=0.39$ ,  $P=2.56E-24$ ,  $h^2=0.34$ ) and MS ( $r^2=0.45$ ,  $P=1.39E-18$ ,  $h^2=0.18$ ). These results are consistent with previous research and expand prior findings by identifying novel phenotype correlations with several autoimmune diseases. Mendelian randomization analyses will help clarify the potentially causal relationship between autoimmune diseases and lung cancer risk.

## PgmNr 769: An integrative multi-omics network-based approach identifies key regulators for breast cancer.

### Authors:

Y. Guo; Y.X. Chen; S.S. Dong; H.M. Niu; Y.J. Zhang; F. Jiang; Y. Rong; T.L. Yang

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China

---

Breast cancer is one of the most common diagnosed cancers and leading causes of cancer-related death in females. Several GWASs have identified over 600 breast cancer susceptibility SNPs, but most SNPs are located within non-coding genomic regions and their functions need to be interpreted. We developed a framework named integrative genomics network-based approach (IGNA) to identify potential regulatory genes based on disease-related biological processes and the network topology. This approach was applied to identify key genes in regulatory networks for breast cancer by integrating interactions and functional data from multiple biological scales, including GWASs, eQTLs, epigenomic elements, transcriptome, protein interactome and chromatin long-range interactions from disease relevant tissues. We collected 1,077,575 SNP-gene mapping pairs by integrating GWASs results with SNPs functional information. To identified biological related gene sets of breast cancer, we collected knowledge-driven pathway and co-expression modules to investigate the genetical perturbations on the regulatory networks. Across 5 breast cancer gene expression profiles, we identified co-expression 357 modules. We gathered 1077 canonical pathways from public databases. Among the total 1,434 gene sets, 12, 31 and 108 gene sets in African, East Asian, European population significantly enriched in disease-associated SNPs with P-value < 0.05. We found 185 overlap of genes across 3 populations and performed pathway enrichment analysis to find the contribution of identified gene sets to biological mechanisms. The top ranked pathways were cancer-related pathways and the shared genes showed the enrichment in several important signaling pathways, cellular processes and viral infection pathways. Moreover, we uncovered key genes for each regulatory network based on GWAS signals and the topology of networks. The top 20 ranked-order genes are *MYC*, *CCND1*, *CDKN1A*, *SMAD3*, *DUSP1*, *RB1*, *STAT1*, *EGFR*, *TGFB1*, *IGF1R*, *TP53*, *STAT3*, *CFLAR*, *MDM2*, *NFKBIA*, *RHOB*, *VEGFA*, *SMAD2*, *GADD45B* and *JUND*. Moreover, we found that abnormal expression levels of the top 20 ranked-order genes were significantly related with poor survival outcomes (P=0.025). Our integrative network-based approach provided a genetic-driven framework to unveil tissue-specific interactions from multiple biological scales and to reveal potential risk genes to complex diseases.

## PgmNr 770: Multiple pathogenic genetic variants in a Puerto Rican (Hispanic) patient with ovarian cancer: Unreported variants and the implications of *de novo* haploinsufficiency in cancer development.

### Authors:

E. Albino <sup>1</sup>; N.J. Arciniegas <sup>2</sup>; S. Carlo <sup>3,4</sup>; A. Cornier <sup>3,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Health Professions, University of Puerto Rico - Medical Sciences Campus, San Juan, PR, USA; 2) Department of Pediatrics, Ponce Health Sciences University, Ponce, PR, USA; 3) Genetics Section, San Jorge Children's Hospital, San Juan, PR, USA; 4) Department of Biomedical Sciences, Ponce Health Sciences University, Ponce, PR, USA; 5) Department of Public Health, Ponce Health Sciences University, Ponce, PR, USA; 6) Department of Pediatrics, Universidad Central del Caribe School of Medicine, Bayamon, PR, USA

---

Forty-seven years old female patient with history of ovarian carcinoma and a pathogenic variant in *BRCA2* gene, c.3922G>T (p.Glu1308\*) was referred to Genetics for evaluation and counseling. Family history was significant for thyroid, lung and ovarian cancer in three generations. A multi-gene cancer panel was requested and the pathogenic variant in *BRCA2* was confirmed. This variant is associated with autosomal dominant hereditary breast and ovarian cancer syndrome, where this sequence change creates a premature translational stop signal, resulting in an absent or disrupted protein product. In addition, a pathogenic variant, c.1187G>A (p.Gly396Asp), was identified in *MUTYH*, which is associated with autosomal recessive *MUTYH*-associated polyposis. Because *MUTYH*-related conditions are autosomal recessive, one pathogenic variant alone may not explain individual's reported condition. However, there is evidence that this monoallelic pathogenic *MUTYH* variant is associated with colon cancer. Also, two variants of uncertain significance were reported, c.4532A>T (p.Asp1511Val) in *ATM* and c.2083G>A (p.Val695Ile) in *BARD1*. The *ATM* gene is associated with an increased risk for autosomal dominant breast, pancreatic and prostate cancers. This variant is not present in population databases but algorithms developed to predict effect of variations suggest that this variant is likely disruptive. The *BARD1* gene is associated with an increased risk for autosomal dominant breast and ovarian cancer in individuals who carry a single pathogenic variant. Further studies of these variants may prove to be an important step toward the clinical and molecular characterization of cancer in our population. Of particular interest is the effect of the loss of heterozygosity (haploinsufficiency) effect as a potentiator of the cancer development process. While targeted gene panels are indicated for analyzing specific mutations, multi-gene panel was an excellent approach in this case due to the clinical manifestations.

## PgmNr 771: Identification of novel potential genetic predictors of urothelial bladder carcinoma susceptibility in Pakistani population.

### Authors:

S. Ali <sup>1</sup>; K. Bangash <sup>2</sup>; A. Rauf <sup>3</sup>; M. Younis <sup>4</sup>; K. Anwar <sup>2</sup>; R. Khurram <sup>1</sup>; M. Khawaja <sup>5</sup>; M. Azam <sup>1</sup>; A. Qureshi <sup>6</sup>; S. Akhter <sup>5</sup>; L. Kiemeny <sup>7</sup>; R. Qamar <sup>1,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biosciences, Faculty of Science, COMSATS University Islamabad, Park Road, Tarlai Kalan, Islamabad 45600, Pakistan; 2) Pakistan Atomic Energy Commission General Hospital, Islamabad 44000, Pakistan; 3) Nishtar Medical College & Hospital, Multan 60000, Pakistan; 4) Department of Urology, Mayo Hospital, Lahore 54000, Pakistan; 5) Department of Urology, Shifa International Hospital, Islamabad 44000, Pakistan; 6) Department of Radiology, The Children's Hospital & The Institute of Child Health, Lahore 54000, Pakistan; 7) Department for Health Evidence and Department of Urology, Radboud University Medical Centre, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands; 8) Al-Nafees Medical College & Hospital Isra University, Islamabad 45600, Pakistan; 9) Pakistan Academy of Sciences, Islamabad 44000, Pakistan

---

Urothelial bladder carcinoma (UBC) is the most common among urinary bladder neoplasms. We carried out a preliminary study to determine the genetic etiology of UBC in Pakistani population. For this purpose, 25 sequence variants from 17 candidate genes were studied in 400 individuals by using polymerase chain reaction-based techniques. Multivariate logistic regression analysis was performed for association analysis of the overall data as well as the data stratified by smoking status, tumor grade and tumor stage. Variants of *GSTM1*, *IGFBP3*, *LEPR* and *ACE* were found to be associated with altered UBC risk in the overall comparison. *CYP1B1* and *CDKN1A* variants displayed a risk modulation among smokers; *IGFBP3* and *LEPR* variants among nonsmokers while *GSTM1* polymorphism exhibited association with both. *GSTM1* and *LEPR* variants conferred an altered susceptibility to low grade UBC; *GSTT1*, *IGFBP3* and *PPARG* variants to high grade UBC while *ACE* polymorphism to both grades. *GSTM1* and *LEPR* variants exhibited risk modulation for non-muscle-invasive bladder cancer (NMIBC); *GSTT1* and *PPARG* variants for muscle-invasive bladder cancer (MIBC), and *ACE* variant for NMIBC as well as MIBC. In general, the susceptibility markers were common for low grade and NMIBC; and distinct from those for high grade and MIBC indicating the distinct pathologies of both groups. In brief, our results conform to reports of previously associated variants in addition to identifying novel potential genetic predictors of UBC susceptibility.

## PgmNr 772: Identification of differential expressed genes in HBV-related hepatocellular carcinoma.

### Authors:

P. An; B. Derrett; C.A. Winkler

[View Session](#) [Add to Schedule](#)

**Affiliation:** Molecular Genetic Epidemiology Section, Basic Science Program, Basic Research Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD

---

Hepatocellular carcinoma (HCC) is the third leading cause of deaths from cancer worldwide. Hepatitis B virus (HBV) infection and hepatitis C virus (HCV) infection are the leading risk factors for developing HCC, particularly in East Asia. There is lack of effective treatment and early detection methods for HCC. Genomic and transcriptomic profiles in HCC tumors are critical in understanding the mechanism of HCC development and providing guidance for novel therapeutic and biomarker development. There are limited transcriptomic studies in HBV-related HCC, most of which underpowered, do not differentiate between HBV- or HCV-associated HCC, and lack validation across studies. We performed a meta-analysis to assess changes in gene expression patterns in HBV-HCC by integrating HBV-HCC samples from TCGA (The Cancer Genome Atlas) and four GEO (Gene Expression Omnibus) datasets. Each of the five datasets had 1000-7000 differentially expressed genes (DEGs) from the tumor/normal tissue comparison (threshold >1.5-fold change, FDR p-value <0.05). In the HBV-HCC TCGA dataset, there were 6603 DEGs. There were 512 DEGs (328 over-expressed, 184 under-expressed) shared across all five datasets in the same direction, indicating 7.8% of DEGs from TCGA HBV-HCC can be replicated. Among top DEGs, several are known to be involved in HCC, e.g. the over-expression of *GPC3* and *PTTG1*, and the under-expression of *CYP1A2* and *CLEC4G*, while most others' role in HBV-HCC have not been defined, including *LYVE-1* involved in tumor metastasis, and *KCNN2*, an ion transporter. Pathway analysis based on overlapping genes revealed that toxin and fatty acid degradation pathways were highly under-expressed (including cytochrome P450 family enzymes that metabolize and degrade toxins, drugs and lipids, e.g. *CYP2A6*) while cell cycle replication and retinoid X receptor (RXR) inhibition pathways are over-expressed (including *Cyclins*, *SLC10A1*, *ALDH8A1*). These identified pathways and genes may have utility in the development of targeted therapies and prevention of HCC. (Funded by the National Cancer Institute, under contract HHSN26120080001E).

## PgmNr 773: Decreased expression of *ZEB1* and *SOX2* in glioma in comparison with meningioma.

### Authors:

S. Bayat<sup>1</sup>; M.T. Raouf<sup>2</sup>; N. Mousavi<sup>1</sup>; S. Panjeh-Shahi<sup>1</sup>; F. Vand-Rajabpour<sup>1</sup>; M. Yaseri<sup>3</sup>; A. Khoshnevisan<sup>2</sup>; M. Tabrizi<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics Department, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran; 2) Department of Neurosurgery, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran; 3) Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran

Gliomas are the most common primary malignant brain tumors in adults. Their aggressiveness results from their high invasive ability which makes their total surgical resection impossible and renders high grade glioma the capability to relapse. Epithelial-to-Mesenchymal Transition (EMT) has been considered as one of the major mechanisms helping glioma aggressiveness. EMT is a complex process modulated by multitude of transcription factors and signaling pathways. This process is known to be implicated not only in malignancy progression, but also in cancer therapy resistance.

Meningioma, the tumor arising from the meninges, is a benign tumor that is efficiently treated by surgical resection with invasiveness occurring in only 15-20 % of cases.

*p21* is an anti-proliferative molecule which arrests the cell cycle. *SOX2* has been proven to interact with *p21* and plays an important role in cell cycle progression and carcinogenesis in addition to its implication in EMT and tumor metastasis. *ZEB1*, an EMT marker, has been proven to be implicated in cancer aggressiveness and metastasis. Therefore, a comparison of the expression of these molecular targets implicated in EMT can help elucidate the underlying mechanisms that contribute to glioma aggressiveness in comparison with the benign tumors of meningioma.

A comparison of the mRNA expression of the molecular targets *ZEB1*, *p21* and *SOX2*, has been carried out in this study using real-time PCR technique in 31 samples of glioma and 44 samples of meningioma. The clinicopathological features of these patients have also been studied.

Based on our Real-Time PCR results, *SOX2* and *ZEB1* expression were significantly lower in glioma in comparison with meningioma. The findings on *ZEB1* and *SOX2* were statistically significant and meaningful. The P value for the difference of *SOX2* and *ZEB1* expression were 0.016 and 0.017 respectively. *p21* expression level was higher in glioma compared to meningioma but the difference was not considered statistically significant.

The difference in the expression of *ZEB1* and *SOX2* between glioma and meningioma samples suggests the fact that there are potential roles for them as targets implicated in aggressive behavior, but the extent up to which these molecules can play their roles needs further elucidation.

\*Shiva Bayat, Mohammad Taghi Raouf and Negar Mousavi are the co-first authors of this article.

\*\*Mina Tabrizi and Alireza Khoshnevisan are the co-corresponding authors of this article.

## PgmNr 774: Polymorphisms of *TAP1* and *TAP2* genes in Taiwanese women with cervical cancer.

### Authors:

T. Chang <sup>1</sup>; Y. Yang <sup>2,4</sup>; Y. Lee <sup>1,3,5</sup>; T. Chen <sup>2</sup>; W. Lin <sup>1</sup>; C. Lin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Research, MacKay Memorial Hospital, New Taipei, Taiwan; 2) Gynecology and Obstetrics Dept, MacKay Memorial Hosp, Taipei, Taiwan; 3) Pediatrics Dept, MacKay Memorial Hospital, Taipei, Taiwan; 4) Gynecology and Obstetrics Dept, Taipei Medical University, Taipei, Taiwan; 5) Pediatrics Dept, Taipei Medical University, Taipei, Taiwan

---

Human papillomavirus (HPV) infection is pivotal in cervical cancer development. However, only a small portion of HPV infected women progress to cervical cancer, suggesting the importance of host genetic makeup. Transporter associated with antigen processing (TAP) plays a central role in MHC I antigen presentation. Defects of these genes have been reported to be common in different types of cancer. This study aims to investigate whether variants of the *TAP1* and *TAP2* genes are associated with cervical cancer in Taiwanese women. The *TAP1* *rs1057141* A/G and *rs1135216* A/G, and *TAP2* *rs2228396* G/A, *rs241447* G/A, *rs4148876* G/A and *rs1800454* C/T polymorphisms were genotyped in 320 cervical squamous cell carcinoma (CSCC) patients and 320 age/sex matched healthy controls. The presence and genotypes of HPV in CSCC patients were also determined. We found no significant association between any polymorphisms or haplotypes examined and CSCC risk. Stratified by the positivity of HPV-16 infection also did not find marked association. Our findings provide no support for the hypothesis that specific polymorphisms of *TAP1* and *TAP2* genes are associated with CSCC risk in Taiwanese women.



## PgmNr 775: Discovering germline metastasis risk variants in colorectal carcinoma via case-case GWAS.

### Authors:

P.Y. Cheah<sup>1,2,3</sup>; L.F. Thean<sup>1</sup>; I. Tan<sup>3,4</sup>; M. Lo<sup>1</sup>; M. Wong<sup>1</sup>; E. Wong<sup>4</sup>; C. Chua<sup>4</sup>; C.L. Tang<sup>1</sup>; EKW. Tan<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Singapore General Hospital, Singapore, Singapore; 2) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 3) Duke-NUS Medical School, National University of Singapore, Singapore.; 4) National Cancer Center Singapore, Singapore

---

Colorectal carcinoma (CRC) is a leading cancer by incidence and mortality worldwide. Majority of death is attributable to distal organ metastasis. Hitherto, germline contribution to metastasis of sporadic CRC is unclear. We aimed to determine the contribution of germline risk variants to CRC metastasis via case-case genome-wide association study (GWAS) in about 3,000 Chinese sporadic CRC patients. Genome-wide genotyping was performed with Affymetrix SNP6 microarray. Metastasis-positive status is confirmed based on distal organ involvement attributable to primary CRC, either from histopathological report or computed tomography/positron emission tomography (CT/PET) scan regardless of the tumor staging at the time of operation. Metastasis-negative status is confirmed with at least 5 years of follow-up without distal organ involvement. The finalized dataset after quality assurance filtering (overall genotypic call rate of >0.9, minor allelic frequency of > 1%, and principal component analysis (PCA)) comprises approximately equal metastasis-positive (n=1282) and metastasis-negative (n=1395) cases. PCA plot with the Singapore Genome Variant Project (SGVP) and the HapMap subjects comprising various ethnicities indicates that there is no population substructure. The Q-Q plot indicates good genomic control ( $\lambda=1.0087$ ) without the need for any PC correction. Single nucleotide polymorphism (SNP) association testing using the additive model indicates several regions of interest (ROI) with the top SNP at  $-\log_{10}p=6.1$  and odds ratio of  $\sim 1.4$  at chromosome 20p12. Notably, sub-group analysis stratifying by tumor site, staging, distal organ metastasis and gender reveal higher  $-\log_{10}p$ -values (with SNPs in three ROIs exceeding genome-wide significance of 7.3) and effect sizes despite smaller sample sizes suggesting that the variants could be specific to the particular subgroup which is more homogeneous. Two ROIs for peritoneal- and bone-specific metastasis have rarer SNPs with high effect-sizes indicating probable population-specific risk variants. Top ranking SNPs from selected ROIs are currently being validated in an independent replication panel of another  $\sim 1,000$  cases.

## PgmNr 776: Variant characterization and effect size estimation of established colorectal cancer susceptibility genes from sequence-based case-control analysis.

### Authors:

J. Chen; Y. Yu; R. Bohlender; C. Huff

[View Session](#) [Add to Schedule](#)

**Affiliation:** Epidemiology, U.T. MD Anderson Cancer Ctr, Houston, Texas.

---

Genes responsible for Hereditary Non-polyposis Colorectal Cancer (HNPCC) have been extensively evaluated in colorectal cancer pedigrees. However, few studies have conducted detailed assessment of these genes in large case-control studies with complete gene sequencing. We conducted a high-throughput sequencing case-control study in 1810 colorectal cases and 5659 controls to further characterize the risks of known pathogenic variants and variants of uncertain significance (VUSs) in these genes. The two genes responsible for the largest proportion of HNPCC cases, *MLH1* and *MSH2*, both exhibited very large effect sizes, with ORs of 32.0 (95% CI:3.9-1462) for *MLH1* and 29.5 (CI:6.4-272.3) for *MSH2*. In contrast, pathogenic and truncating variants exhibited a more modest OR of 5.3 (CI:0.4-73.3) in *PMS2*, with no evidence of association in *MSH6* (OR = 1.2, CI:0.4-3.3). The estimated proportion of familial relative risk (pFRR) attributable to these variants was 1.8% for *MLH1*, 3.3% for *MSH2*, and 0.3% for *PMS2*. Although the associations were not statistically significant, we observed ORs greater than 1.0 for VUSs in the domain regions of each of these three genes, suggesting that a substantial proportion of pathogenic missense variants have yet to be identified. We also report effect size estimates for known pathogenic variants and VUSs in other established colorectal cancer susceptibility genes, including *APC* and *POLD1*. In total, VUSs in our study contributed a pFRR of 0.7%, for a total of 6.1% for all variant categories across all genes. Our results suggest that a substantial proportion of pFRR may be explained by VUSs in well-established high-risk genes while highlighting the need for larger sequence-based association studies to more accurately classify missense variants in established susceptibility genes and to evaluate the contribution of rare variants to cancer heritability.

## PgmNr 777: PRSweb: An interactive visual catalog for polygenic risk scores.

### Authors:

L.G. Fritsche<sup>1,2</sup>; P. VandeHaar<sup>1,2</sup>; S. Patil<sup>1,2</sup>; R.B. Peng<sup>1</sup>; M. Salvatore<sup>1</sup>; L.J. Beesley<sup>1</sup>; B. Mukherjee<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, MI; 2) Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI

---

**Introduction:** Harnessing the rich data resource of the cancer-enriched cohort of the Michigan Genomics Initiative (MGI, <https://precisionhealth.umich.edu/michigangenomics/>), a longitudinal biorepository effort at Michigan Medicine, we pursue an analytic discovery framework to integrate genetic risk information and electronic health records (EHR) to uncover novel association and/or relevant cross-phenotype associations.

**Methods:** Our current framework condenses summary statistics of published genome-wide association studies (GWAS) on cancer traits into polygenic risk scores (PRS) using linkage disequilibrium pruning and p-value thresholding (P&T) or genome-wide posterior mean effect size weighting (LDpred). For each PRS, we determine its predictive performance before performing association scans across the PheCode-based phenomes of MGI and the UK Biobank study.

**Results:** So far, we have mined various sources of available cancer GWAS data, e.g. the Neale Lab UK Biobank GWAS results or the NHGRI-EBI GWAS Catalog and constructed 36 PRS for 19 common cancer traits. Phenome-wide association studies (PheWAS) of these PRS across the MGI phenome of over 38,000 unrelated, European participants included over 1,600 phenotypes and, for some PRS, indicated strong association with their primary cancer trait as well as additional secondary traits. For GWAS-catalog-based PRS, we also validated our approach using the large, population-based UK Biobank study.

**Conclusions:** Since there is an urgent need to better communicate results from such big data exploration with a broad and multi-disciplinary audience, e.g. to inspire hypothesis development, we are developing PRSweb, an interactive webpage for intuitive result browsing. A current working draft already shows downloadable PRS constructions and PRS PheWAS results for all constructed PRS and is available at <http://csg.sph.umich.edu/larsf/PRSweb>. Future implementations will incorporate PRS for a broader set of complex traits beyond cancer and also include tools and tutorials for PRS research.

## **PgmNr 778: Discovery of rare and novel variants associated with breast cancer in Trinidadian families using whole exome sequencing.**

**Authors:**

R. Haraksingh

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dept. of Life Sciences, The University of the West Indies, St. Augustine, Trinidad and Tobago

---

Breast Cancer is the most prevalent cancer that affects women both worldwide and in Trinidad & Tobago. According to the World Health Organization it accounts for 22.1% of all cancer mortality cases as of 2014. This study aims to determine whether there are novel genetic contributors to the development of Breast Cancer in Trinidadian individuals. Approximately 30 genes contributing to Breast Cancer susceptibility have been discovered largely through studies done on Caucasian populations. However, the extent to which these genetic contributors are relevant in Caribbean populations is unknown. 90 Breast Cancer patients in Trinidad have undergone genetic screening for common cancer-causing mutations using the most extensive Breast Cancer screening panel to date. This panel covers 30 genes and 78% of these patients were found to not carry any of these known mutations. As such, we believe that there may be novel genetic variants present in this population that are the potential cause of some hereditary Breast Cancer cases. We performed Whole Exome Sequencing of two members of such a family with evidence of hereditary Breast Cancer. Subsequent bioinformatics analysis revealed potential novel and understudied genetic variants associated with the Breast Cancer condition in this family. This study demonstrates the importance of inclusion of understudied populations in medical research and reveals novel molecular points of entry for developing new Breast Cancer diagnostics and therapeutics for this population.

## PgmNr 779: Early age of onset in lymphoid cancer families after controlling for ascertainment bias.

### Authors:

S. Jones <sup>1,2</sup>; J. Voong <sup>3</sup>; J. Graham <sup>3</sup>; A. Brooks-Wilson <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada; 2) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 3) Department of Statistics and Actuarial Science, Simon Fraser University, Vancouver, BC, Canada; 4) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Vancouver, BC, Canada

---

**Introduction:** Lymphoid cancers are the fourth most common cancer type in the United States. Familial aggregation and early onset of lymphoid cancers are suggestive of shared genetic risk factors. Genetic anticipation, the earlier onset of disease in successive generations, has been reported among lymphoid cancer families; however, the underlying molecular mechanism is unknown. Ascertainment bias and birth cohort effects can be misinterpreted as genetic anticipation.

**Methods:** We reviewed 200 multigenerational pedigrees with 2 or more lymphoid cancer cases including non-Hodgkin lymphoma (NHL), Hodgkin lymphoma (HL), chronic lymphocytic leukemia (CLL) and multiple myeloma (MM), for deviation from the expected population age of onset. SEER population data was used to control for reported sex, ethnicity, lymphoid cancer subtype and age of onset, allowing for the uniform comparison of individuals and cofactors that affect risk of lymphoid cancers. Ascertainment bias was controlled for by excluding probands and adjusting for birth cohort effect.

**Results:** The mean age of onset varied by type of lymphoid cancer ( $p < 0.0001$ ), and remained significant after controlling for sex, ethnicity, and subtype ( $p = 0.0023$ ). Median familial NHL ( $p < 0.0001$ ), HL ( $p = 0.0019$ ), CLL ( $p < 0.0001$ ) and MM ( $p = 0.0096$ ) ages of onset are substantially earlier than comparable population data. Anticipation for NHL ( $p < 0.0001$ ), HL ( $p = 0.0008$ ) and CLL ( $p = 0.0255$ ) (but not MM) was observed, and remained significant for NHL ( $p < 0.0001$ ) and HL ( $p = 0.0410$ ) after excluding probands. Anticipation for NHL ( $p = 0.0060$ ) and lymphoid cancers collectively ( $p < 0.0001$ ) was observed after excluding young birth cohorts.

**Conclusions:** The median age of onset in families was substantially younger than population cases when controlling for known cofactors that affect risk. Evidence for anticipation remained even after controlling for cofactors and excluding probands and young birth cohorts. These observations support the application of genomic methods to identify genes and genetic variants that underlie familial lymphoid cancers. Awareness of patterns such as anticipation in lymphoid cancer families may help predict risk in family members.

## PgmNr 780: Germline DNA repair gene mutations in multiple myeloma.

### Authors:

V. Joseph <sup>1,2</sup>; S.M. Lipkin <sup>2</sup>; K. Offit <sup>1,2</sup>; O. Landgren <sup>1,2</sup>; N. Camp <sup>3</sup>; C. Vachon <sup>4</sup>; M. Hildebrandt <sup>5</sup>; E.E. Brown <sup>6</sup>; J. McKay <sup>7</sup>; J. Garber <sup>8</sup>; R.J. Klein <sup>9</sup>; V. Ravichandran <sup>1</sup>; A. Norman <sup>4</sup>; M. Adams <sup>10</sup>; K. Doheny <sup>10</sup>; B. Thomas <sup>11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York.; 2) Department of Medicine, Weill Cornell Medical College, New York, NY; 3) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 4) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 5) Department of Epidemiology, MD Anderson Cancer Center, Houston, TX; 6) Department of Pathology, University of Alabama at Birmingham, Birmingham, AL; 7) Genetic Cancer Susceptibility Group, International Agency for Research on Cancer, Lyon, FR; 8) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA; 9) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY; 10) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 11) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

---

Multiple Myeloma (MM) accounts for almost 14% of all hematologic cancers and is essentially incurable. A clear racial disparity in incidence and mortality is observed in MM. The precursor state Monoclonal Gammopathy of Uncertain Significance (MGUS) shows a progression risk of 1% each year. Several low penetrant common variations near transcriptional regulators and tumor suppressors associated with risk of sporadic MM have been identified from genome-wide association studies. A few monogenic high penetrant pathogenic germline mutations were also identified in European American familial and sporadic cases. We have created the Multiple Myeloma Genetics Consortium (MMAGIC) to address genetic MM predisposition using germline whole exome sequence (WES) data. Using the approved resources of the Center for Inherited Disease Research sequencing facility, we conducted a pilot study of 155 WES from MSKCC and WCMC. We analyzed the QC/QA metrics derived from the Illumina NovaSeq 6000 100 bp paired-end sequenced samples on 155 individuals, with capture using the exome bait sets from Twist Biosciences. We compared these with HiSeq 4000 sequencing on a subset of individuals. The samples included in the pilot were a few families and otherwise early onset sporadic cases of MM. We then ran our automated variant curator PathoMAN on the resulting variant data. We observed one rare recurrent variant in *RAD51B* and a few other rare variants in other DNA repair pathway genes such as *BLM* and *RECQL*. Our preliminary analysis shows the presence of rare pathogenic DNA repair gene mutations in MM.

## PgmNr 781: Putative non-coding regulatory driver discovery using allele specific expression in unmatched tumor-healthy samples.

### Authors:

N.M.R Lykoskoufis; H. Ongen; E.T. Dermitzakis

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetic Medicine and Development, University of Geneva, Geneva, Geneva, Switzerland

---

Cancer development is not fully understood, and evidence suggest that non-coding regulatory drivers play a key role in tumorigenesis. Protein-coding drivers being the sole forces of tumorigenesis increasingly appears as biologically implausible at least for some cancers. However, given the complexity of non-coding DNA, regulatory driver discovery requires new methods.

We had developed a method that uses allele specific expression (ASE) as a proxy to discovering genes with allelic dysregulation impacted by non-coding somatic mutations, that requires matched healthy-tumor samples (Ongen et al., 2014). However, matched cohorts are not easy to obtain. Thus, we developed a novel method following the same principles as previously, that can be used with reference unmatched samples, from RNA-seq datasets, from the same tissue of origin.

Regulatory somatic mutations are expected to be in heterozygosity, so they will affect ASE. Hence, under selection, we expect genes to show increased ASE in tumor compared to healthy samples and a directional change of expression levels. Thus, to detect regulatory drivers, we devised our method based on three key filtering steps.

We use colorectal cancer data from the SYSCOL project comprising of matched healthy-tumor samples and GTEx data from colon sigmoid and colon transverse. We perform three unmatched analyses that we compare with previous results from a SYSCOL matched design. At 5% FDR, we discover 360 genes when treating SYSCOL samples as unmatched, 360 genes combining tumor with colon sigmoid and 409 genes combining tumor with colon transverse. Of these, 287 genes overlap all three analyses. We compare them to the 105 genes discovered in the SYSCOL matched design and observe an overlap of 95 genes, indicating that our method efficiently detects regulatory drivers. Some identified genes are already implicated in cancer like *KRAS* and we observe an enrichment for colorectal cancer involved pathways like the APC and MAPK pathways.

We had showed that the impact of copy number alterations (CNA) is minimal. However, we are improving our analytical framework to account for it. We extend our analysis to prostate and other cancers and plan on using WGS to discover the actual non-coding somatic mutations that may be causing the observed ASE difference.

All in all, this method provides a great opportunity to learn more about regulatory drivers that can be broaden to multiple cancers and allow us to obtain the full picture of tumorigenesis.



## PgmNr 782: An integrative genetic epidemiologic approach identifies new low- and medium-penetrance susceptibility genes for breast cancer.

### Authors:

R. Moslehi <sup>1</sup>; H. Tsao <sup>1</sup>; A. Dzutsev <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Public Health and Cancer Research Center, University at Albany, State University of New York, Albany, NY; 2) Cancer Inflammation Program, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD

**Introduction:** A large fraction of common (peri- and post-menopausal) breast cancer cases in the general population may be due to segregation of genes with low and moderate risk [relative risk  $\leq 4$ ] for breast cancer; the majority of these genes remain unidentified due to methodological challenges presented by small effect sizes. We used a hypothesis-driven integrative genetic epidemiologic approach aimed at identifying low-penetrance breast cancer susceptibility genes.

**Methods:** We tested the association of common breast cancer with genes in a candidate biologic pathway by analyzing multiomics data including genome-wide association studies (GWAS), gene expression microarray, transcriptome, and proteome datasets. We restricted analysis to Caucasian cases with invasive ductal breast carcinoma diagnosed at  $\geq 50$  years of age and healthy controls. Raw data from four gene expression microarray datasets (98 cases, 163 controls) were used for pathway and meta-analysis. Raw data from four GWAS datasets (2961 cases, 4335 controls) were analyzed using logistic regression to calculate odds ratios (OR) and 95% confidence intervals (CI) adjusting for covariates. We also analyzed The Cancer Genome Atlas (TCGA) breast tumor data (522 cases) by comparing frequency of mutations in our candidate genes with all other genes and with known breast cancer susceptibility genes.

**Results:** Meta-analysis of gene expression microarray datasets identified significant upregulation of several of our candidate genes including *ERCC6* ( $p=7.95 \times 10^{-6}$ ) and *ERCC8* ( $p=4.67 \times 10^{-6}$ ). Candidate-gene analysis of GWAS data identified increased risk for variations in those genes in all datasets; for example, increase in risk associated with one *ERCC6* variation ranged from 35% (OR=1.35, 95%CI:1.01-1.80,  $p=0.04$ ) to 3-fold (OR=2.91, 95%CI:1.05-8.06,  $p=0.04$ ) in different datasets. Statistically-significant increased risk was also found with specific haplotypes in each gene and with joint *ERCC6-ERCC8* diplotypes (up to 5-fold,  $p \leq 0.034$ ). Analysis of TCGA data supported our findings; for example, *ERCC6* mutation frequency in breast tumors (1.80%) was comparable to known high- and medium-penetrance breast cancer susceptibility genes such as *BRCA1* (2.90%), *BRCA2* (2.90%), *BLM* (1.90%), and *CHEK2* (1.00%).

**Conclusions:** Using a hypothesis-driven integrative genetic epidemiologic approach to analysis of multiomics data, we have identified several previously unreported low- and medium-penetrance breast cancer susceptibility genes.



## PgmNr 783: Evaluating association of colon cancer SNPs with colonoscopy results and precancerous polyps.

### Authors:

Qin <sup>1,2</sup>; E. Hauser <sup>1,2</sup>; B. Sullivan <sup>1,2</sup>; A. Madison <sup>1</sup>; K. Sims <sup>1</sup>; T. Redding <sup>1</sup>; M. O'Leary <sup>1</sup>; A. Stone <sup>3</sup>; W. Carter <sup>3</sup>; D. Weiss <sup>4</sup>; D. Lieberman <sup>5,6</sup>; D. Provenzale <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) VA Cooperative Studies Program Epidemiology Center, Durham VA Health Care System, Durham, North Carolina; 2) Duke University, Durham, NC; 3) Cooperative Studies Program Pharmacogenomics Analysis Lab, Central Arkansas Veterans Health System, Little Rock, AR; 4) Perry Point VA Medical Center, Perry Point, MD; 5) VA Portland Health Care System, Portland, OR; 6) Oregon Health & Science University, Portland, OR

---

Recent meta-GWAS of 125,478 colorectal cancer (CRC) cases and controls has increased the number of replicated SNPs associated with CRC to 212, resulting in a predictive CRC genetic risk score (GRS) comprised of 120 SNPs (Hyghe et al. Nature Genetics, 2018). The goal of our study was to evaluate associations between these CRC SNPs and longitudinal outcomes in a prospective CRC screening cohort followed for at least 10 years using the VA medical record. The VA Cooperative Studies Program (CSP) 380 study was initiated between 1994-1997 and enrolled 3,121 healthy veterans at average risk of CRC at baseline. This cohort includes an associated biorepository with DNA, serum, polyp and normal tissue. Genome-wide genotyping was performed on DNA from 612 individuals representing 1,220 colonoscopies over a 10-year period. The primary outcome of this study was the finding of advanced neoplasia (AN) at any colonoscopy, defined as any tubular adenoma  $\geq 10$  mm, villous histology, high-grade dysplasia or CRC. Each colonoscopy provided a phenotype for the presence or absence of AN. We identified 238,254 SNPs genotyped in the CSP 380 dataset representing SNPs +/- 2 mb from each CRC SNP in the literature. We performed a genetic association analysis in the subset of 500 European-Americans for cases versus those without AN using logistic regression with an additive genetic model, adjusting for age and two genetic principal components (PCs) as implemented in PLINK version 1.09. We also performed an analysis of AN at each colonoscopy using a linear mixed model as implemented in R glmmPQL. Fourteen SNPs in eight regions had p-values  $< 10^{-4}$ . Of particular interest are associations of AN with four CRC regions: 1) chromosome 6p21.32 with 4 CRC SNPs within 2 mb of rs12207382 (OR=1.76); 2) chromosome 17q25.3 with 2 CRC SNPs within 2 mb of rs9914652 (OR=1.77), rs62073391 (OR=2.20) and rs56975743 (OR=1.91); 3) chromosome 19p13.11 with 1 CRC SNP within 2 mb of rs2733762 (OR=1.78); and 4) chromosome 20q13.13 with 3 CRC SNPs within 2 mb of rs230013 (OR=1.78). These regions may represent genetic variants associated with early preneoplastic stages in the progression to CRC and may be particularly useful in early risk stratification at the time of colonoscopy. Next steps include evaluation of the utility of the CRC GRS in prediction of participants with AN and development of a suitably large mega-cohort of individuals with longitudinal colonoscopies.

## PgmNr 784: Germline genotyping association with tumor immune RNAseq signatures in NSCLC.

**Authors:**

D. Quarless; Z. Dezso; A.J. Grundstad; D. Masica; J. Waring; J.W. Davis; M. Ceccarelli; B. Riley-Gillis

[View Session](#) [Add to Schedule](#)

**Affiliation:** Abbvie INC, North Chicago, Illinois.

---

Improvements in the clinical efficacy of immunoncology have invigorated immunotherapeutic treatments of multiple forms of cancer. Tumor immune signatures as biomarkers have been utilized to investigate both the tumor immune landscape and microenvironment. Furthermore, the composition of the tumor immune infiltrate has increasingly demonstrated predictive capabilities for immunotherapeutic response in cancer patients.

Germline polymorphisms are known to influence drug efficacy and cancer susceptibility. Germline variation and somatic tumor DNA alterations interactions from The Cancer Genome Atlas (TCGA) have demonstrated the influence of inherited germline polymorphisms on the landscape of somatic mutation profiles in tissue and gene-specific manner. Additional TCGA studies reveal the germline SNP influence on the cancer-immune phenotype and therapeutic response.

Here, we explore associations of germline SNPs with tumor immune RNAseq signatures by leveraging pre-treatment tumor transcriptomic data and matched germline genotype data from clinical samples with non-small cell lung cancer (NSCLC). Twenty-three immune enrichment scores were calculated from processed RNAseq data for the following immune sets: tumor inflammation signature (TIS), the immunologic constant of rejection (ICR), exhaustion signature, and the Bindea immune cell signatures. Immune scores, genotypes, and study covariates for the clinical studies were analyzed in a linear association framework (n=385 total patients).

Genome-wide analysis revealed 2 statistically significant association regions ( $p < 5e-08$ ) to the central memory T cell (TCM) and natural killer cells with a lower surface density of CD56 antigen (NK.CD56dim) immune scores. To restrict multiple testing burden, germline SNPs were filtered to +/-50kb of 113 immunoncology-related genes and 2 additional significant associations with the B-cell and T effector memory cell (TEM) signatures were identified ( $p < 8e-07$ ). Interestingly, eQTL effects do not explain these associations as the implicated genes are not part of the signatures. Ongoing analyses of associated regions aim to uncover potential genetic mechanisms between SNPs, genes, and signatures.

Our results support a germline genetic influence on the tumor immune environment. These findings implicate germline SNPs as a potentially important factor for the tumor immune landscape and could aid efforts to stratify patients receiving immunotherapies.

## PgmNr 785: Functional characterization of non-coding regulatory drivers in chronic lymphocytic leukemia.

### Authors:

A. Real <sup>1</sup>; H. Ongen <sup>1</sup>; N.M.R. Lykoskoufis <sup>1</sup>; C. Borel <sup>1</sup>; G. Puga Yung <sup>2</sup>; J.D. Seebach <sup>2</sup>; E.T. Dermitzakis <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Department of Medicine, University of Geneva and University Hospitals Geneva (HUG), Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, Switzerland; 4) SIB, Swiss Institute of Bioinformatics, Geneva, Switzerland

---

The role of coding mutations in cancer development has been widely studied. However, the identification of the non-coding regulatory drivers and the genes mediating their effects, remains poorly understood. In a previous study (BioRxiv 174219), we identified 19 local modules of coordinated non-coding regulatory elements of the genome (cis-regulatory domains or CRDs, FDR 5%) that accumulated an excess of somatic mutations in 150 patients with chronic lymphocytic leukemia (CLL), a cancer affecting B-cells. We hypothesize that the CRDs with an excess of somatic mutations (after controlling for mutational confounders) can have a driver role in CLL development and we aim to perform functional characterization of their effect on the 45 genes inferred to be regulated by these putative driver CRDs.

In this work, we present the development of an in vitro cancer-like model in which transformed lymphoblastoid cell lines (LCLs) are evaluated for their ability to reproduce cancer-like phenotypes such as increase in cell migration and proliferation and decrease of apoptosis. Among the 45 genes identified, 35 genes showed differential expression between CLL and LCLs. For 21 genes with increased expression due to hypermutated CRDs, we setup a multiplexing approach to knock-in several genes at once using a single lentiviral gene expression vector and evaluate the phenotypic changes. For 14 genes with observed decreased expression, we use a multiplex CRISPR-Cas9 approach based on a single lentiviral vector coding for the Cas9 protein together with multiple gRNA guides. The presence of a GFP tag will allow FACS-based screening for the selection of cells with high GFP expression and thus high multiplicity of infection.

Limitations of this approach are the difficulties to transfect LCLs, partially solved by the decision to use a lentiviral vector increasing the vector delivery, and the complexity to mimic a disease using an in vitro model.

The identification of changes in transformed B-cells towards a cancer-like phenotype will provide a functional characterization of non-coding mutations in CLL and it will potentially identify new genes driving tumorigenesis. This study could also provide a strong functional and experimental framework to explore and dissect the contribution of the non-coding cancer genome not only in CLL but also in other cancers, helping to understand the biological complexity of these effects. Experiments are in process and first results will be presented.

## PgmNr 786: Cancer spectrum in male *BRCA1/BRCA2* mutation carriers: Results from the Consortium of Investigators of Modifiers of *BRCA1/BRCA2* (CIMBA).

### Authors:

V. Silvestri <sup>1</sup>; J. Kirk <sup>2</sup>; S.L. Neuhausen <sup>3</sup>; T.R. Rebbeck <sup>4,5</sup>; M. Tischkowitz <sup>6</sup>; G. Chenevix-Trench <sup>7</sup>; A.C. Antoniou <sup>8</sup>; E. Friedman <sup>9,10</sup>; L. Ottini <sup>1</sup>; CIMBA

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular Medicine, Sapienza University of Rome, Rome, Rome, Italy; 2) Familial Cancer Service, Weatmead Hospital, Wentworthville, New South Wales, Australia; 3) Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA, USA; 4) Harvard T.H. Chan School of Public Health, Boston, MA, USA; 5) Dana-Farber Cancer Institute, Boston, MA, USA; 6) Department of Medical Genetics, University of Cambridge, Addenbrookes Hospital, Cambridge, UK; 7) Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 8) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 9) The Susanne Levy Gertner Oncogenetics Unit, Chaim Sheba Medical Center, Ramat Gan, Israel; 10) Sackler Faculty of Medicine, Tel Aviv University, Ramat Aviv, Israel

Men with *BRCA1/2* mutations represent a unique group in terms of clinical management. The paucity of data on cancers arising in male *BRCA1/2* mutation carriers has limited the development of clinical guidelines for men with *BRCA1/2* mutations. We characterized the cancer spectrum observed in male *BRCA1/2* mutation carriers collected within the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA) and compared *BRCA1* with *BRCA2* mutation carriers in terms of number, site and age of cancer diagnoses.

This study encompassed data from 6902 men, 3651 (52.9%) with *BRCA1* and 3251 (47.1%) with *BRCA2* mutations. A total of 1376 men (19.9%) had at least one cancer diagnosis, the majority (922/1376, 67%) carrying a *BRCA2* mutation. Compared with men carrying *BRCA1* mutations, men carrying *BRCA2* mutations are more likely to be affected with cancer [Odds Ratio (OR) = 3.20, 95% Confidence Interval (CI) 2.79-3.67,  $p < 0.0001$ ] and to develop multiple tumors (OR for trend = 2.77, 95%CI 2.46-3.11,  $p < 0.0001$ ). A total of 1634 cancers were reported in the 1376 affected individuals. Breast and prostate cancers were significantly more frequent (OR for breast cancer = 5.49, 95%CI 4.03-7.48,  $p < 0.0001$ ; OR for prostate cancer = 1.42, 95%CI 1.11-1.82,  $p = 0.006$ ), whereas all cancers other than breast and prostate taken together were significantly less frequent (OR = 0.22, 95%CI 0.18-0.28,  $p < 0.0001$ ), in *BRCA2* compared with *BRCA1* mutation carriers. In analyses restricted to cancers other than breast and prostate, colorectal cancer was less frequent (OR = 0.47, 95%CI 0.29-0.77,  $p = 0.003$ ), whereas pancreatic cancer was more frequent (OR = 3.04, 95%CI 1.57-5.88,  $p = 0.001$ ) in *BRCA2* compared with *BRCA1* mutation carriers.

Mean ages at diagnosis of breast and prostate cancers were similar for *BRCA1* and *BRCA2* mutation carriers, while for other malignancies taken together, age at diagnosis was older in *BRCA1* (59 years) compared with *BRCA2* mutation carriers (55.6 years,  $t$ -test  $p$ -value=0.0036).

Our results highlight significant differences in male *BRCA2* mutation carriers compared with *BRCA1* mutation carriers, that may actually help refine existing recommendations in terms of specifying

distinct surveillance guidelines for men with either *BRCA1* or *BRCA2* mutations.

As under the present study design it was not possible to assess the hypothesis that *BRCA1/2* mutations are associated with elevated cancer risks for male carriers compared to non-carriers, future studies are planned to address these issues.

## PgmNr 787: Identification and functional characterization of germline variants predisposing patients to congenital anomalies and childhood cancer.

### Authors:

S. Sisoudiya<sup>1,2</sup>; H. Li<sup>3</sup>; J. Schraw<sup>4</sup>; A. Sabo<sup>3</sup>; P.J. Lupo<sup>2</sup>; S.E. Plon<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College Medicine, Houston, TX; 2) Section of Hematology-Oncology, Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Department of Medicine, Baylor College of Medicine, Houston, TX

**Introduction:** Little is known about the role of non-chromosomal birth defects on cancer risk and the underlying genetic mechanism. In the **Genetic Overlap Between Anomalies and Cancer in Kids** (GOBACK) Study, we performed association analysis between birth defects (BD) and childhood cancers (CC) by leveraging birth defect and cancer registries from Texas, Michigan, Arkansas and North Carolina. We evaluated more than 10 million births which enabled us to identify specific BD-CC patterns (Lupo et al., JAMA Oncology, *in press*). Based on these associations, proband-parent trios from the GOBACK Study are being enrolled for a family-based whole-genome sequencing (WGS) study to evaluate the underlying genetic mechanisms.

**Method:** WGS (40X) was performed on DNA extracted from salivary samples of the enrolled families. Single nucleotide variants and short insertions/deletions were called using Platypus. Structural variants (SVs) were called using multiple callers, followed by merging and filtering via SURVIVOR. A consensus call was made by at least 2 callers.

**Results:** Our epidemiology study identified a novel association between choanal atresia and acute leukemia (HR=9.2, 95% CI: 3.8-22.1). SV analysis of the WGS data identified a ~5kb *de novo* heterozygous deletion in *USP9X* in a female proband with multiple anomalies including choanal atresia and acute leukemia (B-ALL). Furthermore, the patient's clinical phenotype was indicative of a CHARGE-like syndrome with negative *CHD7* analysis. *De novo* LOF mutations in *USP9X* were recently implicated in a female-specific CHARGE-like syndrome with overlapping features to our patient including two patients with childhood cancer (one each with B-ALL and osteosarcoma). *USP9X* is also recurrently somatically mutated in sporadic ALL. It encodes a highly conserved deubiquitinating enzyme. We have successfully used CRISPR/Cas9 to create the heterozygous deletion and a full knock-out of *USP9X* in B-cell derived lymphoblastoid cells. Clones of these edited cells will be assayed for alterations in proliferation and transformation characteristics as well as changes in stability of *USP9X* targets to understand the potential role of *USP9X* as a novel ALL susceptibility gene.

**Conclusion:** Our results suggest that *USP9X* represents a novel ALL susceptibility gene and also plays an important role in sporadic ALL. The experiments underway should provide insights into the molecular mechanisms of *USP9X*-mediated leukemogenesis.

## **PgmNr 788: Mutational signatures and mutagenic impacts associated with betel quid chewing in oral squamous cell carcinoma.**

### **Authors:**

S. Su <sup>1</sup>; L. Chang <sup>2</sup>; S. Yang <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Chang Gung Memorial Hospital, Keelung, Taiwan; 2) Florida Atlantic University, Florida, US; 3) Chung Shan Medical University, Taichung, Taiwan

---

Betel quid (BQ) chewing is a prevailing risk for oral squamous cell carcinoma (OSCC) in Central and Southeast Asia. Yet, the detailed mechanisms by which BQ chewing damages the genome and creates somatic alterations that ultimately cause OSCC are still not fully understood. Through performing exome sequencing of tumor-normal pairs from 196 OSCC patients, including 95 habitual BQ chewers and 101 non-BQ users, we conducted a quantitative survey of mutational signatures and genomic aberrations and explored their association with BQ chewing. Elevated mutation rate was seen in cancers of the tongue but not in overall OSCC with respect to the use of BQ. Additionally, we identified a specific mutational signature that is enriched in tumors from BQ users. Moreover, increased numbers of small insertions and deletions (INDELs) and breakpoints derived from structural variations (SV) were observed, whereas decreased extent of loss of heterozygosity was detected in the coding regions of BQ-related OSCC genomes. However, neither the number of base substitutions and microsatellite instability events nor the extent of copy-number alterations differed between BQ-related and -unrelated OSCC. In concordance with the proposition that BQ chewing increases OSCC risk as a mutagen, our results unveil a BQ-associated mutational pattern and indicate mutagenic impacts of BQ chewing on preferentially eliciting small INDELs and SV-related breakpoints in OSCC genomes.



## PgmNr 789: Rethinking models to study the metastatic and metastatic suppressive molecular pathways in skin and brain.

### Authors:

F. Vand-Rajabpour<sup>1</sup>; P. Noormohammadpour<sup>2,3</sup>; S. Panjeh-Shahi<sup>1</sup>; A. Khoshnevisan<sup>5</sup>; R. Ahmadi<sup>1</sup>; J. Rahmati<sup>3</sup>; Z. Tabrizi<sup>1</sup>; F. Tabrizi<sup>1</sup>; M. Daneshpajoo<sup>2,3</sup>; K. Kamyab-Hesari<sup>3</sup>; M. Yaseri<sup>4</sup>; S. Bayat<sup>1</sup>; M.T. Raouf<sup>5</sup>; M. Tabrizi<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics Department, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran; 2) Autoimmune Bullous Disease Research Center, Department of Dermatology, Razi Hospital, Tehran University of Medical Sciences, Razi Hospital-Vahdate Eslami Street, Tehran, Iran; 3) Tumor Clinic, Pathology Department and the Department of Plastic and Reconstructive Surgery, Razi Dermatology Hospital, Tehran University of Medical Sciences, Tehran, Iran.; 4) Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran; 5) Department of Neurosurgery, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran

Studying the molecular mechanism of the least metastatic cancers could be as significant as studying highly metastatic cancers for molecular mechanistic targeting clues. We hypothesized that spectrum of cancers, based on metastatic and invasion features, could be a good *in vivo* model and the molecular mechanism of the least metastatic cancer to elucidate key regulatory molecules in EMT and stemness on the road to development of the most metastatic cancer in spectrum of cancers. This hypothesis was examined in two spectrums of cancers, from the least metastatic to the most metastatic cancer in skin cancers and brain tumors with both organ systems originating from the embryonic ectodermal layer.

Firstly, 443 samples of the spectrum of skin cancers, from basal cell carcinoma (BCC-the least metastatic cancer), to squamous cell carcinoma (SCC-often metastatic cancer) and to melanoma (most metastatic cancer) were collected to study the expression level of ten mRNA/lncRNA as EMT and stemness key regulatory molecules in skin cancers compared to related controls (*BMI1*, *TWIST1*, *SNAI2*, *SNAI1*, *EZH2*, *HIF1a*, *SOX9*, *p16*, *FAL1* and *ANRIL*). We demonstrated downregulation of *SNAI1* ( $p=0.005$ ), *EZH2* ( $p=0.022$ ), *HIF1a* ( $p=0.008$ ), *TWIST1* ( $p=0.017$ ), *BMI1* ( $p=0.000$ ), *ANRIL* ( $p=0.004$ ) and *FAL1* ( $p=0.935$ ) and also upregulation of *p16* ( $p=0.000$ ), *SOX9* ( $p=0.0128$ ) and *SNAI2* ( $p=0.000$ ) in non-metastatic BCCs. In SCCs, upregulation of *p16* ( $p=0.001$ ), *HIF1a* ( $p=0.001$ ), and *ANRIL* ( $p=0.002$ ) were seen. In melanomas, upregulation of *p16* ( $p=0.035$ ) and *ANRIL* ( $p=0.000$ ) were shown.

Subsequently, 75 samples of the spectrum of brain tumors (glioma and meningioma) alongside with the 27 BCCs and 12 controls were collected and the mRNA expression of *p21*, *ZEB1* and *SOX2* were analyzed. The expression of *SOX2* significantly decreased in BCCs, as the least metastatic cancer ( $p<0.001$ ), compared with controls. The expression of *SOX2* and *ZEB1* significantly increased in meningiomas, as the least aggressive tumor, compared to gliomas, as the aggressive tumor ( $p=0.016$ ,  $p=0.017$ , respectively).

These findings stress the importance of further studies on spectrum of cancers, based on



metastatic/invasion features, which could be a good *in vivo* model to shed some light on the molecular mechanism of metastasis/aggressiveness of cancers. Also, our findings emphasized that BCC could be a unique *in vivo* cancer model to comprehend the intrinsic metastatic attenuating regulatory mechanisms.

## PgmNr 790: Whole-exome sequencing analyses identify risk genomic loci for nasopharyngeal carcinomas.

### Authors:

Y. Zeng <sup>1</sup>; M. Lee Kiang Chua <sup>2,3</sup>; C. Chuen Khor <sup>4</sup>; Y. Zeng <sup>5</sup>; J. Bei <sup>5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-Sen University, China, Guangzhou, Guangdong, China; 2) Division of Radiation Oncology and Division of Medical Sciences, National Cancer Centre Singapore, Singapore 169610.; 3) Duke-NUS Medical School, Singapore 169857.; 4) Human Genetics, A\*STAR, Genome Institute of Singapore, Singapore, 138672, Singapore.; 5) Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Guangdong Key Laboratory of Nasopharyngeal Carcinoma Diagnosis and Therapy, Guangzhou 510060, P. R. China.; 6) Center for Precision Medicine, Sun Yat-sen University, Guangzhou 510080, P. R. China

### Background:

Nasopharyngeal carcinoma (NPC) featured a relatively high prevalence in southern China, southeast Asia and Northern Africa, yet our current understanding of its genetic etiology is confined to common variants which were identified through genome-wide association studies. It has been estimated that the genetic variation contributes 61.3% of the risk to develop NPC (narrow sense heritability), with only 16.9% of this genetic contribution attributable to the common variants, indicating an urgent need to study genetic effects from germline rare variants for this disease.

### Methods:

Exome-wide sequencing was conducted for blood samples of 2465 NPC patients and 2304 healthy controls recruited from south China and southeast Asia. This collection also includes 354 NPC patients with a family history. A SNP-set (Sequence) kernel association test (SKAT) was applied to localize aggregate genetic effects from germline rare variants in coding genes on NPC. A Generalized gene-set analysis was used to identify the combined genetic effect from both common and rare variants at pathway level on NPC.

### Results:

The gene-based SKAT analysis identified seven genes displaying consistently elevated frequency of rare mutations in NPC patients across samples as compared with healthy controls. A subset of these genes additionally showed a higher frequency of mutation in subgroups of patients with family history. The exon-based SKAT analysis combined with single variants association tests localized major contributing rare variants for the association observed at gene level. A proportion of these contributing rare variants showed distinguished minor allele frequency in Asia population as compared with European population. The generalized gene-set analysis identified functional pathways showing increased frequency of risk alleles from both common and rare variants in NPC patients.

### Conclusions:

This study identified important germline rare mutations, coding genes and molecular pathways contributing to the genetic etiology of NPC. We provided potential biomarkers for personalized NPC

genetic risk evaluation and potential explanations of the regional differences of NPC prevalence. Future studies of functional consequences of these molecular variation are required to elucidate the causal pathways connecting these germline mutations and NPC.

## PgmNr 791: Integrating multi-omics data identified novel susceptibility genes for colorectal cancer.

### Authors:

K. Zhang; H. Wu; J. Guo; S. Yao; T.L. Yang

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shanxi, China

---

Colorectal cancer (CRC) is the third most common cancer and the fourth leading cause of cancer-related deaths worldwide. Recent studies of the genomics, transcriptomics, proteomics and metabolomics landscapes of human CRC have identified many alterations and molecular markers of the disease. However, the rapidly accumulating omics data have yet to bring novel biomarkers and drug targets to the clinic. To better understand the disease, we integrated multi-omics data of CRC to identify novel susceptibility genes. The data we used included the genome-wide association studies (GWAS) summary data of the malignant neoplasm of colorectum from UK Biobank in a total of 3,825 CRC patients and 448,439 controls and the summary-level GWAS data of blood metabolites involving 363 metabolites. The protein expression data of 90 tumor specimens and 30 normal tissues was derived from Clinical Proteomic Tumor Analysis Consortium (CPTAC), and the transcriptome data of CRC with 645 tumor specimens and 51 normal tissues was from The Cancer Genome Atlas (TCGA). We began an exploration of the causal associations between blood metabolites and CRC with two-sample Mendelian randomization (MR) method and identified that 31 blood metabolites had causal effect on CRC with a suggestive P value ( $P < 0.05$ ). Enrichment and pathway analyses showed that these metabolites were involved in 39 pathways and functional processes, such as glutathione metabolism, glycerophospholipid metabolism, urea cycle and energy metabolism, which referred 684 genes defined by Kyoto Encyclopedia of Genes and Genomes (KEGG). Moreover, the different expression analyses both in proteomic and transcriptomic landscapes indicated 128 CRC-associated genes. Therein, *ANPEP* and *GGCT*, the glutathione regulatory genes, were confirmed had an effect on patient survival and implicated novel targets for CRC. In summary, we performed integration analysis comprising genomics, transcriptomics, proteomics and metabolomics, and found that *ANPEP* and *GGCT* may be novel susceptibility genes for CRC.

## PgmNr 792: 16p11.2 microdeletion syndrome associates with neuroblastoma.

### Authors:

L.E. Egolf<sup>1,2,3</sup>; Z. Vaksman<sup>2,3,4</sup>; G. Lopez<sup>2,3,4</sup>; J.L. Rokita<sup>2,3,4</sup>; A. Modi<sup>2,3,5</sup>; P.V. Basta<sup>6,7</sup>; H. Hakonarson<sup>8,9</sup>; A.F. Olshan<sup>6,7</sup>; S.J. Diskin<sup>1,2,3,4,5,8,10</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.; 2) Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA.; 3) Center for Childhood Cancer Research, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA.; 4) Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA.; 5) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.; 6) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina-Chapel Hill, Chapel Hill, NC, 27599, USA.; 7) Lineberger Comprehensive Cancer Center, School of Medicine, University of North Carolina-Chapel Hill, Chapel Hill, NC, 27599, USA.; 8) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.; 9) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA.; 10) Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.

**Background.** Neuroblastoma is a malignancy that arises during sympathetic nervous system development and causes 12% of pediatric cancer deaths. Neuroblastoma has a complex genetic basis: approximately 1-2% of cases result from rare mutations that are inherited in a Mendelian autosomal dominant fashion, but the vast majority arise sporadically without family history. Although genome-wide association studies have identified common variants that confer risk for sporadic neuroblastoma at over a dozen loci, the contribution of rare variants has not yet been well-characterized due to limited sample sizes.

**Methods.** To identify large, rare germline copy number variants (CNVs) associated with neuroblastoma, we analyzed two independent, multi-ethnic cohorts totaling 5,585 neuroblastoma patients and 23,505 cancer-free controls genotyped on matched single-nucleotide polymorphism (SNP) arrays. Large (>500 kb), rare (<1% in controls) CNVs were tested for association using Fisher's exact test with a Bonferroni-corrected significance threshold. Whole-genome sequencing (WGS) was used to validate candidate CNVs and assess heritability in a subset of patients.

**Results.** Out of 40 CNVs tested, the only CNV reaching genome-wide significance in the discovery cohort was a 550-kb deletion on chromosome 16p11.2 (0.36% of cases and 0.02% of controls;  $p=8.3 \times 10^{-76}$ ). Notably, this CNV corresponds to a known microdeletion syndrome that confers risk for diverse developmental phenotypes including autism spectrum disorder and other neurodevelopmental disorders. We validated this association in the independent replication cohort ( $p=3.0 \times 10^{-6}$ ). Meta-analysis yielded a combined p-value of  $3.3 \times 10^{-9}$  and odds ratio of 13.9 (95% confidence interval=5.8-33.4), suggesting a substantial impact on neuroblastoma risk. The association remained significant when we restricted our analysis to individuals of European ancestry to mitigate potential confounding by population stratification (0.42% of cases and 0.03% of controls;  $p=4.10 \times 10^{-8}$ ). We validated the deletion in paired germline and tumor DNA from twelve patients by

WGS. Finally, WGS of four patient-parent trios revealed that the deletion primarily arose *de novo* without maternal or paternal bias.

**Conclusion.** This finding constitutes the first association of 16p11.2 microdeletion with cancer and suggests that disruption of the 16p11.2 region may dysregulate neurodevelopmental pathways that influence both neurological phenotypes and neuroblastoma.

## **PgmNr 793: Are we ready to offer polygenic risk score tests in the clinic? Expanding the utility of prostate cancer polygenic risk score in multiple ethnicities and clinical best practices.**

### **Authors:**

L. Sharma <sup>2</sup>; P. Hulick <sup>1</sup>; B. Helfand <sup>1</sup>; E. Levin <sup>2</sup>; A. Bolze <sup>2</sup>; K. Dunaway <sup>2</sup>; J. Lu <sup>2</sup>; J. Xu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) NorthShore University HealthSystem, Evanston, Illinois; 2) Helix, San Carlos, CA.

---

Pathogenic variants explain only a small portion of prostate cancer (PCa), however the utility of disease risk measured by polygenic risk score, has been consistently demonstrated for many common diseases, including PCa. Polygenic tests are currently being developed, offered and used in the clinical space. In collaboration with experts in the field, we have developed a comprehensive clinical polygenic PCa test that looks at SNPs stratified by risk among different ethnicities, namely, East Asian, African American and Non-Hispanic White. The individuals are categorized as average, moderately-high or high risk for developing PCa. Medical management recommendations like earlier and more frequent screenings for PCa are included in the clinical report. These recommendations are based on the individuals' risk, and generally follow the risk-based PCa screening guidelines of the U.S. Preventive Service Task Force. Here we propose best practices for developing a clinical polygenic risk score test in multiple ethnicities as well as provide recommendations on counseling patients that have undergone such tests.

For test development, we used the odds ratio (OR)-weighted and population-standardized genetic risk score (GRS) method. In this method, each SNP is first standardized against the general population and then multiplied for all SNPs. Thus its expected mean in the general population will always be 1.0, regardless of the number of SNPs used in calculation, and its values can be simply interpreted as relative risk to that of the general population. These two important features of population-standardized GRS makes the interpretation and implementation of an individual's risk comprehensible. We also applied two benchmarks, baseline and calibration, to ensure the reliability of GRS values.

This work aims to provide procedures to build a simple and interpretable risk score in multiple ethnicities and provide guidance on medical management. We will also discuss challenges and future work to address the gaps in the process of clinical test development and return of results.

## PgmNr 794: Germline genetic variation in 3 multiplex bladder cancer pedigrees.

### Authors:

A. Pemov<sup>1</sup>; T. Wegman-Ostrosky<sup>1,2</sup>; J. Kim<sup>1</sup>; B. Douthitt<sup>1</sup>; M.H. Greene<sup>1</sup>; N.E. Caporaso<sup>1</sup>; J.F. Fraumeni Jr.<sup>1</sup>; D.R. Stewart<sup>1</sup>; Cancer Genomics Research Laboratory, DCEG, Frederick National Laboratory for Cancer Research

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Division of Research, Instituto Nacional de Cancerologia, Ciudad de Mexico, Mexico

---

**Background.** Bladder cancer (BC) is the sixth most common cancer in the US. Familial BC is rare and implies a genetic component. An excess of BC is observed in Lynch syndrome (LS) (*MLH1*, *MSH2*, *MSH6*, *PMS2*). In addition, *GSTM1*, *NAT2*, *UGT1A8/10*, *ERCC2*, *CCNE1*, *TERT*, *TP63* and others are implicated by GWAS. **Methods.** To identify BC-associated germline variation, we exome-sequenced 2-3 affected individuals (with DNA availability) from 3 multiplex BC pedigrees, each with 3-6 affecteds; Family A was the first published BC pedigree (1967). We evaluated pedigree-specific segregating variants and pathogenicity using InterVar, ClinVar, REVEL, CADD, MetaSVM and HGMD. **Results.** We identified 72, 134 and 92 segregating variants in the 3 families respectively. We observed 1-4 likely pathogenic (LP) variants in each family; the majority of variants were of unknown significance (VUS), likely benign (LB) or benign (B). There was no overlap in the pedigrees among the segregating variants; 4 genes were shared by the pedigrees in pair-wise comparisons. Comparison of segregating variants in the 3 families to known significant BC GWAS hits identified only *NBN* in Family B. We observed 2 B variants in *MLH1* (although predicted deleterious by REVEL, CADD and MetaSVM and reported as disease-causing (DM) in HGMD) and 1 VUS in *MSH2* (although predicted deleterious by REVEL and MetaSVM and reported as likely DM in HGMD) in families B and C respectively. We identified 1 *NBN* (LP) variant (Nijmegen breakage syndrome 1 gene, critical for chromosomal stability) and 3 *BRCA1* (all B) alleles in pedigrees B and C respectively. No LS-associated variants were found in Family A; however, we observed a heterozygous LP variant in *CFTR* and variants in 3 DNA repair-related genes: *RECQL* (LB), *SPDR* (VUS) and *CUL4A* (VUS). Aside from DNA repair genes, we observed predicted deleterious alleles in *ME1*, *STEAP3* (Family A); *IDH1*, *ACADS* (Family B); and *VAV2*, *ITGAV*, *PRKCE* (Family C). **Conclusions.** We identified multiple variants segregating in familial BC that were pedigree-specific. Genes implicated in DNA repair and the DNA damage response may play an important etiologic role in BC. However, other biological pathways, e.g. angiogenesis (*VAV2*, *ITGAV*, *PRKCE*), apoptosis and cell cycle (*STEAP3*) and acetyl-CoA-mediated oxidation (*ME1*, *IDH1*, *ACADS*) may be involved as well.



## PgmNr 795: Polygenic risk score for prostate cancer by ~80 loci identified by genome-wide association study in Japanese population.

### Authors:

M. Fujita <sup>1</sup>; R. Takata <sup>1,2</sup>; A. Takahashi <sup>1</sup>; Y. Momozawa <sup>1</sup>; K. Maejima <sup>1</sup>; E.J. Saunders <sup>3</sup>; C.A. Haiman <sup>4</sup>; K. Matsuda <sup>5</sup>; S. Egawa <sup>6</sup>; O. Ogawa <sup>7</sup>; W. Obara <sup>2</sup>; Y. Kamatani <sup>1</sup>; S. Akamatsu <sup>1,7</sup>; H. Nakagawa <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 2) Department of Urology, Iwate Medical University, Morioka, Japan; 3) The Institute of Cancer Research, Sutton, UK; 4) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA; 5) Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan; 6) Department of Urology, Jikei Medical University, Tokyo, Japan; 7) Department of Urology, Kyoto University Graduate School of Medicine, Kyoto, Japan

---

Genome-wide association studies (GWAS) in the European and multi-ethnic populations have discovered ~170 loci associated with genetic risk for prostate cancer (PCa). In contrast, genetic studies on PCa risk in the Asian populations still lag behind. To expand our knowledge of PCa genetics in the Japanese population, we conducted a GWAS (5,088 cases and 10,682 male controls) and a replication study (4,818 cases and 73,261 male controls). As a result, we identified 14 novel SNPs were associated with PCa risk. They included rs1125927 (*TMEM17*,  $P = 3.95 \times 10^{-16}$ ), rs73862213 (*GATA2*,  $P = 5.87 \times 10^{-23}$ ), rs77911174 (*ZMIZ1*,  $P = 5.28 \times 10^{-20}$ ), and rs138708 (*SUN2*,  $P = 1.13 \times 10^{-15}$ ). For seven of the 14 novel loci, minor allele frequencies were significantly lower in Caucasians than in Japanese. Furthermore, by combining the newly identified SNPs with previously known ones, we computed polygenic risk score based on 82 SNPs and stratified the polygenic risk for Japanese PCa cases. Early-onset PCa cases and cases with family PCa history were significantly enriched in the genetically high-risk (upper 5%) population ( $P = 0.00221$  for cases with age < 60 years old and  $P = 0.00339$  for cases with PCa family history). Our study provides important insight into genetic mechanisms of PCa and facilitates PCa risk stratification in the Japanese population.

## PgmNr 796: Mitochondrial DNA haplogroups and susceptibility of neuroblastoma.

### Authors:

Chang <sup>1</sup>; H. Qu <sup>1</sup>; Y. Liu <sup>1</sup>; J. Glessner <sup>1</sup>; R. Komal <sup>2</sup>; S. Diskin <sup>2,3</sup>; J. Maris <sup>2,3</sup>; P. Sleiman <sup>1,2,3</sup>; H. Hakonarson <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.; 2) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA.

Neuroblastoma is a childhood malignancy that arises from the developing sympathetic nervous system. Although mitochondrial dysfunctions have been implicated in the pathophysiology of neuroblastoma, the role of mitochondrial DNA (mtDNA) has not been extensively investigated.

To determine whether mtDNA haplogroups influence risk of developing neuroblastoma. We conducted a case-control study to explore potential association of major European mtDNA haplogroups with the susceptibility of neuroblastoma. The mtDNA haplogroups were identified from genotyping SNP array data, of two independent neuroblastoma cohorts. Our results indicate that mtDNA haplogroup K was significantly associated with reduced risk of neuroblastoma in the discovery cohort consisting 1,474 cases and 5,699 controls (odds ratio 0.72,  $P = 0.004$ ). The association was replicated in an independent cohort (odds ratio 0.69,  $P = 0.014$ ) of 930 cases and 3,611 controls. Meta-analysis was performed by combining the two studies. The association remained highly significant after correction for multiple testing (odds ratio = 0.71,  $P = 2.02 \times 10^{-5}$ ,  $P_{corrected} = 1.62 \times 10^{-4}$ ).

To examine if haplogroup K is associated with a particular subtype of neuroblastoma, patients were further classified into low, intermediate, and high risk groups based on the Children's Oncology Group criteria, and each of three risk groups were tested against controls. The association between haplogroup K and high-risk neuroblastoma was significant in both discovery and replication samples. This association was further confirmed by a meta-analysis of the two datasets (odds ratio 0.58,  $P = 1.79 \times 10^{-4}$ ).

To analyze the impact of mtDNA haplogroups on gene regulation in neuroblastoma, transcriptome profiles of 99 neuroblastoma primary tumors, including five samples of haplogroup K, were characterized by RNA sequencing. This identified 708 differentially expression genes. Gene functional enrichment analysis further indicated that up-regulated genes are significantly enriched in 'immune system process' (GO: 0002376) and 'T cell aggregation' (GO: 0071593) pathways, while down-regulated genes are enriched in 'neurogenesis' (GO:00220088) and 'nervous system development' (GO: 007399) pathways in K haplogroup neuroblastoma when compared with non-K haplogroup neuroblastoma.

These findings provide new insights into the genetic basis of neuroblastoma, implicating

mitochondrial DNA encoded proteins in the etiology of neuroblastoma.

## PgmNr 797: Point mutation and somatic copy number analysis reveals drivers of clonal expansions in normal tissue adjacent to head and neck squamous carcinoma.

### Authors:

Y.A. Jakubek <sup>1</sup>; K. Chang <sup>1</sup>; S. Sivakumar <sup>1</sup>; Y. Yu <sup>1</sup>; J. Fowler <sup>1</sup>; C.D. Huff <sup>1</sup>; E. Vilar <sup>2</sup>; H. Kadara <sup>3</sup>; P. Scheet <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, U.T. MD Anderson Cancer Ctr, Houston, Texas.; 2) Clinical Cancer Prevention, U.T. MD Anderson Cancer Ctr, Houston, Texas.; 3) Translational Molecular Pathology, U.T. MD Anderson Cancer Ctr, Houston, Texas.

---

The presence of genetically distinct cell populations within an individual, known as clonal mosaicism, has long been cataloged in the context of cancer. Recent advances in technologies and statistical methods have enabled more sensitive detection of clonal mosaicism in non-cancerous tissues, of which blood has been the most comprehensively profiled.

To gain a broader view of clonal mosaicism in other tissues, we conducted a pan-tissue survey of somatic mega-base scale chromosomal alterations (sCNAs) in normal tissue adjacent to tumor (NAT) of TCGA patients using a sensitive haplotype-based method for detection of allelic imbalance. Our results showed significant differences in the rate and genomic distribution of sCNAs across different tissues. NAT samples from head and neck squamous carcinoma (HNSC) patients exhibited one of the highest mosaicism rates (25%) and enrichment for sCNAs in 9q ( $p = 1e-7$ ). To better catalog mosaicism in HNSC NAT, we expanded upon our survey of sCNAs to include single nucleotide variants and indels (SNVs). We generated SNV calls for NAT using exome sequencing data (59 NAT with paired blood; median = 18 SNVs in NAT). For contrast, we profiled SNVs in matched tumor tissues (58 sets; median = 179 SNVs).

The presence of sCNAs in NAT was positively associated with a higher SNV burden ( $p = 0.008$ ). For tumors, sCNA and SNV burden did not show an association. To gain insights into drivers of clonal expansions, we tested for genes under positive selection by modeling the ratio of synonymous to non-synonymous mutations. In HNSC NAT, *PPM1D* ( $q = 2.3e-4$ ) and *FAT1* ( $q = 0.02$ ), showed evidence of positive selection. *FAT1* mutations were exclusive to NAT with sCNAs. In the adjacent tumor tissues, we observed evidence of positive selection for *FAT1* ( $q = 5.6e-17$ ), but not for *PPM1D*. In NAT, sCNAs spanned SNVs in HNSC driver genes (*EP300*, *FAT1*, *NOTCH1*, *TP53*) that were not detected in the adjacent tumor. We tested for enrichment of HNSC driver mutations in NAT with sCNAs. Putatively deleterious SNVs in HNSC driver genes showed a positive association with sCNAs ( $p = 0.002$ ; adjusted for SNV burden). This association, and the overall higher SNV burden in NAT with sCNAs, suggests that sCNAs may serve as a biomarker for clonal expansions with a higher potential for progression towards malignancy. Our findings corroborate previous reports of a positive association between sCNA burden in oral premalignant lesions and progression towards invasive disease.

## PgmNr 798: Genetic association of cytochrome and DNA repair gene variants with concomitant cisplatin-based chemoradiotherapy (CRT) in cervical cancer patients.

### Authors:

Monisha. Banerjee <sup>1</sup>; Mohd. Abbas <sup>1</sup>; Maneesh. Gupta <sup>1</sup>; Kirti. Srivastava <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics Laboratory, Department of Zoology, University of Lucknow, Lucknow-226007, India; 2) Department of Radiotherapy, King George's Medical University, Lucknow-226003, India

---

### Abstract

Efficacy and toxicity due to Concomitant cisplatin-based chemoradiation (CRT) limit therapeutic effectiveness. This is attributed to genetic variability and is a major challenge in current clinical practice. Single nucleotide polymorphisms (SNPs) in cytochrome and DNA repair genes were analyzed for their association with treatment outcome of CRT in cervical cancer patients.

**Purpose of the Study:** Genetic associations may help to predict treatment outcomes and can be used as biomarkers to screen patients prior to treatment.

**Objective:** To genotype several SNPs in cytochrome (*CYP1A1*, *CYP2E1*, *CYP2D6*) and DNA repair (*XRCC1*, *XRCC2*, *XRCC3*, *XRCC4*, *RAD51*, *OGG1*, *MGMT*) genes in cervical cancer patients and correlate with treatment outcome viz. clinical response, vital status, toxicity, and overall survival after CRT.

**Methods used:** All patients received pelvic external beam radiotherapy (total dose of 50 Gy in 25 fractions) with weekly concomitant cisplatin (40 mg/m<sup>2</sup>) followed by three applications of high dose rate (HDR) intracavitary brachytherapy 7 Gy/fraction at a one-week interval. Weekly hemogram and renal function tests were done in all patients. SNPs in cytochrome and DNA repair genes were successfully genotyped by PCR-RFLP, ARMS-PCR methods and analyzed for their association with treatment outcome of CRT. The follow-up included response evaluation criteria in solid tumors (RECISTver1.0) after one month of treatment and primary endpoint was taken as overall survival (OS). Genotypic effects on OS were evaluated by Kaplan-Meier function and the Cox proportional hazards model. SPSSver21.0 was used with  $P < 0.05$  as statistically significant.

**Key Results:** Out of 227 cases, 85.5% were responders (CR+PR) and 14.5% poor/non-responders. Disease recurrence was significantly decreased in women with 'TA+AA' genotypes of *CYP2E1*+7678T/A while those with 'TC+CC' of *CYP1A1*-3801T/C showed better overall survival. Most of the cases with 'GC+CC' genotypes of *RAD51*+135G/C and 'GA+AA' of *XRCC1*+399G/A were found to be alive at the end of the study period. Individuals with GA+AA of *XRCC2*+31479G/A showed significantly increased risk of cervical cancer but no association with survival.

**Conclusion:** *RAD51*+135G/C and *XRCC1*+399G/A showed significantly better overall survival in cervical cancer patients after CRT. Such genetic studies will lead to the development of prognostic regimens and treatment strategies, a step towards personalized medicine.

## PgmNr 799: Common variant genetic background modifies risk of breast cancer among carriers of pathogenic germline risk variants.

### Authors:

J.R. Homburger <sup>1</sup>; C. Lai <sup>1</sup>; C.L. Neben <sup>1</sup>; A.Y. Zhou <sup>1</sup>; A.V. Khera <sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Color Genomics, Burlingame, California.; 2) Center for Genomic Medicine and Cardiology Division, Department of Medicine, Massachusetts General Hospital, Boston, MA; 3) Cardiovascular Disease Initiative of the Broad Institute of MIT and Harvard, Cambridge, MA; 4) Harvard Medical School, Boston, MA

DNA sequencing of individuals affected by breast cancer reveals a pathogenic or likely pathogenic (P/LP) variant in 5-10% of individuals. While carriers of a P/LP variant have a several-fold increased risk of breast cancer, many do not develop cancer. The causes of this incomplete penetrance are currently unknown. Here, we demonstrate that common variant background, assessed in the form of a genome-wide polygenic score (GPS), modifies the penetrance of P/LP variants for risk of breast cancer. We performed panel-based genetic testing on 13,914 women, 1,805 of whom had a P/LP variant in one of 12 genes known to be associated with hereditary breast cancer: *BRCA1*, *BRCA2*, *TP53*, *PTEN*, *STK11*, *CDH1*, *PALB2*, *CHEK2*, *ATM*, *NBN*, *BARD1*, or *BRIP1*. We also performed low-coverage whole genome sequencing at a mean depth of 0.2X and subsequent imputation to calculate a recently validated GPS for breast cancer (Mavaddat 2019 AJHG) on all 13,914 women. We found no significant interaction between GPS and presence of a P/LP variant ( $p=0.384$ ) in a multivariate logistic regression, suggesting that GPS is an independent risk factor for breast cancer. The odds ratio (OR) increase per standard deviation change in GPS was similar among P/LP carriers (OR per SD of 1.46, 95% CI 1.29-1.70) and non-carriers (OR 1.59, 95% CI 1.49-1.70). When subsetting to only *BRCA1* and *BRCA2* carriers, GPS increased breast cancer odds by 1.41 per SD increase (95% CI 1.15-1.72). We investigated whether this effect varied by gene, assessing GPS effects in four genes with over 200 P/LP carriers. We observed no interaction between GPS and carrier status in carriers of P/LP variants in *BRCA1* ( $p=0.31$ ), *BRCA2* ( $p=0.22$ ), *ATM* ( $p=0.43$ ), and *CHEK2* ( $p=0.07$ ). P/LP variant carriers in the lowest quintile of the GPS had risk comparable to that of noncarriers, with an odds ratio 1.29 (95% CI 0.92-1.79). By contrast, individuals who carried both a P/LP variant and a high GPS -- defined as the top quintile of the GPS -- had 3.73-fold increased odds for breast cancer (95%CI 2.94-4.74).

In conclusion, our results demonstrate that rare P/LP variants and common variant genetic background additively impact risk for breast cancer and that a lower GPS can offset the risk conferred by a P/LP variant. Our results suggest that a risk model that integrates both monogenic and polygenic risk can improve risk stratification and counseling for breast cancer, especially for individuals with moderate penetrance P/LP variants.

## PgmNr 800: Validation of PRS scores for breast cancer risk in carriers of *BRCA1/2* pathogenic variants.

### Authors:

A. Toland <sup>1,3</sup>; A. Willoughby <sup>1</sup>; A. Webb <sup>2</sup>; G. Brock <sup>2</sup>; D. Kinnamon <sup>3</sup>; R. Pyatt <sup>4</sup>; M. Daly <sup>5</sup>; K. Manickam <sup>6</sup>; T. Pal <sup>7</sup>; K. Confab <sup>8</sup>; M.B. Terry <sup>9</sup>; L. Senter <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cancer Biology & Genetics, Ohio State Univ, Columbus, Ohio.; 2) Center for Biostatistics, Ohio State Univ, Columbus, Ohio; 3) Division of Human Genetics, Department of Internal Medicine, Ohio State Univ, Columbus, Ohio; 4) Kean University, Union, NJ; 5) Fox Chase Cancer Center, Philadelphia, PA; 6) Nationwide Children's Hospital, Columbus, OH; 7) Department of Medicine, Hereditary Cancer Clinic, Vanderbilt University, Nashville, TN; 8) kConfab, Melbourne, Australia; 9) Department of Epidemiology, Columbia University, New York, NY

---

The penetrance of breast cancer in carriers of pathogenic variants (PVs) in *BRCA1* and *BRCA2* (*BRCA*) varies widely. Factors influencing penetrance include birth cohort, pregnancy and breast feeding history, and genetic modifiers. Genome-wide association studies have identified risk variants for breast cancer specific to *BRCA* PV carriers as well as for breast cancer in the general population. Polygenic risk score (PRS) models have been developed to predict cancer risk in *BRCA* PV carriers, however most of the validation studies have been done on women ascertained in high-risk clinical settings due to previous cancer diagnoses. The goal of this study was to evaluate PRS models and identify those most predictive of breast cancer risk in *BRCA* PV carriers unaffected at ascertainment. Genotypic and phenotypic data for evaluating models were ascertained through the OSU Clinical Cancer Genetics, Geisinger MyCode, kConfab, Inherited Cancer Registry and the Breast Cancer Family Registry studies. Only women who were unaffected at the time of ascertainment or testing were included in the study with the exception of some women with cancer in the Geisinger population-based cohort. As genotypes from existing GWAS level-data were used for analyses, individuals/studies that were missing more than 20% of genotypes for any given model were excluded. Multiple published models were used including *BRCA1* specific breast cancer, *BRCA2* specific breast cancer, 77-single nucleotide variant (SNV) and estrogen-receptor negative (ER-) models. For *BRCA1* carriers, 390 women and 148 breast cancer events were included in the analysis. The model with the highest Area under the Curve (AUC) was the ER- model (AUC=0.61, p-value<0.0001). For *BRCA2*, 369 women were included and 140 breast cancer events occurred. The most predictive model for *BRCA2* PV carriers was a *BRCA2* specific breast model (AUC=0.56, p-value=0.029). We also separated individuals into groups by risk (bottom 20%, top 20% and middle 60%) and performed Kaplan Meier analysis for cancer free survival. The ER- model separated the bottom 20% of *BRCA1* PV carriers from the other risk groups (p=0.037). The *BRCA2* specific model separated the top 20% risk group from the other groups, but the low risk group had more events than the middle group. In summary, PRS models for breast cancer risk in unaffected *BRCA* PV carriers show promise for more refined risk estimates, but there is room for improvement.



## PgmNr 801: Exploratory whole-transcriptome profiling of functional neural markers of cognitive function during aromatase inhibitor therapy for breast cancer.

### Authors:

J.D. Merriman <sup>1</sup>; R. Lam <sup>2</sup>; G. Yu <sup>1</sup>; C.M. Bender <sup>3</sup>; Y.P. Conley <sup>3</sup>; B.E. Aouizerat <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Meyers College of Nursing, NYU, New York, NY; 2) Bluestone Center for Clinical Research, NYU, New York, NY; 3) School of Nursing, University of Pittsburgh, Pittsburgh, PA

---

**Purpose:** Most breast cancer survivors are postmenopausal and receive endocrine therapy with an aromatase inhibitor (AI), which may influence cognitive function. Between 25% and 60% of breast cancer survivors report cognitive problems before and during AI therapy. Cognitive problems may include difficulty concentrating and remembering. Dysregulated gene expression may be associated with functional neural markers of these cognitive problems. The purpose of this study was to explore relationships between variability in whole-transcriptome differential gene expression and neural activation before and during AI therapy.

**Design:** This exploratory, hypothesis-generating study included 11 postmenopausal women with breast cancer recruited from a Comprehensive Cancer Center and 9 age-matched healthy women (controls). Patients were assessed before AI therapy, and 6 and 18 months after starting AI therapy. Controls were assessed at yoked intervals.

**Methods:** Blood was collected immediately proximal to neuroimaging using a 3-Tesla Siemens Trio magnetic resonance imaging scanner. RNA isolation and RNA-Seq data collection were completed in bulk. After quality control processing, sequence reads were aligned to the human genome and mapped to known genes based on annotations using Ensembl GRCh38 v95. Differential gene expression for patients versus controls and pathway analysis were done using DESeq2 and GAGE, respectively. SPM-12 was used to extract neural activation in brain regions of interest during the Emotional Faces N-Back task. Normalized gene counts were correlated with fMRI parameter estimates using Spearman's rank correlation.

**Results:** At baseline, pathways involved in sensory perception, stimulus detection, neural function, and cytokine function were perturbed (FDR  $q < 0.1$ ). *FDXR*, involved in mitochondrial function, was upregulated in patients versus controls (FDR  $q = 0.009$ ). We found trends for correlations between *FDXR* gene expression and neural activation in the left amygdala ( $\rho = 0.39$ ,  $p = 0.086$ ) and hippocampus ( $\rho = 0.38$ ,  $p = 0.098$ ). Six months after initiating AI therapy, pathways involved in sensory and neural function remained perturbed (FDR  $q < 0.05$ ). *FDXR* gene expression no longer differed between patients and controls. Eighteen months after beginning AI therapy, pathways involved in sensory function remained perturbed (FDR  $q < 0.001$ ).

**Conclusion:** Mitochondrial dysfunction may play a role in cognitive problems experienced by patients with cancer before systemic therapy.

## PgmNr 802: Enrichment of cancer related genes and signaling cascades in the acquired CNVs of human induced pluripotent stem cell lines.

### Authors:

K. Kanchan<sup>1</sup>; K. Iyer<sup>1</sup>; L.R. Yanek<sup>1</sup>; M.A. Taub<sup>2</sup>; C. Malley<sup>1</sup>; K. Baldwin<sup>3</sup>; L.C. Becker<sup>1</sup>; U. Broeckel<sup>4</sup>; L. Cheng<sup>1</sup>; C. Cowan<sup>5</sup>; M.D. Antonio<sup>6</sup>; K.A. Frazer<sup>6</sup>; I. Carcamo-Orive<sup>7</sup>; J.W. Knowles<sup>7</sup>; T. Quertermous<sup>7</sup>; G. Mostoslavsky<sup>8</sup>; G. Murphy<sup>8</sup>; M. Rabinovitch<sup>9</sup>; D.J. Rader<sup>10</sup>; M.H. Steinberg<sup>11</sup>; E. Topol<sup>12</sup>; W. Yang<sup>13</sup>; C.E. Jaquish<sup>14</sup>; I. Ruczinski<sup>2</sup>; R.A. Mathias<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Johns Hopkins Univ School of Medicine, Baltimore, Maryland.; 2) Department of Biostatistics, Bloomberg School of Public health, Johns Hopkins University, Baltimore, MD, USA; 3) Department of Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA, USA; 4) Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI, USA; 5) Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA; 6) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA; 7) Division of Cardiovascular Medicine, Cardiovascular Institute and Diabetes Research Center, Stanford University School of Medicine, Stanford, CA, USA; 8) The Center for Regenerative Medicine, Boston Medical Center, Boston University School of Medicine, Boston, MA, USA; 9) Department of Pediatrics, Stanford University, Stanford, CA, USA; 10) Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 11) Department of Medicine, Section of Hematology-Oncology, Boston University School of Medicine, Boston, MA, USA; 12) Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA, USA; 13) Penn Center for Pulmonary Biology and Institute for Regenerative Medicine, University of Pennsylvania, Philadelphia, PA, USA; 14) National Heart, Lung, and Blood Institute, NIH, Bethesda, MD, USA

We have previously described the genomic integrity of human induced pluripotent stem cells (hiPSCs) generated within the NHLBI NextGen program. Here, we focus on the genomic regions with instability to identify potential enrichments of biological processes and pathways. We performed gene set enrichment analysis (GSEA) for genes mapping to regions with acquired copy number variations (CNVs) in the hiPSCs. GSEA was done sequentially to understand differences in regions with repetitive CNV acquisition over the 149 lines with at least one acquired CNV: Set 1/ Set 2/ Set 3/ Set 4 included genes mapping to regions of acquired CNVs in  $\geq 1$ ,  $\geq 2$ ,  $\geq 3$  and  $\geq 4$  lines. Using the 'GO-Slim Biological Process' annotation and an FDR of 5%, the set of 470 annotated genes showed enrichment in a variety of biological processes. Among these, we identified two cancer related signaling cascades that were overrepresented: "JNK cascade" (GO:0007254) (FDR-P=0.01) and "I-kappaB kinase/NF-kappaB signaling cascade" (GO:0007249) (FDR-P=0.04). Interestingly, the fold enrichment (FE) of these cascades increased markedly as we went from Set 1 through to Set 4. For the JNK cascade the FE is 5.51 for the genes in Set 1, 8.07 for Set 2, 17.46 for Set 3, and 26.35 for Set 4. Similarly, for I-kappaB kinase/NF-kappaB cascade, FE increased from 4.64, to 8.07, 13.81 and 18.94, respectively. In addition, we also identified the overrepresentation of "IL-6 production" (GO:0032635) (FDR-P=0.002) with FE raising from 12.93 to 21.05, 45.52 and 68.72. Elevated levels of IL-6 stimulate hyperactivation of JAK/STAT3 signaling pathway that promotes tumor growth and progression. The sub-chromosomal

region chr20q11.2 was the most recurrent aberrant region and was detected in 16 hiPSC lines (15 duplications and 1 deletion). The core overlap region of the duplications encompasses 15 genes including cancer associated loci *ID1*, *BCL2L1* and *TPX2*. Overall, hiPSCs acquired deletions included tumor suppressor genes whereas the acquired duplications included oncogenes. In conclusion, our results confirm that some sub-chromosomal regions are particularly susceptible to genomic instability, and that in this large set of hiPSCs from NextGen, the acquired CNVs show a specific enrichment for cancer related genes.

## PgmNr 803: Comprehensive transcriptome-wide association study of prostate cancer risk reveals large number of non-coding gene associations.

### Authors:

N.B. Larson <sup>1</sup>; S.K. McDonnell <sup>1</sup>; Z. Fogarty <sup>1</sup>; S.N. Thibodeau <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Health Sciences Research, Mayo Clinic, Rochester, Minnesota.; 2) Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota.

---

The genetic etiology of prostate cancer is highly polygenic, and over 100 independent genome-wide significant risk loci have been identified by large GWAS to date. However, most of these loci do not lie in close proximity to protein-coding genes and are presumed to be regulatory in nature. Recent methodological advances for integrating tissue-specific gene expression eQTL analysis results with GWAS summary statistics has led to improved identification of disease gene associations mediated through transcriptomic dysregulation. In this study, we leveraged a large normal prostate tissue eQTL dataset to conduct a transcriptome-wide association study (TWAS) of prostate cancer risk based on SNP association results from the recent PRACTICAL case-control GWAS meta-analysis. Our eQTL dataset consisted of 471 samples of normal prostate tissue from prostate/bladder cancer patients with available RNA-Seq and imputed Illumina Infinium 2.5M genotype data. RNA-Seq data were processed using ENSEMBL gene annotation to comprehensively characterize expression patterns of all RNA species. In addition, 441 samples also had corresponding small RNA-Seq performed to characterize miRNA expression. We applied the FUSION TWAS software to train expression prediction models for all expressed genes with a minimum degree of heritability ( $h^2$  range: 0.02 – 0.98), resulting in 13049 total genes eligible for TWAS analysis. Based on a significance threshold of  $FDR < 0.05$ , we identified 703 significant genes associations with prostate cancer risk: 6 miRNAs, 204 lncRNAs (known and novel), 383 protein-coding genes, and 110 other RNA species. Among protein-coding genes, gene-set analysis via WebGestalt revealed significant enrichment ( $FDR < 0.05$ ) for genes related to Gene Ontology terms involving embryonic morphogenesis and development (e.g., GO:0060173, GO:0048568, GO:0035108). Among the 320 total non-coding genes associated in our TWAS, 83 (26%) had no local protein-coding gene TWAS hits within 500kb, suggesting the risk effects conferred at those loci are mediated through these non-coding transcripts. Other loci demonstrated clustered dysregulation of multiple proximal genes, with over 50 gene associations identified in the MHC region alone. These findings suggest that additional efforts untangling the gene expression regulatory networks of the prostate transcriptome, including the roles of non-coding genes, will be critical to fully elucidating the molecular mechanisms of prostate cancer risk.

## PgmNr 804: Genetic ancestry versus environmental differences in oncogenic mutations and driver genes in lung adenocarcinoma.

### Authors:

Y. He <sup>1</sup>; A. Gusev <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Harvard Medical School, Boston, MA.; 2) Dana Farber Cancer Institute, Boston, MA.

---

Cancers differ significantly between ethnic backgrounds, but the etiology and impact of these differences are poorly understood. We estimate genetic ancestry of over 1000 individuals with lung adenocarcinoma (LUAD) cancer in the Dana-Farber *Profile* tumor sequencing cohort and quantify the influence of genetic ancestry on recurrent somatic alterations. From roughly 700 mutation phenotypes (including somatic single-point mutations, insertions/deletions, and fusions), we identify 31 mutations significantly associated with at least one of five major ancestry groups (FDR adjusted  $p < 0.05$ ). We replicate the established association between mutation in EGFR gene and Asian ancestry ( $p = 2.2 \times 10^{-7}$ ). Surprisingly, we also find that non-Asian ancestries, such as Ashkenazi Jewish ( $p = 4.6 \times 10^{-3}$ ) and African ancestry ( $p = 4.0 \times 10^{-3}$ ), are significantly associated with EGFR as well. In addition to EGFR, we identify many novel genes significantly associated with ancestry differences (FDR adjusted  $p < 0.05$ ), including the association of Native American ancestry with KEAP1 mutations ( $p = 5.9 \times 10^{-5}$ ), and Asian ancestry with NOTCH2 mutations ( $p = 3.2 \times 10^{-5}$ ) and ARID1 mutations ( $p = 1.6 \times 10^{-3}$ ). Using multivariate models, we identify multiple mutations independently associated with each ancestry, including six mutations with Jewish ancestry, five mutations with African ancestry, and four mutations with European ancestry and East Asian ancestry. We also find significant differences between all continental ancestry groups for tumor mutational burden (TMB), which is an emerging prognostic indicator for immunotherapy and may point to ancestry-specific treatment selection criteria. Lastly, to better understand genetic versus environmental contributions, we compare discrepancies between self-reported ethnicity and derived ancestry in mutation phenotypes and find that most are explained by genetic differences. Taken together, our results suggest that somatic mutations differ between ancestries at many more genes than previously thought.

## PgmNr 805: Genome-wide association study identifies common genetic variants associated with prognosis of hepatocellular carcinoma.

### Authors:

H. Chen; D. Jiang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, China

---

### Aim

Hepatocellular carcinoma (HCC) is one of the most common cancers and the third most common cause of cancer mortality worldwide. The prognosis of HCC is poor, with a 5-year survival rate less than 30%. As an important aspect of tumor heterogeneity, genetic variation may influence the proliferation and metastasis of HCC. The aim of this study is to identify single nucleotide polymorphisms (SNPs) associated with overall survival of HCC patients.

### Methods

464 Chinese HCC patients with surgical treatments from two cohorts (OPS-1, OPS-2) and 641 Chinese HCC patients without surgical treatment from NonOPS cohort were enrolled in this study. Genome-wide genotyping analysis for the patients in OPS-1 and NonOPS was conducted using Illumina Human OmniExpress BeadChips, while for those in OPS-2 was conducted using Illumina Asian Screening Array. The following inclusion criteria were used to filter SNPs: 1) call rate > 95%, 2) minor allele frequency > 0.1 and 3) P for Hardy-Weinberg equilibrium > 0.001. Imputation analysis was done by referencing Haplotype Reference Consortium (HRC1.1). Cox proportional hazards regression under three models (additive, dominant and recessive) with adjustment for age at onset and sex, followed by meta-analysis, were used to analyze data.

### Results

After quality control, 4,830,338 SNPs, 4,908,943 SNPs and 3,133,930 SNPs were left for OPS-1, NonOPS and OPS-2, respectively. Of which, 95 SNPs in 19 blocks (additive model), 70 SNPs in 20 blocks (dominant model) and 145 SNPs in 42 blocks (recessive model) showed consistently suggestive significance ( $P < 0.05$ ) with the same direction of associations across the three cohorts. When a joint analysis of the patients in the three cohorts was performed, rs10020174, a SNP located in the intron region of *MARCH1* was significantly associated with overall survival of HCC under recessive model ( $HR = 8.50$ ,  $P_{meta} = 4.40 \times 10^{-10}$ ). Besides rs10020174, another SNP located in *PRKD1* (rs17096078) achieved a suggestive level of significance ( $HR = 1.29$ ,  $P_{meta} = 7.1 \times 10^{-6}$ ) under dominant model.

### Conclusion

By integrating three HCC cohorts, we identified several common genetic variants that potentially influence overall survival of HCC patients. Our findings may improve the understanding of the biology of the disease and suggest biological targets for pharmaceutical intervention.

## PgmNr 806: Germline testing in African American women with invasive breast cancer.

### Authors:

R. Ellsworth <sup>1</sup>; L. Lovejoy <sup>2</sup>; S. Rummel <sup>2</sup>; C. Shriver <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Breast Carr Project, Murtha Cancer Center, Windber, Pennsylvania.; 2) Chan Soon-Shiong Institute for Molecular Medicine, Windber. PA; 3) Clinical Breast Carr Project, Murtha Cancer Center/Uniformed Services University, Bethesda, MD.

---

**Background:** Although African ancestry is not included in the NCCN guidelines for BRCA1/2 genetic testing, African American women (AAW) are at increased risk of being diagnosed at a young age and/or with triple negative breast cancer. Improved understanding of the frequency and nature of cancer predisposition genes may improve the care of women with breast cancer and the development of risk reduction strategies, we evaluated the utilization and results of germline testing of cancer predisposition genes in AAW with invasive breast cancer.

**Methods:** 546 self-described AAW with invasive breast cancer enrolled in the Clinical Breast Care Project from 2001-2017. Eligibility status for genetic testing was assessed using the NCCN guidelines version 1.2019. In patients without clinical panel testing performed and with an available genomic DNA sample, targeted sequencing across 94 cancer predisposition genes was performed. .

**Results:** Of the 60% of patients meeting NCCN criteria, 37% of underwent clinical testing, resulting in a mutation frequency of 13%, including mutations in ATM (n=2), BRCA1 (n=7), BRCA2 (n=5) and TP53 (n=1). High-risk women tested in the research setting had a mutation frequency of 9% including BRCA1 (n=2) and BRCA2 (n=2) and single carriers of mutations in BLM, CHEK2, MLH1, MUTYH, PALB2, PMS2 and RECQL4. Within 111 low-risk women, the mutation frequency was 5% with mutations in BRCA2 (n=3), PMS2 (n=1), SBHD (n=1) and TSC2 (n=1). Six women had previously unreported variants in BRCA1, BRCA2, TP53, ALK and RB1 and 29% of the women had VUS in genes including BRCA1 and BRCA2.

**Conclusions:** Translating NCCN eligibility into pursuit of germline testing remains a challenge as only 33% of the high-risk women underwent clinical testing and 40% of the mutation carriers within the high-risk group were detected in the research setting. Furthermore, 5% of low-risk women were mutation carriers. In addition, while the majority (66%) of mutations detected were in BRCA1 or BRCA2, multi-gene testing allowed for an additional 50% of mutation carriers to be identified. Because identification of germline mutations may alter patient management and prevent secondary cancers, it is critical to identify and address barriers such as evolving NCCN guidelines, cost, and patient preference, to testing and optimize the selection of AAW for genetic testing.



## PgmNr 807: A custom genotyping array for detecting disease associations in men of African descent reveals population-level heterogeneity in the genetic risks of prostate and other cancers.

### Authors:

M. Harlemon <sup>1,2</sup>; O. Ajayi <sup>3</sup>; P. Kachambwa <sup>3</sup>; M. Kim <sup>1</sup>; C. Simonti <sup>1</sup>; M. Quiver <sup>1</sup>; T. Rebbeck <sup>4</sup>; L. Peterson <sup>3</sup>; J. Lachance <sup>1</sup>; MADCaP Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia; 2) Biological Sciences, Clark Atlanta University, Atlanta, Georgia; 3) The Centre for Proteomics and Genomics Research, Cape Town, South Africa; 4) Dana-Farber Cancer Institute and Harvard TH Chan School of Public Health, Boston, Massachusetts

---

Although prostate cancer is the leading cause of cancer mortality for African men, the vast majority of known disease associations have been detected in individuals who have European ancestry. Furthermore, most genome-wide association studies have used genotyping arrays that are hindered by SNP ascertainment bias. To overcome these disparities in genomic medicine, the Men of African Descent and Carcinoma of the Prostate (MADCaP) Network has developed a genotyping array that is optimized for studies of African populations. The MADCaP Array contains more than 1.5 million markers and an imputation backbone that successfully tags over 94% of common genetic variants in African populations. This array also has a high density of markers in genomic regions near known cancer associations, including 8q24. We assessed the effectiveness of the MADCaP Array by genotyping over 800 prostate cancer cases and controls from seven urban study sites in sub-Saharan Africa. We find that genes mirror geography: samples from Ghana and Nigeria cluster together and samples from Senegal and South Africa yield distinct ancestry clusters. Cases and controls are ancestry-matched. We identify highly differentiated loci in Africa (including rs3817963, rs2294008, rs5919432, and rs6465657), and include allele frequencies at cancer-associated loci as an additional, novel resource. Genetic risks of prostate cancer are heterogeneous across African populations: polygenic risk scores for prostate cancer are lower in Senegal and higher in Nigeria.

Significance: The MADCaP Array enables investigators to identify novel Africa-specific disease associations and to fine-map genetic loci that are associated with prostate and other cancers.



## PgmNr 808: MEDUSA: Phylogenetic analysis of mesothelioma tumours by multiregional sampling, whole exome sequencing, and copy number analysis.

### Authors:

E.J. Hollox <sup>1</sup>; A. Nakas <sup>2</sup>; M. Jama <sup>1</sup>; A. Sharkey <sup>2,3</sup>; A.G. Dawson <sup>2,3</sup>; J. Le Quesne <sup>3</sup>; L. Brannan <sup>4</sup>; F. Dudbridge <sup>4</sup>; D. Fennell <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Genome Biology, University of Leicester; 2) University Hospitals of Leicester NHS Trust, UK; 3) Leicester Cancer Research Centre, University of Leicester; 4) Department of Health Sciences, University of Leicester, UK

---

### Background:

The Mesothelioma evolution: Drugging somatic alterations (MEDUSA) project aims to investigate the genomic evolution and heterogeneity of malignant pleural mesothelioma and identify genomic changes early in mesothelioma evolution that can be targeted by drugs. For 20 malignant pleural mesothelioma patients, we have analysed the exomes of at least four regions of the tumour and paired whole blood.

### Methods

Using paired tumour-normal analysis with the software Sequenza, we have called copy number alterations specific to the tumour, and used the software Tumult to reconstruct a phylogeny of the tumour for each of the 20 patients.

### Results

We show that mesothelioma shows extensive heterogeneity in copy number changes, and accumulates typically between 100-200 copy number gains and losses while evolving in a branching pattern. We identify and validate copy number alterations that occur truncally, early in the evolution of the tumour, and are recurrent across patients, including homozygous loss of *CDKN2A* and *MTAP*, heterozygous loss of *MTOR*, and heterozygous loss of *BAP1*. Losses of these key genes are observed in some other patients, but only in a subset of regions, suggesting that they have occurred later in the evolution of the tumour compared to truncal changes.

### Conclusions

As truncal changes are likely to be present throughout the tumour, identifying them highlights potential Achilles' heels for drug targeting and treatment.

## PgmNr 809: Polygenic risk scores generated from European GWAS poorly predict prostate cancer risks in African populations.

### Authors:

M. Kim <sup>1</sup>; C. Simonti <sup>1</sup>; M. Harlemon <sup>1,2</sup>; L. Petersen <sup>3</sup>; I. Agalliu <sup>4</sup>; T. Rebbeck <sup>5</sup>; J. Lachance <sup>1</sup>; MADCaP Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA; 2) Clark Atlanta University, Atlanta, GA; 3) Centre for Proteomics and Genomics Research, Cape Town, South Africa; 4) Albert Einstein College of Medicine, Bronx, NY; 5) Dana Farber Cancer Institute, Harvard Medical School, Boston, MA

---

Men of African descent have the highest incidence and mortality of prostate cancer compared to men with other ancestries. However, most GWAS have used European study cohorts. Because of this, GWAS may do not always replicate well in other populations, and this can lead to misestimated polygenic risk scores (PRS). To alleviate disparities in genomic medicine, the Men of African Descent and Carcinoma of the Prostate (MADCaP) Network has developed a novel genotyping array that is optimized for studying prostate cancer in men of African descent. Applying the MADCaP Array to 399 African cases and 403 African controls, we test whether previously ascertained disease associations are able to successfully replicate across continents. For example, rs5919432 (a CaP-associated SNP that at Xq12) has elevated risk allele frequencies in MADCaP cases compared to African controls. We also examine whether prostate cancer associations near recombination hotspots are less likely to be replicated across continents. Using a curated set of 141 prostate cancer-associated loci from Schumacher et al. 2018, PRS of 802 men of African ancestry from the MADCaP Network and 240 men of European ancestry from 1000 Genomes Project were calculated. Comparing PRS distributions of individuals from seven African study sites and five European populations, we find that the predicted risks of prostate cancer are significantly higher for African men than European men. Among MADCaP study sites, individuals from Dakar, Senegal and Abuja, Nigeria show the lowest and highest predicted risk of prostate cancer, respectively. Overall, we find that PRS generated from European GWAS do a poor job of predicting prostate cancer risks in urban African populations. Using ADMIXTURE, we identify genetic ancestries that are correlated with increased risk of prostate cancer. Additional analyses include testing the effectiveness of PRS generated from non-European GWAS, assessing whether individuals with severe prostate cancer (i.e., higher Gleason scores) have higher PRS scores, and correcting for biases that can occur when generating PRS.

## PgmNr 810: SJLIFE WGS: A resource of curated genetic variation data generated from whole-genome sequencing of >3,000 long-term survivors of pediatric cancer.

### Authors:

S. Lei <sup>1</sup>; X. Zhou <sup>1</sup>; Z. Wang <sup>1</sup>; A. Thrasher <sup>1</sup>; M.C. Rusch <sup>1</sup>; K. Szlachta <sup>1</sup>; S.V. Rice <sup>1</sup>; S.W. Brady <sup>1</sup>; X. Ma <sup>1</sup>; Y. Yasui <sup>2</sup>; M.M. Hudson <sup>2</sup>; L.L. Robison <sup>2</sup>; J. Zhang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN; 2) Department of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN

---

Germline mutations in cancer predisposition genes contribute to ~10% of pediatric cancers, but their full impact on subsequent neoplasms and chronic health conditions after treatment is largely unknown. To enable genetic research on pediatric cancer survivors, we generated whole-genome sequencing (WGS) data at >30X coverage for 3006 participants enrolled in the St Jude Lifetime Cohort Study (SJLIFE) comprised of 2433 Caucasians, 308 African-Americans, 11 Asians, and others with admixed ethnicity. WGS was also performed for a community control group (SJLIFE controls) of 341 individuals without a history of childhood cancer. Genotype calls were determined by simultaneously optimizing the thresholds of read count and variant allele fraction and maximizing genotype concordance with SNP array. Over 93 million single-nucleotide variants (SNVs) were identified; 5.8 million of which were classified as “invalid” using a curation pipeline, which evaluates distributions of read coverage and variant allele fraction across samples, as well as Hardy-Weinberg disequilibrium and excess of heterozygosity for each site. Interestingly, 0.8 million of our invalid SNVs were considered passing quality control in gnomAD, suggesting that additional curation of public databases may be needed. Compared to gnomAD, over 30 million SNVs are unique in SJLIFE, the majority (80%) of which were singletons. We were able to replicate previously reported associations of *ARID5B* (17 SNPs in one LD block) with acute lymphocytic leukemia (ALL). However, we also found SNVs with significant associations by comparing to gnomAD, but the results were not replicated using other control cohorts. For example, the alternate allele frequency (AF) of rs374896001 is 0.09 in SJLIFE survivors of ALL while the AF in gnomAD, TOPMed, and SJLIFE controls are 0.0002, 0.27, and 0.16, respectively. These results demonstrate that SJLIFE WGS complements existing public data sets by enriching the content of genetic variation in human population. The data set not only serves as an important resource for genetic research on pediatric cancer and survivorship, but also provides a useful control for other genetic studies. SJLIFE genomic data are publicly available on the St. Jude Cloud ([https://platform.stjude.cloud/requests/cohorts?dataset\\_accession=SJC-DS-1002](https://platform.stjude.cloud/requests/cohorts?dataset_accession=SJC-DS-1002)) and full access has already been granted to 15 investigators from 11 research institutions.

## PgmNr 811: RCCD1: A potential pleiotropic target for ovarian and breast cancers.

### Authors:

N. Safi

[View Session](#) [Add to Schedule](#)

**Affiliation:** cedars Sinai Medical Center, Los Angeles, CA.

---

Authors: Nikoo Safi<sup>1</sup>, Justyna Kanska<sup>1</sup>, Stephanie Chen<sup>1</sup>, Jasmine Plummer<sup>1</sup>, Kruttika Dabke<sup>1</sup>, Siddhartha Kar<sup>2</sup>, and Simon Gayther<sup>1</sup>

<sup>1</sup>Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA;

<sup>2</sup>Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK;

### Introduction:

Genetic variance in RCCD1 is identified as a potential pleiotropic risk factor of breast, ovarian and prostate cancers. Quantitative trait association of RCCD1 with expression changes of susceptibility genes suggests the putative role of this gene in cancer acquisition. This finding merited experimental evaluation of this gene as a novel therapeutic target of breast and ovarian cancer. RCCD1 is involved in numerous biological pathways including estrogen receptor, P53 signaling, Epithelial-to-Mesenchymal Transition via TGFb, and DNA damage responses. Recently we used chromosome conformation capture (4C) to investigate the interaction of RCCD1 promoter with possible candidate causal variants at the 15q26 pleiotropic risk locus in breast and ovarian normal cell lines (MCF10A and FT246) and cancer cell lines (BT549, MCF7 and UWB1.289, KURAMOCHI). In this study, we created and genotyped cell line models of RCCD1 KO and OX to functionally validate the initial 4C results. Selected cell lines were analyzed for downstream targets of RCCD1. Immunofluorescent analysis of polyclonal RCCD1 KO cell lines showed that RCCD1 depletion is correlated with E-cadherin colocalization. E-cadherin is regulated by RCCD1 and is related to Mesenchymal-to-Epithelial Transition. Our results showed that RCCD1 expression is also correlating with P53 mutated, N-cadherin, and Vimentin, which suggest the causative role of RCCD1 in development of breast and ovarian cancer. Further experiments will validate if RCCD1 can act as a therapeutic or preventive target for these cancers.

## PgmNr 812: Tumor evolution in differentially growing metastases with heterogeneous immune microenvironments in a pancreatic neuroendocrine tumor patient.

### Authors:

Tang <sup>1</sup>; Y. Shao <sup>2</sup>; S. Bian <sup>3</sup>; D. Zhao <sup>4</sup>; L. Jiang <sup>2</sup>; T. Wen <sup>2</sup>; L. Chen <sup>2</sup>; L. Wan <sup>2</sup>; Z. Chen <sup>2</sup>; L. Xiang <sup>5</sup>; Q. Li <sup>6</sup>; X. Wang <sup>6</sup>; A. Jin <sup>7</sup>; F. Lv <sup>8</sup>; J. Li <sup>9</sup>; L. Chang <sup>9</sup>; Y. Guan <sup>9,10</sup>; R. Chen <sup>9</sup>; X. Xia <sup>9</sup>; X. Yi <sup>9,10</sup>; B. Hao <sup>3</sup>; S. Liao <sup>3</sup>; D. Li <sup>2</sup>; K. Ding <sup>11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetic Institute of Henan Province, Henan Provincial People's Hospital, Zhengzhou, Henan, Henan, China; 2) Department of Hepatobiliary Surgery, The First Affiliated Hospital of Chongqing Medical University, Chongqing, P.R. China; 3) Henan Medical Genetics Institute, People's Hospital of Zhengzhou University, Henan University, Zhengzhou, Henan 450003, P.R. China; 4) Department of Pathology, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, P.R. China; 5) The First Clinical College of Chongqing Medical University, Chongqing, P.R. China; 6) Department of Pathology, School of Basic Medicine, Chongqing Medical University, Chongqing, P.R. China; 7) Chongqing Key Laboratory of Basic and Clinical Tumor Immunology, Department of Immunology, School of Basic Medicine, Chongqing Medical University, Chongqing, P.R. China; 8) Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, P.R. China; 9) Geneplus Beijing Institute, Beijing, P.R.China; 10) Department of Computer Science and Technology, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xian, Shanxi, P.R.China; 11) Department of Bioinformatics, School of Basic Medicine, Chongqing Medical University, Chongqing, P.R. China

---

Pancreatic neuroendocrine tumor (PNET), arising from the endocrine tissues of the pancreas, is often diagnosed inadvertently and present with multiple metastases in the liver and local lymph node. The evolutionary principles underlie the emergence of metastatic traits in PNET remain exclusive. Here, we present a rare case of PNET with multiple metastases in the liver and lymph nodes over a period of 12 years. We investigated both genetic and immunologic components in longitudinally collected tumors using whole-exome and immuno-repertoire sequencing. A hypermutation in the primary tumor (i.e., a prevalence of 84.0 per MB) and significantly reduced mutation burden in metastases showed that metastasis occurred before the accumulation of mutation in the primary tumor. We demonstrated that polyclonal seedings have contributed to multiple metastases by both linear and branching spread patterns. In addition, the progression of metastatic tumors was modulated by heterogeneous immune microenvironments with varied T cell receptor (TCR) repertoires. Our findings provided an important implication for the evolution of long-term survival tumors, and the anti-tumor immunotherapy for metastatic tumors especially at later stages such as PNET.

## PgmNr 813: Scans of positive selection in urban African populations reveal recurrent targets of adaptation.

**Authors:**

M. Quiver; J. Lachance; MADCaP Consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** Georgia Institute of Technology, Atlanta, Georgia.

---

Most existing studies of natural selection in Africa have focused on rural populations, and the evolutionary genetics of urban African populations are relatively understudied. Here, over 800 individuals from sub-Saharan Africa were genotyped using the Men of AAfrican Descent and Carcinoma of the Prostate (MADCaP) Array. These individuals were sampled from seven urban study sites. MDS and ADMIXTURE analyses reveal three major clusters in the genetic data analyzed in this study. Using the MADCaP Array, population branch statistics (PBS) were calculated for 1.5 million loci to identify targets of positive selection in African populations. PBS scores were calculated for three evolutionary branches: Senegal, South Africa, and Ghana-Nigeria. Integrating scans of positive selection with PheWAS data from the UK Biobank, we identify phenotypes that tend to be associated with selected loci in urban African populations. Many positive selected loci are in the located MHC region of chromosome 6. This pattern is especially strong for the Senegalese branch. In addition, X chromosomal loci are enriched for high PBS scores. This may be due to selection acting on recessive alleles in hemizygous males. We also find that many positively selected loci appear to be involved in blood-related phenotypes, including hematocrit and neutrophil percentage. SNPs with high PBS scores in urban African populations include rs17202519, rs113430757, rs9786075, rs12851390, and rs9991733.

## PgmNr 814: Therapeutic yield of canonical splice variants among 325 solid tumor cases.

### Authors:

M. Prego; A. Hesse; K. Kelly; H. Chandok; H. Reddi

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical Laboratory, The Jackson Laboratory for Genomic Medicine, Farmington, CT.

---

In contrast to being routinely evaluated for disease association in inherited disorders, canonical splicing mutations are often not evaluated for precision medicine in clinical oncology, despite tumorigenic importance. To assess the relevance of canonical splice variants in clinical oncology, we retrospectively analyzed their therapeutic potential across 325 patients tested at our laboratory. Fastq data from our 212 gene ActionSeq™ NGS panel was mapped, aligned and called using a customized BWA/GATK pipeline and variants annotated using snpEff. The resulting +/- 1,2 variants were first functionally assessed by predictive tools and literature review, and then tier-classified according to joint AMP/ASCO/CAP cancer interpretation guidelines. Therapeutic significance (Tier I and II) was determined by the availability of indicated FDA approved therapies or clinical trials recruiting for aberrations of the identified gene. 46 splice variants were detected in 40/325 (12.3%) cases. 28/46 (61%) variants across 26/325 (8%) patients were determined to have therapeutic potential with clinical trials recruiting for aberrations in the gene. These 28 variants were distributed across a variety of primary tumors including brain (50%), lung and ovarian (11%), and fallopian tube (7%). *TP53* was the most recurrently mutated gene in this cohort with 9/28 splice variants detected, followed by *NF2* 3/28. 2/46 variants met Tier I criteria (*MSH6* and *BRCA2*), and 26/46 were classified as Tier II. A single variant was classified as Tier IV (benign) and the remainder (17/46) were classified Tier III (unknown significance). Putative access to 15 off-label FDA drugs in the clinical trial setting were identified for 13/28 variants. An average of 2.5 trials per variant were identified with 34 unique trial associations in total. These trials produced an average of 2.3 targeted drugs per variant with 30 unique drugs across 13 unique drug classes in total. 5/25 patients with therapeutic splice variants detected would have had a tier III only report otherwise. Similarly, 10/25 had 1 therapeutic variant reported originally. These results support the clinical validity of canonical splice evaluation, suggesting its incorporation in routine clinical reporting for solid tumors would be beneficial. This notion is bolstered by the fact that the yield rate for splicing mutations was greater than other commonly reported alterations in this cohort including CNV deletions, fusions, and in-frame indels.

## PgmNr 815: Landscape of germline genomic variations across 31 cancer types.

### Authors:

A. Abhyankar; H. Poisner; A. Rehman; S. Guha; A. Thomas; K. Wrzeszczynski; V. Jobanputra

[View Session](#) [Add to Schedule](#)

**Affiliation:** Molecular Diagnostics, New York Genome Center, New York, New York.

---

Germline predisposition alleles have been identified in several tumor suppressor genes and oncogenes. While most of the studies have focused on analysis of a single cancer type, germline variants explain only a small fraction of heritable cancers. To expand the current catalogue of germline cancer predisposition alleles we analyzed whole genome sequencing of tumor and matched normal samples on 200 clinical cases representing 31 different cancer types. Tumor tissue was sequenced to achieve mean genome coverage of 80x and normal tissue at 40x. Germline variant calling was performed on the matched normal samples to identify SNVs, indels, CNVs and structural variants. Mitochondrial DNA heteroplasmy and copy number were estimated using in-house developed pipeline. Variant calls were annotated, filtered and prioritized in relation to the tumor type. Prioritized germline variants were evaluated in the tumor sample to assess their biological impact. Following ACMG guidelines for variant classification we identified pathogenic and likely-pathogenic germline variants that were either shared by multiple samples with the same cancer type or shared across cancer types indicating common etiology. Within specific individual samples we identified synergistic combinations between germline and somatic variants and their possible mechanism of action. We further investigated the contribution of inherited mutations towards somatic tumor mutation burden, tumor heterogeneity and clonality. Germline variants were discovered in genes such as AIP, APC, BRCA2, CHEK2, EPHB2, GPR101, MLH1, MSH2, MSR1, MUTYH, PALB2, PIGN, PPP2R1B, PTEN, RAD51C, SDHA, SDHB, SMO, TP53 in diverse cancer types such as pancreatic cancer, renal cancer, lung adenocarcinoma, hepatocellular carcinoma, meningioma, papillary thyroid carcinoma, melanoma, neuroblastoma, osteosarcoma, and desmoplastic small round cell tumor. Here, we present our findings of how the identification and classification of inherited mutations can provide additional understanding of individual cancers and can potential clinically benefit the treatment of cancer patients.



## PgmNr 816: Integrating multi-platform genomic datasets for kidney renal clear cell carcinoma subtyping using stacked denoising autoencoders.

### Authors:

T. Gu <sup>1</sup>; X. Zhao <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) ICBR, University of Florida, Gainesville, FL; 2) Department of Ophthalmology & Visual Sciences, University of Michigan, Ann Arbor, MI

---

Clear cell renal cell carcinoma (ccRCC) is highly heterogeneous and is the most lethal cancer of all urologic cancers. We developed an unsupervised deep learning method, stacked denoising autoencoders (SdA), by integrating multi-platform genomic data for subtyping ccRCC with the goal of assisting diagnosis, personalized treatments and prognosis. We successfully found two subtypes of ccRCC using five genomics datasets for Kidney Renal Clear Cell Carcinoma (KIRC) from The Cancer Genome Atlas (TCGA). Correlation analysis between the last reconstructed input and the original input data showed that all the five types of genomic data positively contribute to the identification of the subtypes. The first subtype of patients had significantly lower survival probability, higher grade on neoplasm histology and higher stage on pathology than the other subtype of patients. Furthermore, we identified a set of genes, proteins and miRNAs that were differential expressed (DE) between the two subtypes. The function annotation of the DE genes from pathway analysis matches the clinical features. Importantly, we applied the model learned from KIRC as a pre-trained model to two independent datasets from TCGA, Lung Adenocarcinoma (LUAD) dataset and Low Grade Glioma (LGG), and the model stratified the LUAD and LGG patients into clinical associated subtypes. The successful application of our method to independent groups of patients supported that the SdA method and the model learned from KIRC are effective on subtyping cancer patients and most likely can be used on other similar tasks. We supplied the source code and the models to assist similar studies at [https://github.com/tjgu/cancer\\_subtyping](https://github.com/tjgu/cancer_subtyping).

## PgmNr 817: Deep learning prediction of chemotherapy response using multi-omics features.

### Authors:

S. Guo<sup>1</sup>; S.J. Schrod<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, WI; 2) Computation and Informatics in Biology and Medicine, University of Wisconsin-Madison, Madison, WI

---

Combining deep learning algorithms with multi-omics data provide an opportunity to improve precision medicine, especially for personalized therapy treatment. In this study, we investigated the performance of deep-learning algorithms applied to multi-omics genomic data from the Cancer Genome Atlas (TCGA) project. Drug response status including complete response (n=815), partial response (n=102), stable disease (n=182) and clinical progressive (n=892) across 32 cancers were evaluated. For computational efficiency and classifier performance, we proposed an approach using Sure Independence Screening (SIS) and Gini impurity index (SII) for feature selection from ultrahigh dimensional genomic feature set followed by conditional generative adversarial networks (CGAN) to produce predictive models for chemotherapy response, internally validated by 10-fold cross-validation. mRNA, miRNA, and methylation datasets were used both independently and jointly. For the miRNA-based prediction, we identified 23 highly informative features, including miR-141, miR-200c, miR-205, miR-9, and miR-338, which, in combination, produced an AUC=0.64 for discriminating complete and partial response from stable disease and clinical progressive endpoints. Using mRNA-seq data, we identified 264 highly informative features including *MYCBP*, *KLF15*, *IGIP*, and *GRIA3*. The mRNA-based predictive model yielded an average AUC=0.71. The DNA methylation model attained a higher level of performance with AUC=0.81 (95% CI: 0.78-0.84). Combining mRNA-seq, miRNA, and methylation data improved the predictive performance with an accuracy of 84.2% and an AUC=0.86 (95% CI: 0.82-0.90). Functional enrichment analysis for mRNA and DNA methylation selected features showed enrichment for the antioxidant response pathway and basal transcription factors associated with the platinum drug resistance (PDR) pathway. Importantly, our chemotherapy response classifier substantially outperforms a predictive model focused on mRNA and methylation features within PDR pathway (hsa01524 in KEGG) AUC=0.62, demonstrating that taking a broad approach using genome-wide multi-omics data dramatically improves discrimination. In summary, this work demonstrates that applying deep-learning algorithms to multi-omics data can generate informative predictive models for chemotherapy response across numerous cancers.

## PgmNr 818: Identifying novel therapeutic targets in cancers by analysis of genes insensitive to gene dosage changes.

### Authors:

V. Mohanty <sup>1</sup>; F. Wang <sup>1</sup>; G. Mills <sup>2</sup>; K. Chen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, TX.; 2) Cell Development and Cancer Biology, Oregon Health & Science University, Portland, OR

---

Structural changes introduced by chromosomal instability are a hallmark of cancer. They are thought to result in tumor growth advantage by modulating the copy number of driver genes (oncogenes and tumor suppressors). Unlike point mutation and indels, these structural changes are generally large and influence the copy number of hundreds of genes including both driver and non-driver genes. Understanding the functional consequences of these structural changes is one of the fundamental challenges hindering the advances of genomic medicine.

We find that many of these induced gene dosage changes can be detrimental to tumor fitness. The expanding tumor clones often uncouple the expression of these genes from their copy numbers through additional regulatory changes. We hypothesize that by restoring the expressions of these genes, we can induce a flux of signaling detrimental to the tumor, thus achieving therapeutic benefits.

To discover the relevant regulatory changes, we use machine learning to build cancer specific regulatory networks from gene expression, methylation, miRNA, promoter and enhancer activity data. We model the effects of perturbations on the transcription factors that regulate the expressions of uncoupled genes on various phenotypes (i.e., immune infiltration, proliferation and cell cycle) determining tumor fitness. We introduce a Bayesian approach that quantifies the degree to which a copy number predicts the expression of a gene in an individual sample, allowing us to identify targets in individual samples. Finally, We systematically examine autologous SNP 6.0 array and RNA sequencing data in the cancer genome atlas (TCGA) using our approach. We discover a list of novel targets such as *TAP1/2*, whose gene expressions are consistently uncoupled from copy numbers in an appreciable number of patients. Functional validations are underway as part of the NCI cancer target discovery and development (CTD<sup>2</sup>) project.

## **PgmNr 819: Germline testing in colorectal cancer: Increased yield and precision therapy implications of comprehensive multigene panels.**

### **Authors:**

S. Yang; E. Esplin; S. Michalski; D. Pineda; S. Lincoln; R. Nussbaum

[View Session](#) [Add to Schedule](#)

**Affiliation:** Invitae, San Francisco, California.

---

**Background:** Studies suggest that the prevalence of germline abnormalities in homologous recombination deficiency (HRD) genes and other cancer genes not traditionally associated with colorectal cancer (CRC) may be more common in patients with CRC than previously appreciated. Herein, we investigate the landscape of pathogenic/likely pathogenic (P/LP) variants in CRC and the efficacy of comprehensive multigene panels identifying candidates for precision therapies.

**Methods:** DNA sequencing and exon-level copy number analysis were performed at a commercial diagnostic laboratory in a convenience sample of >9000 patients (pts) referred because of a personal history of colon cancer between 2013 and 2018. The genes requisitioned varied but consistently included 14 genes on a hereditary CRC panel; the patient data were de-identified and further analyzed for all 83 genes on a large hereditary cancer syndrome panel under an IRB-approved protocol.

**Results:** P/LP findings were identified in 2101 of 9669 pts (21%), 1838 (19%) pts when MUTYH heterozygotes are excluded. When restricted to five Lynch syndrome (LS) genes, only 9% of patients had a P/LP finding, which increased to 15% when 19 guidelines-based CRC genes were assessed. 137 pts (1.4%) had two or more P/LP variants. P/LP variants were in MLH1, MSH2, MSH6, PMS2, CHEK2, APC, MUTYH, BRCA2, ATM, BRCA1, PALB2, RAD50, BRIP1, TP53, EPCAM, among others, of which 1.4% were BRCA1/2.

When patients heterozygous for recessive conditions are excluded, 8% had P/LP variants in genes classically associated with CRC (i.e. LS genes, APC) or were biallelic for MUTYH, and 9% had P/LP variants in cancer risk genes not commonly associated with CRC. Using a comprehensive multigene panel approach, P/LP variants in genes with known therapeutic implications, such as in HRD and mismatch repair deficiency, were detected in 1408 (14%) of patients, and 1670 (17%) had P/LP variants in genes with established clinical management guidelines.

**Conclusions:** This study suggests that 1 in 5 patients with CRC harbor actionable germline variants, up to one-half of which remain undetected when only LS genes are tested. Over 50% of patients eligible for germline-based precision cancer therapy would be missed by germline testing limited to the LS genes. Comprehensive panel testing identified candidates for precision treatment and established management recommendations, and have clinical implications for both pts and their at-risk family members.

## PgmNr 820: Computational identification of mutator-derived lncRNA signatures of genome instability for improving clinical outcome of breast cancer.

### Authors:

H. Zhao <sup>1,3</sup>; S. Bao <sup>1</sup>; J. Yuan <sup>1</sup>; D. Fan <sup>1</sup>; J. Liu <sup>2</sup>; N. Wu <sup>3,4,5</sup>; Z. Wu <sup>4,5,6</sup>; G. Qiu <sup>3,4,5</sup>; J. Su <sup>1</sup>; M. Zhou <sup>1</sup>; DISCO (Deciphering disorders Involving Scoliosis and Comorbidities) study group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Biomedical Engineering, Wenzhou Medical University, Wenzhou 325027, P. R. China; 2) Department of Breast Surgical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, P. R. China; 3) Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China; 4) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Beijing 100730, China; 5) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences, Beijing 100730, China; 6) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China

---

**Motivation:** Emerging evidence revealed the critical roles of long non-coding RNAs (lncRNAs) in maintaining genomic instability. However, genome instability-associated lncRNAs and their clinical significance in cancers remain largely unexplored.

**Results:** We developed a mutator hypothesis-derived computational frame combining lncRNA expression profiles and somatic mutation profiles in a tumor genome and identified 128 novel genomic instability-associated lncRNAs in breast cancer as a case study. We then constructed a genome instability-derived two lncRNA-based gene signature (GILncSig) which stratified patients into high- and low-risk groups with significantly different outcome and was further validated in the independent patient cohort. Furthermore, the GILncSig correlated with genomic mutation rate in both ovarian cancer and breast cancer, indicating its potential as a measurement of the degree of genome instability. Interestingly, we found that the GILncSig was able to divide TP53 wild-type patients into two risk groups, with the low-risk group showing significantly improved outcome and the high-risk group showing no significant difference compared with those with TP53 mutation.

**Conclusions:** This study provided a critical approach and resource for further studies examining the role of lncRNAs in genome instability, and identified a genome instability-derived lncRNA signature which may have important value in outcome prediction and customized decision-making in breast cancer.

**Keywords:** genome instability; mutator phenotype; long non-coding RNAs

## **PgmNr 821: Prevalence of *UGT1A1\*28*, *UGT1A1\*93*, *UGT1A1\*60*, *ABCC5*, and *ABCG1* and their impact on toxicity in Iraqi cancer patients treated with irinotecan.**

### **Authors:**

A. Al-Talqani; S. Kathem

[View Session](#) [Add to Schedule](#)

**Affiliation:** College of Pharmacy, University of Baghdad, Baghdad, Iraq

---

Irinotecan-induced neutropenia and diarrhea are life-threatening toxicities that limit the use of the drug. These toxicities influenced by distinct interindividual variability due to gene polymorphism. In this prospective, longitudinal and multicenter study, 64 patients were participated to estimate the prevalence of *UGT1A1\*28*, *UGT1A1\*93*, *UGT1A1\*60*, *ABCC5* and *ABCG1* allele polymorphisms in Iraqi cancer patients treated with irinotecan-based therapy. In addition, the study aims to assess the relation between the polymorphic alleles and toxicity in standard- and reduced- dose of irinotecan. We showed that the prevalence of one vs two polymorphic allele of *UGT1A1\*28*, *UGT1A1\*93*, *UGT1A1\*60*, *ABCC5* and *ABCG1* were (35.9 vs 9.4 %), (54.7 vs 3.1 %), (17.9 vs 4.9 %), (39.1 vs 40.6 %) and (48.4 vs 0.0 %) respectively. Results revealed a significant ( $P<0.05$ ) relation between *UGT1A1\*28* allele and incidence of severe diarrhea and neutropenia (grade 3 and 4). In addition, we report a significant ( $P<0.05$ ) relation between the occurrence of *UGT1A1\*93* allele or *UGT1A1\*60* allele and incidence of severe (grade 3 and 4) neutropenia or diarrhea, respectively. Further, when Irinotecan dose reduced by 25-35%, we reported a significant ( $P<0.05$ ) reduction in the incidence of severe (grade 3 and 4) neutropenia and diarrhea in patients having *UGT1A1\*93* allele polymorphism, while non-significant change observed with other polymorphic alleles. In summary, we found that polymorphic alleles of *UGT1A1\*28*, *UGT1A1\*93*, *UGT1A1\*60* are strong determinants of toxicity that should be well considered in the treatment with irinotecan that necessitates dose modification.

## PgmNr 822: Results from clinical exome sequencing for cancer risk assessment in primary care patients who screen positive for genetic risk.

### Authors:

L.M. Amendola <sup>1</sup>; B.A. Rolf <sup>1</sup>; E. Shuster <sup>2</sup>; K. Mittendorf <sup>2</sup>; T.L. Kauffman <sup>2</sup>; K. Anderson <sup>3</sup>; S. Okuyama <sup>3</sup>; K. Goddard <sup>2</sup>; B. Wilfond <sup>4</sup>; M. Dorschner <sup>5</sup>; G.P. Jarvik <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Division of Medical Genetics, University of Washington Medical Center, Seattle; 2) Kaiser Permanente Center for Health Research, Portland; 3) Denver Health and Hospital Authority, Denver; 4) Seattle Children's Research Institute, Seattle; 5) Department of Pathology, University of Washington Medical Center, Seattle

---

Background: Identifying hereditary cancer conditions provides the opportunity for screening or surgical intervention in those at risk. The Cancer Health Assessments Reaching Many (CHARM) study is performing clinical exome sequencing for 880 participants who score at increased risk for hereditary cancer using the B-RST<sup>TM</sup> 3.0 (positive or moderate risk) or PREMM<sub>5</sub> ( $\geq 2.5\%$  risk threshold) screening tools for hereditary breast and ovarian cancer syndrome and Lynch syndrome, respectively, or have limited family information. Here we summarize the findings reported for the first 200 CHARM study participants. Methods: Clinical exome sequencing includes annotation of 39 high and moderate penetrance hereditary cancer associated genes and 78 adult onset, medically actionable incidental finding (IF) genes. All pathogenic and likely pathogenic (P/LP) variants identified in hereditary cancer associated genes are reported; variants of uncertain significance (VUS) are only reported in genes related to the participant's history. P and loss-of-function LP variants are reported in IF genes. Results: Of 200 participants, 142 (71%) screened at increased risk on the B-RST<sup>TM</sup> 3.0 and/or PREMM<sub>5</sub> tool. Twenty of 200 participants (10%) had a P/LP variant in a hereditary cancer gene reported; 9 were in genes related to their personal or family history. Five of the remaining 11 participants with a P/LP cancer associated gene variant had limited family information. Six of 118 (5%) participants with a positive or moderate risk B-RST<sup>TM</sup> 3.0 screen had a P/LP breast or ovarian cancer variant compared to 3 of 82 (4%) participants who did not score at increased risk. Three of 51 (6%) participants with a positive PREMM<sub>5</sub> screen had a P/LP colon cancer variant compared to 6 of 149 (4%) participants who did not score at increased risk. Fifteen participants (8%) had a cancer risk related VUS reported. Two participants (1%) had an IF P variant unrelated to cancer reported. Discussion: The rate of findings reported to date in the different results categories is consistent with previous estimates for hereditary cancer testing and non-cancer IFs. While the proportion of participants with a P/LP variant was slightly higher in those who scored at increased risk, half of the P/LP results were in individuals without a known related personal or family history. These findings have implications for implementing cancer risk assessment screening tools and targeting genetic testing at a population level.



## **PgmNr 823: *Functional annotation of cancer somatic variants to guide radiotherapeutic decisions: The radiogenomic atlas.***

### **Authors:**

J.A. Castrillon <sup>1</sup>; B.D. Yard <sup>1</sup>; A.P. Petty <sup>1</sup>; M.E. Abazeed <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Translational Hematology Oncology Research, Lerner Research Institute, Cleveland, Ohio. 9500 Euclid Avenue/R40, Cleveland, Ohio 44195, USA.; 2) Department of Radiation Oncology, Cleveland Clinic, 9500 Euclid Avenue/T2, Cleveland, Ohio 44195, USA.

---

**Background:** Despite effectively guiding targeted drug therapy, genetic biomarkers do not currently inform the personalized use of radiotherapy. An understanding of the interplay between the cancer genome and ionizing radiation has the potential to predict cellular response to radiotherapy and therefore guide its clinical use.

**Methods:** We previously developed a high-throughput cancer cell line profiling platform that measures radiation survival and leverages cancer genomic data to advance knowledge of radiation tumor biology. We analyzed the radiation response of 533 cancer cell lines annotated by the Cancer Cell Line Encyclopedia (CCLE) and identified single-nucleotide variants (SNVs), copy number alterations and gene expression changes associated with radiation sensitivity. To annotate functional variants, we developed a new profiling schema that permits the unary assessment of SNVs in SV40 immortalized BEAS2B lung cells. To achieve this, vector constructs from 96 genes, comprising >500 unique SNVs, were generated using a high-content, site-directed mutagenesis approach that creates and integrates the variant into lentiviral vectors resulting in its stable expression under a PGK or EF1a promoter. The radio-protective *NFE2L2 E79K* gain-of-function variant and isogenic controls were used to calibrate relative radiation survival of individual SNVs tested in multiple tranches. Nominated SNVs were validated by colony forming assay. Selective SNVs that were confirmed to alter radiation sensitivity were profiled by 28 bioactive compounds in order to query the genetic dependency similarities between commonly used anti-tumor drugs and radiation.

**Results:** We show that top radioresistant alleles include both common and rare cancer variants. SNVs within *ARAF*, *BRAF*, *CTNNB1*, *EGFR*, *MAP2K1* and *PIK3CA*, among others were identified. A subset of KEAP1 loss of function variants were found to confer a dominant-negative phenotype by stabilizing NRF2 levels. Using our integrated drug sensitivity platform, we show that the association between the sensitivities of cells to radiation and anti-cancer drugs vary on the basis of the SNV in a drug class specific manner (e.g. DNA crosslinkers, epigenetic modifiers, mitotic spindle and DNA replication inhibitors, *et cetera*).

**Conclusion:** We present the largest radiogenomic profiling effort of cancer SNVs to date. Our atlas is poised to serve as a key resource that will help guide radiation treatment delivery.



## PgmNr 824: Exploring the role of mutant *p53* in somatic osteosarcoma.

### Authors:

D. Chachad <sup>1</sup>; H. Beird <sup>2</sup>; C. Vera <sup>3</sup>; W. Wang <sup>3</sup>; G. Lozano <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, UT MD Anderson Cancer Center, Houston, TX.; 2) Genomic Medicine, UT MD Anderson Cancer Center, Houston, TX; 3) Bioinformatics and Computational Biology, UT MD Anderson Cancer Center, Houston, TX

---

Osteosarcomas are malignant bone tumors that show rapid progression and pulmonary metastases. Studies in the field of osteosarcoma have not conclusively addressed important questions cooperating alterations that lead to tumorigenesis and metastasis, or the cell of origin for osteosarcomas. However, a growing body of research shows *p53* to be one of the most frequently mutated genes in osteosarcomas. Spontaneous mutations in the *p53* DNA-binding domain are missense in nature and exhibit gain of function (GOF) activities which are mediated mostly through its interaction with other transcriptional regulators, via its functional and highly active transactivation domain. *p53* mutants are classified as conformational (*p53*<sup>R172H</sup>) or contact mutants (*p53*<sup>R245W</sup>). Previous studies indicate that these mutants manifest transformation in unique ways. However, the mechanisms by which these hotspot mutants exert their tumorigenic potential remains to be evaluated. We hypothesize that *p53*<sup>R172H</sup> and *p53*<sup>R245W</sup> are independent tumor-initiating events that regulate different transcriptional programs for osteosarcoma formation. We have developed genetically engineered mouse cohorts that express either *p53*<sup>R172H</sup> or *p53*<sup>R245W</sup> specifically in osteoblasts after Cre-mediated recombination, resulting in 100% penetrant osteosarcomas. Additionally, after Cre-mediated recombination, while one allele of *p53* expresses the mutant, the other allele is deleted, which allows us to study GOF for that specific mutant.

The murine primary tumors exhibit extensive copy number variation thereby recapitulating human osteosarcomas. Preliminary data from our cohorts show differences in terms of survival and metastatic incidence, with *p53*<sup>R172H</sup> being more potent than *p53*<sup>R245W</sup>. RNA-seq analyses of primary tumors also suggest that the transcriptomes regulated by these respective mutants are different. Additionally, immunohistochemistry analyses show that *p53*<sup>R245W</sup> is more stable than *p53*<sup>R172H</sup>. Studies are ongoing to identify dysregulated pathways and mechanisms by which these mutations in *p53* give rise to metastatic osteosarcoma. These studies will help us identify novel targets mediating transcriptional GOF of respective mutants.

## PgmNr 825: Common non-coding genetic variation as a clinically actionable biomarker of melanoma immunotherapy toxicity.

### Authors:

V. Chat<sup>1,2,3</sup>; R. Ferguson<sup>1,2,3</sup>; L. Morales<sup>1,2,3</sup>; K. Monson<sup>1,2,3</sup>; E. Cohen<sup>1,2,3</sup>; E. Esteva<sup>1,2,3</sup>; U. Moran<sup>1,3,4,5</sup>; A. Pavlick<sup>3,4</sup>; P. Ascierto<sup>6</sup>; I. Osman<sup>1,3,4,5</sup>; J. Weber<sup>1,3,4</sup>; T. Kirchhoff<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Perlmutter Cancer Center, NYU Langone Health, New York, New York.; 2) Departments of Population Health and Environmental Medicine, New York University School of Medicine, New York, New York; 3) The Interdisciplinary Melanoma Cooperative Group, New York University School of Medicine, New York, New York; 4) Department of Medicine, New York University School of Medicine, New York, New York; 5) Ronald O. Perleman, Department of Dermatology, New York University, New York, New York; 6) Istituto Nazionale Tumori-Istituto di Ricovero e Cura a Carattere Scientifico, Fondazione G. Pascale, Naples, Italy

While immunotherapy (IT) improved the clinical outcome of metastatic cutaneous melanoma (CM), ~60% of patients do not respond to treatment. Furthermore, 65-80% of patients treated with IT experience immune-related adverse events (irAEs). We have recently shown that host immunity may affect IT response and prognosis by demonstrating the association of these outcomes with immune-related expression quantitative trait loci (ieQTLs). In this study, we investigated whether ieQTLs also impact risk of irAEs in IT-treated CM patients. By interrogating a healthy twin cohort expression dataset (MuTHER), we have identified 50 ieQTLs most significantly associated with the expression of 265 immune genes. These 50 SNPs were genotyped using the MassARRAY system in 195 anti-CTLA-4-treated patient samples collected as part of a multi-institutional collaboration. Multivariable logistic regression adjusting for age at treatment, study center and gender was used to test for association of germline variants with irAEs. We found a significant association with severe irAEs in anti-CTLA-4-treated patients for rs7036417 (OR=8.62; 95%CI=3.02-24.64; p=5.77E-05). In this analysis, having two copies of the alternate allele of rs7036417 results in patients being 8.6 times more likely to develop severe irAEs. Having two copies of rs7036417 in lymphoblastoid cell lines is associated with higher expression of SYK, suggesting that high SYK expression is associated with increased risk of severe irAEs. In this study, we report that rs7036417, an eQTL in SYK, is associated with an increased risk of severe irAEs. SYK is known to interact with B-cell and Fcγ receptors and its inhibition has been effective in patients with rheumatoid arthritis, one of the most common irAEs. Our study points to the need of a systematic investigation of non-coding genome in discoveries of novel biomarkers in pharmacogenomics applications. The effects observed for the associated ieQTLs with ICI toxicity, further support the notion that the clinical relevance of non-coding common genetic variation in pharmacogenomics loci, likely due to the absence of selection pressure, will be substantially enhanced, as opposed to common variants associated with disease risk.

## PgmNr 826: Mutation burden and I index for detection of microsatellite instability in colorectal cancer by targeted next-generation sequencing.

### Authors:

S.M. Chun <sup>1</sup>; J.E. Kim <sup>2</sup>; Y.S. Hong <sup>2</sup>; K.P. Kim <sup>2</sup>; J. Kim <sup>1</sup>; C.O. Sung <sup>1</sup>; E.J. Cho <sup>3</sup>; T.W. Kim <sup>2</sup>; S.J. Jang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Pathology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea.; 2) Department of Oncology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea.; 3) Asan Medical Institute of Convergence Science and Technology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea.

---

ext-generation sequencing (NGS) panels are widely used for defining tumor mutation profiles and determining treatment approaches. We performed targeted NGS with 382 colorectal cancer genes with known microsatellite instability (MSI). After exclusion of germline alterations, the load of somatic mutations and small insertion/deletion (indel) alterations were determined. In the test set, 79 patients with 41 microsatellite-stable (MSS) and 38 MSI tumors were included. There were 120 MSS and eight MSI-high tumors in the validation set. The number of somatic mutations of whole samples were distinguished into three groups: mutant functional polymerase epsilon catalytic subunit, MSI, and MSS tumors. The median numbers of somatic and indel mutations in MSI tumors were higher. The indel mutation to whole mutation ratio (I index) was higher in MSI tumors. Hypermutation and low I index of polymerase epsilon catalytic subunit mutant tumors, a somatic mutation load cut-off of  $\geq 40$ , and an I index of  $\geq 9\%$  were selected as the criteria for detecting MSI tumors with high sensitivity and specificity. With the analysis of alteration patterns of homopolymer genes, a higher median number of homopolymer mutations in MSI tumors was observed. Mutated homopolymer  $\geq 5$  was selected as the criterion for detecting MSI tumors. MSI in colorectal cancer can be detected by targeted NGS panels with high sensitivity and specificity using somatic mutation load and I index. Additionally, further investigation was performed with same criteria for analysis of targeted NGS using cell-free DNA.

## PgmNr 827: Development of a large scale liquid biopsy platform for classification of microsatellite instability.

### Authors:

C. Cibulskis<sup>1</sup>; Y. Markuva<sup>1</sup>; J. Grimsby<sup>1</sup>; B. Blumenstiel<sup>1</sup>; M. Defelice<sup>1</sup>; J. Abreu<sup>1</sup>; M. Fleharty<sup>1</sup>; J. Tsui<sup>1</sup>; M. Cipicchio<sup>1</sup>; S. Hamilton<sup>1</sup>; S. Dookran<sup>1</sup>; G. Ha<sup>1</sup>; V. Adalsteinsson<sup>1,3</sup>; G. Getz<sup>1,2,4</sup>; N. Lennon<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Eli and Edythe L. Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, 02142 MA USA; 2) Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, 02215 MA USA; 3) Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, 500 Main Street, Cambridge, 02142 MA USA; 4) Massachusetts General Hospital, 55 Fruit Street, Boston, 02129 MA USA

---

Identifying patients whose cancer is driven by microsatellite instability and mismatch repair deficiency may lead to successful treatment with FDA approved immune checkpoint blockade therapy. We aim to identify such patients using a low cost, sensitive and highly accurate liquid biopsy whole genome assay.

To date we have generated >7000 cfDNA whole genomes at ultra low pass coverage (0.1x - 1x) from a wide range of cancer indications. Using the iChor analysis tool we are able to routinely generate tumor purity estimates and copy number profiles at large scale as well as provide research teams with important quality control metrics for selecting the most useful cases for additional deep sequencing assays (exomes or targeted panels). We find that only 17% of blood samples assayed by our lab contain > 10% ctDNA and the remaining will require ultra sensitive approaches to mutation profiling and MSI classification. Current platforms for determining MSI status in tumor samples (tissue staining or DNA sizing assays) are not transferable to circulating cell free DNA applications. Moreover, standard next generation sequencing protocols pose a challenge to classify MSI cases because PCR errors inherent in the library preparation and sequencing process can mimic MSI deletion events. A genome-wide approach to compute sample level MSI scores from >23 Million potential MSI loci may offer enough statistical power to accurately detect MSI samples at low tumor purity. To test this, we generated mixtures of MSI cell line DNA with normal DNA and performed whole genome sequencing to depths ranging from 0.1x - 0.5x coverage. Using this data and the novel algorithm MSIdetect developed in the Broad Cancer Program, we demonstrate accurate classification of MSI cases down to 0.5% tumor fraction. With further optimization of analysis algorithms and additional sequencing innovations it may be possible to push the limit of detection to 0.1% and lower. This platform and analysis pipeline together will form the foundation for a CLIA grade assay capable of delivering both tumor purity estimates and MSI classification on patient cfDNA samples.

## PgmNr 828: Integrated analysis of genomic data in ovarian cancer to facilitate precision medicine.

### Authors:

S.J. Cooper <sup>1</sup>; A. Montgomery <sup>2</sup>; E. Gordon <sup>1</sup>; A. Londono <sup>2</sup>; B. Johnston <sup>1</sup>; W. Goldsberry <sup>2</sup>; R. Arend <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) HudsonAlpha Institute, Huntsville, AL.; 2) University of Alabama - Birmingham, Birmingham, AL

---

Ovarian cancer is the deadliest of the gynecological malignancies due to the prevalence of late-stage diagnosis and chemoresistance. The development of precision medicine for ovarian cancer has the potential to positively impact patient survival. In order to advance the goal of precision medicine, we used transcriptomic profiling and metabolomic analysis to characterize heterogeneity among tumors pre- and post-treatment with standard of care platinum-based chemotherapy. We recapitulate subtypes of ovarian cancer described by previous groups using gene expression. In addition, we generated metabolomic profiles for ovarian tumors. Integration of gene expression and metabolomic data showed that TCA cycle metabolism is linked to increased cellular proliferation in tumor tissue. Further, we show that gene expression patterns associated with DNA damage and stress response are induced following platinum therapy. Induction of immune response is variable among patients following platinum treatment and proliferation is generally reduced after therapy. We also show that over half of patients change gene-expression subtype following platinum treatment. While current standard of care treats patients of each subtype similarly, looking toward the future, these data show that initial treatment and response may help direct subsequent selection of therapy. We also profiled patient ascites to determine the extent to which gene expression signatures useful for subdividing patients persist in metastatic cells. We show that in nearly half of our cases, the signature of the subtype remains detectable in the ascites. However there are some significant changes associated with ascites compared to tumor samples from the same patient. Namely there is up-regulation of genes associated with proliferation and cell migration as well as up-regulation of signaling pathways such as MAPK and TGF $\beta$ . These analyses suggest profiling of metastatic cells might also be informative in determining therapies that specifically target metastatic cells. Our data provide further insight into the heterogeneity that exists ovarian cancer tissues and analysis of matched post-treatment tissues and ascites cells gives a more complete view of transcriptomic and metabolomic profiling that can lead to delineation of the clinically relevant profiles to be used in the pursuit of precision medicine for ovarian cancer.

## PgmNr 829: Effect of *KRAS* mutations on extracellular vesicle content and function in human pancreatic ductal epithelial cells.

### Authors:

R.L. Dittmar<sup>1,2</sup>; M.C. Tai<sup>1</sup>; A. Maitra<sup>1,3,4</sup>; A.M. Killary<sup>1,2</sup>; S. Sen<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Genetics and Epigenetics Program, The University of Texas MD Anderson UTHealth Graduate School of Biomedical Sciences, Houston, TX; 3) Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX; 4) Sheikh Ahmed Bin Zayed Al Nahyan Center for Pancreatic Cancer Research, The University of Texas MD Anderson Cancer Center, Houston, TX

---

Activating mutations in *KRAS* occur in over 90% of pancreatic ductal adenocarcinomas (PDAC), the most deadly cancer by organ site with a five year survival of 9%. Mutations in this “undruggable” PDAC driver gene are believed to be the initiating genetic alteration in PDAC, potentially occurring years before malignancy. *KRAS* is mutated in bona fide PDAC precursor lesions, including mucinous cystic lesions, and promotes many hallmarks of cancer, including tumor microenvironment (TME) remodeling. While *KRAS* mutations are known to alter the TME, their effect on extracellular vesicles (EVs), small membrane-bound mediators of cell-to-cell communication in TME have not been well characterized. We hypothesize that *KRAS* mutations not only alter EV content, but also affect the TME, driving cystic lesions towards PDAC development. Given that EV cargo reflects the content of the cells releasing them, EVs may provide a “snapshot” of the biology of tumor initiating cells, potentially making them circulating biomarkers for early PDAC detection.

To address this hypothesis, we profiled microRNAs (miRNAs) in EVs from a human pancreatic ductal epithelial (HPDE) cell line with wildtype *KRAS* and a cell line derived from these same cells stably expressing *KRAS*<sup>G12V</sup> that were injected into mice and formed tumors (KRAST). After isolating EVs, we profiled their miRNA content using the HTG miRNA Whole Transcriptome Assay, which simultaneously profiles 2,083 miRNAs. Using this technique, we identified 18 miRNAs significantly differentially abundant in EVs from KRAST cells v. HPDE cells with FDR < 0.1. Additionally, we isolated EVs from 22 pancreatic cyst fluid samples and profiled miRNA content. A total of 330 miRNAs were significantly differentially abundant between pre-malignant cyst fluid EVs and benign cyst fluid EVs, while a total of 923 miRNAs were significantly differentially abundant between malignant cyst fluid EVs and benign cyst fluid EVs (FDR < 0.1). Several of these miRNAs were also significantly differentially abundant in EVs from pancreatic cyst fluid samples, whose *KRAS* mutation statuses were detected either by ddPCR or targeted sequencing. We are currently doing functional studies to identify the impact of these alterations on the pancreatic ductal epithelial cell microenvironment. Understanding the effects of *KRAS* mutations on EV content and function should lead to urgently needed sensitive and specific cyst fluid biomarkers for early detection of PDAC.

## PgmNr 830: Germline sequencing of advanced prostate cancer patients in the BARCODE2 trial.

### Authors:

R. Eeles <sup>1</sup>; S. Benafif <sup>1</sup>; R. Rageevakumar <sup>1</sup>; E. Saunders <sup>1</sup>; I. Whitmore <sup>1</sup>; C. Cieza-Borella <sup>1</sup>; E. McGrowder <sup>1</sup>; B. Jenkins <sup>1</sup>; S. Burnett <sup>1</sup>; A. Reid <sup>2</sup>; A. Falconer <sup>4</sup>; J. Walker <sup>1</sup>; G. Attard <sup>3</sup>; E. Bancroft <sup>1</sup>; A. Borges Da Silva <sup>1</sup>; C. Goh <sup>5</sup>; C. Peckitt <sup>2</sup>; S. Wallace <sup>6</sup>; E. Hall <sup>1</sup>; Z. Kote-Jarai <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) The Royal Marsden Hospital, Sutton, Surrey, UK; 3) UCL Cancer Institute, London, UK; 4) Imperial College Healthcare NHS Trust, London, UK; 5) Royal Surrey County Hospital, Guildford, UK; 6) University of Leicester, Leicester, UK

---

**Background:** The BARCODE2 trial (NCT02955082) is a European Research Council supported study investigating the response to platinum chemotherapy in men with advanced prostate cancer (PrCa) and a germline mutation in a DNA repair gene. We present the germline DNA sequencing results for 100 men who underwent genetic screening.

**Methods:** Eligible men with advanced PrCa consented to trial entry and provided a blood sample for DNA extraction and sequencing. All patients were known to have metastatic castration resistant PrCa (mCRPC) and had been treated with one or both of: Docetaxel chemotherapy and an androgen receptor targeted agent. Next generation sequencing (NGS) using a customised 115 gene panel (Agilent Technologies) was carried out on a MiSeq machine. Protein truncating variants (PTVs) in DNA repair genes by NGS were validated by Sanger sequencing. Patients found to carry a PTV were eligible to receive carboplatin upon disease progression.

**Results:** 100 patients have undergone genetic testing. The mean age of patients entering the trial was 67 (range 46-84). 22 patients (22%) carried one or more PTVs in a DNA repair gene that would qualify them for carboplatin treatment in the study: 4 carriers of a *BRCA2* variant and 3 carriers of a variant in *ALKBH3*, 2 carriers of an *ATM* variant and 2 carriers of a *PALB2* variant. One of the *ATM* variant carriers and one of the *BRCA2* variant carriers also carried a PTV in *POLQ* and 2 further carriers of the same *POLQ* variant were identified. One carrier was identified to carry a variant in each of the following genes: *MRE11A*, *PARP2*, *BLM*, *CHEK2*, *EXO1*, *LIG4*, *FANCD2*, *MSH5* and *PMS1*. Of the 21 identified PTVs, 11 are classified as pathogenic or likely pathogenic in ClinVar. There was no significant difference in the age at PrCa diagnosis between carriers and non-carriers (mean age 59 years vs 60; t-test p=0.29). There was also no difference in the presence of a family history of cancer in a first degree relative.

**Conclusion:** The initial NGS results in the BARCODE2 trial demonstrate a higher frequency of germline mutations in mCRPC patients compared with previously published data. This study is ongoing and aims to sequence 450 patients. If the frequency of germline mutations in men with advanced PrCa is confirmed to be 20% or higher, this would support the roll-out of routine genetic testing in the oncology setting to identify men who may benefit from targeted treatments such as platinum chemotherapy and PARP inhibitors.



## PgmNr 831: Genomic and immune landscape characteristics associate with improved overall survival of melanoma patients.

### Authors:

F. Ferguson<sup>1,2,3</sup>; L. Morales<sup>1,2,3</sup>; D. Simpson<sup>1,2,3</sup>; J. Cadley<sup>1,2,3</sup>; E. Esteva<sup>1,2,3</sup>; K. Monson<sup>1,2,3</sup>; V. Chat<sup>1,2,3</sup>; C. Martinez<sup>1,2,3</sup>; J. Weber<sup>1,3,4</sup>; I. Osman<sup>1,3,4,5</sup>; T. Kirchhoff<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Perlmutter Cancer Center, NYU Langone Health, New York, New York.; 2) Department of Population Health and Environmental Medicine, New York University School of Medicine, New York, New York; 3) The Interdisciplinary Melanoma Cooperative Group, New York University School of Medicine, New York, New York; 4) Department of Medicine, New York University School of Medicine, New York, New York; 5) Ronald O. Perelman, Department of Dermatology, New York University, New York, New York

---

Tumor-specific indicators, such as tumor mutation burden (TMB) have been shown to affect overall survival (OS) in melanoma. Recently, pan-cancer analyses from The Cancer Genome Atlas (TCGA) discovered specific tumor immune signatures predictive of overall survival (OS), yet it is unclear how these interact with other prognostic markers, independently of immunotherapy (IT). We aimed to combine the immune landscape signatures with TMB and other prognostic markers to improve melanoma OS prediction in patients, independent of IT. We examined the whole-exome data in conjunction with the molecular, clinical and immune features from 278 metastatic melanomas from TCGA, not treated by IT, to develop an improved prognostic model of melanoma OS. Using the discovery (N = 139) and validation (N = 139) design we performed multivariate Cox proportional hazards (Cox HR) models, adjusted for age and tumor stage at primary diagnosis, to identify interaction between TMB and melanoma immune features (n=59), refining the prediction of melanoma OS. We identified 4 immune features that were significantly associated with OS in both the discovery and validation cohorts. The multivariate Cox HR models revealed that IFN- $\gamma$  response (IFN- $\gamma$  R) and macrophage regulation (MR) signatures in combination with TMB were the most significantly associated with OS ( $p = 8.80E-14$ ). After further refinement, we observed that patients with high TMB, high IFN- $\gamma$  R and high MR had significantly better OS compared to high TMB, low IFN- $\gamma$  R and low MR (HR=2.8,  $p = 3.55E-08$ ). This association was not observed in low TMB patients. We show, for the first time, that genomic (TMB) and tumor immune features are significantly associated with improved OS, independent of IT. Further analysis of patients revealed that high TMB associates with improved OS in patients with high IFN- $\gamma$  R and MR but not in low IFN- $\gamma$  R and MR. Hence, this data provides first evidence that patients with high TMB have distinct OS outcome depending on other tumor immune features. Beside the biological link between TMB and IFN- $\gamma$  and MR, our data suggest that these associations may significantly improve the current melanoma prognostic models.



## PgmNr 832: Assessment of germline structural variants in pediatric cancer survivors for cancer predisposition.

### Authors:

D.A. Flasch <sup>1</sup>; E. Davis <sup>1</sup>; S.V. Rice <sup>1</sup>; H. Mulder <sup>1</sup>; D. Kennetz <sup>1</sup>; R. Mostafavi <sup>2</sup>; C.I. Wilson <sup>3</sup>; J. Easton <sup>1</sup>; K.E. Nichols <sup>2</sup>; Z. Whang <sup>3</sup>; Y. Yasui <sup>3</sup>; L.L. Robison <sup>3</sup>; M.M. Hudson <sup>2</sup>; J. Zhang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computational Biology, St. Jude Children's Research Hospital, Memphis, TN.; 2) Oncology, St. Jude Children's Research Hospital, Memphis, TN.; 3) Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN.

---

Few studies have comprehensively evaluated the prevalence and potential impact of germline structural variants (SVs) on cancer predisposition. Whole genome sequencing (WGS) studies of germline samples from cancer patients provides an opportunity to explore the currently underappreciated prevalence of germline SVs and their role in cancer predisposition.

Here we examine germline WGS data from 2,986 pediatric cancer survivors along with 344 healthy controls from the St. Jude Lifetime Cohort Study to identify germline SVs (e.g., deletions, inversions, insertions and translocations) in coding and noncoding potentially-regulatory regions of the genome associated with cancer predisposition. Secondary neoplasms were reported in 439 of these survivors allowing further assessment of genetic predisposition caused by germline SVs. We applied a copy number variant (CNV) detection algorithm CONCERTING in conjunction with three SV detection algorithms, CREST, Delly, and a novel algorithm SEASULT to identify germline SV candidates. SEASULT identifies SVs by *de novo* assembly of soft-clipped and discordant reads which improves both sensitivity and accuracy of SV analysis. Curated public polymorphic SV and CNV datasets, in addition to control WGS data, aided removal of nonpathogenic polymorphic SVs. Patient clinical and family history data and publicly available genomics data facilitated assignment of pathogenicity likelihood of the remaining SVs. Initial analysis revealed 1,017 deletion SVs in potentially-regulatory (74), exonic (230), or intronic (713) regions of a set of 1,028 cancer-related genes. Two survivors harbored deletion SVs, previously unreported, which included the 5'UTR and first exon of the gene related to their primary diagnoses (e.g., *RB1* in retinoblastoma and *WT1* in Wilms tumor). Two potentially-regulatory and one exonic *BRCA* deletion SVs were observed in survivors with a family history of breast cancer, suggesting increased likelihood of a secondary neoplasm. Other survivors displayed hemizygous loss of genes (e.g., *CDKN2AIP*, *CFHR1*, *MLH3*, *PTCH2*, and *SBDS*) that may be involved in cancer predisposition. Furthermore, preliminary analysis of germline insertion SVs revealed a 565bp insertion of *ROBO1* on chr3.p12.3 within the second exon of *NF1* on chr17q11.2 in a patient presenting with café-au-lait spots and a bilateral optic nerve glioma. These findings highlight the importance of examining germline SVs for comprehensive analysis of cancer predisposition.

## PgmNr 833: Identification of *CLDN6* as a novel tumor-promoting gene in gastric cancer.

### Authors:

I. Imoto <sup>1,2</sup>; T. Kohmoto <sup>1,3</sup>; K. Masuda <sup>4</sup>; K. Shoda <sup>5</sup>; E. Otsuji <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Molecular Genetics, Aichi Cancer Center Research Institute, Nagoya, Aichi, Japan; 2) Department of Cancer Genetics, Nagoya University Graduate School of Medicine, Nagoya, Japan; 3) Department of Human Genetics, Graduate School of Biomedical Sciences, Tokushima University, Tokushima, Japan; 4) Kawasaki Medical School, Kurashiki, Japan; 5) Division of Digestive Surgery, Department of Surgery, Kyoto Prefectural University of Medicine, Kyoto, Japan

---

In this study, we aimed to identify novel tumor-promoting drivers that would be highly expressed in gastric cancer (GC) and contribute to the worse prognosis in patients with GC. Through genome-wide screening for genes overexpressed in GC tumors compared to adjacent non-tumorous tissues and genes whose expression levels were positively correlated with worse overall survival using publicly available datasets, we identified *CLDN6* as one of the candidate genes promoting GC, particularly Lauren's intestinal subtype of GC. *CLDN6* immunoreactivity was detected in a part of Lauren's intestinal type of GC, and correlated with worse overall survival. Silencing endogenous *CLDN6* in *CLDN6*-expressing GC cell lines inhibited cell proliferation and cell migration/invasion abilities. Microarray-based comprehensive gene expression analysis revealed that inhibition of *CLDN6* induced suppression of the Hippo signaling pathway through suppressing transcription of the *YAP* gene. Our findings suggested that aberrant expression of *CLDN6* might play an important role in malignant progression of GC, particularly its intestinal type, through Hippo signaling pathway at least in part; furthermore, *CLDN6* might be a possible prognostic marker and therapeutic target in a subset of GC.

## PgmNr 834: A variational autoencoder model for accurate imputation of drug response.

### Authors:

P. Jia <sup>1</sup>; G. Pei <sup>1</sup>; R. Hu <sup>1</sup>; Z. Zhao <sup>1,2,3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA

---

Accurate imputation of drug response in large cohort is critical for the understanding of mechanisms underlying drug sensitivity and resistance. Most cancer pharmacogenomics data are obtained from preclinical studies, such as cell lines and mouse models and suffer from limitations such as small sample sizes and heterogeneity. In this study, we developed a deep variational autoencoder model followed by Elastic Net regression to predict drug response based on baseline gene expression profile. With systematic evaluation and computational benchmark, we demonstrated high accuracy rates of the prediction models that outperformed current in-silico methods [7/24 drugs with in-sample Pearson Correlation Coefficient (PCC) > 0.7 and 21/24 with PCC > 0.6] and then applied them in The Cancer Genome Atlas (TCGA) data for 33 cancer types. We found some drugs, especially those of targeted therapies, achieved the best performance when their prediction model were trained using only cell lines with solid tumor origin while some drugs, especially those with cytotoxicity, achieved the best performance when using all types of cell lines. With the TCGA multi-omics data, we found the response profile of two groups of compounds showed significantly different molecular associations, where samples sensitive to cytotoxicity compounds had an increased tumor mutation burden but such a pattern was not observed in targeted compounds. Both known mutation-drug associations (e.g., Lapatinib with HER2+, PF2341066 with MET, and genes with BRAF inhibitors) and novel associations were identified, which were further broken down to mutation clusters (hot spots). For example, FGFR3 was found sensitive to the MEK inhibitor AZD6244 with its two clusters: one on the 248-249<sup>th</sup> amino acids and the other on the 370:371:373<sup>th</sup> amino acids; and MDM4, instead of MDM2, was found most significantly associated with Nutlin-3, which is a MDM2 inhibitor. In addition, mutant tumor suppressor genes were depleted from sensitive samples to EGFR/BRAF/MEK inhibitors. A systematic screen of gene expression signatures revealed biomarkers for each drug, including significant associations between increased Tumor Inflammation Score (TIS) with RAF265 (a BRAF inhibitor), Paclitaxel, and Nutlin-3, and a significant association between decreased expression of stroma genes and Lapatinib (HER2 inhibitor) in breast cancer. Collectively, these results provided valuable insights to understand the mechanisms underlying drug response.

## PgmNr 835: Assessing a causal relationship between circulating lipids and breast cancer risk via Mendelian randomization.

### Authors:

K.E. Johnson <sup>1</sup>; K.M. Siewert <sup>2</sup>; D. Klarin <sup>3,4,5</sup>; S.M. Damrauer <sup>6,7</sup>; K.M. Chang <sup>6,8</sup>; P.S. Tsao <sup>9,10</sup>; T.L. Assimes <sup>9,10</sup>; K.N. Maxwell <sup>8,11</sup>; B.F. Voight <sup>6,11,12,13</sup>; VA Million Veterans Program

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 2) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 3) Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Boston VA Healthcare System, Boston, MA, USA; 6) Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, USA; 7) Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 8) Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 9) Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; 10) VA Palo Alto Health Care System, Palo Alto, CA, USA; 11) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 12) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 13) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

### Importance

A number of epidemiological and genetic studies have attempted to determine whether levels of circulating lipids are associated with risks of various cancers, including breast cancer. If alteration of lipid levels reduced risk of breast cancer, this could present a target for disease prevention.

### Methods

We assessed a potential causal relationship between genetic variants associated with plasma lipid traits and risk for breast cancer using Mendelian randomization (MR). Data from genome-wide association studies in 210,967 subjects from the Million Veterans Project were used to construct genetic instruments for plasma lipid traits. The effect of these instruments on breast cancer risk was evaluated using genetic data from the BCAC consortium based on 137,045 breast cancer cases and 119,078 controls. We employed single-trait MR for each lipid trait (HDL, LDL, and TG), and multivariable MR accounting for correlation between lipid traits and body mass index. We also used independent associations at core HDL or LDL genes for locus-specific MR. We tested for genome-wide genetic correlation between each lipid trait and breast cancer risk using LD-score regression, and local genetic correlation using p-Hess.

## Results

We observed that a 1-SD genetically determined increase in HDL or LDL levels is associated with an increased risk for all breast cancers (HDL: OR=1.09, 95% CI=1.03-1.15,  $P=1.7\times 10^{-3}$ ; LDL: OR=1.06, 95% CI=1.02-1.12,  $P=4.7\times 10^{-3}$ ). Multivariable MR analysis, which adjusted for the effects of LDL, TG, and body mass index, further improved the strength of this observation for HDL (OR=1.07, 95% CI=1.03-1.10,  $P=8.6\times 10^{-5}$ ). We did not observe a difference in this relationship when stratified by breast tumor estrogen receptor status. In our locus-specific MR of independent associations at core HDL genes, we found locus-specific associations with breast cancer at *ABCA1*, *APOE-APOC1-APOC4-APOC2* and *CETP*; and in meta-analysis across loci (OR=1.11, 95% CI=1.06-1.16,  $P=1.5\times 10^{-6}$ ). Finally, we find evidence for local genetic correlation between lipid levels and breast cancer at the *ABO* locus.

## Conclusions

Genetically elevated plasma HDL levels appear to increase breast cancer risk. Future studies are required to understand the mechanism underlying this putative causal relationship, with the goal to develop potential therapeutic strategies aimed at altering the HDL-mediated effect on breast cancer risk.

## PgmNr 836: Neoantigen discovery workflow: Accurate prediction of immuno-oncology targets.

### Authors:

V. Kovačević; V. Josipović; S. Stojanović; N. Skundrić; J. DiGiovanna; B. Davis-Dusenbery

[View Session](#) [Add to Schedule](#)

**Affiliation:** Seven Bridges, New Belgrade, Serbia, Serbia

---

Neoantigens are peptides on the surface of cancer cells that may activate the immune system. Multiple novel therapeutic approaches involve identifying neoantigens and predicting which neoantigens can be leveraged to trigger immunity-induced tumor regression.

We present a novel method (workflow) for neoantigen discovery and prioritization using Next Generation Sequencing (NGS) data which process tumor-normal pairs of whole exome sequencing (WES) samples and tumor gene expression data. Our method considers multiple dimensions of genomic features including phasing of mutations, the influence of both somatic and germline mutations, and positions of all transcripts and mutations on start and stop codons. This Neoantigen workflow (also known as a pipeline) is a combination of tools (e.g. BWA, GATK4) and is described in Common Workflow Language (CWL). Here, individual tools are associated with Docker containers enabling both configurability of all relevant parameters and full reproducibility across an abundance of computational environments.

This neoantigen discovery workflow has been tested throughout the last two rounds (Round 2 and Round X) of the Tumor Neoantigen Selection Alliance (TESLA) Challenge organized by the Parker Institute for Cancer Immunotherapy. Within Round X, TESLA provided flow-cytometry-validated peptide-HLA sets for more than 10 patients. Our neoantigen discovery workflow detected the majority of confirmed candidates. Detailed results will soon be published by the TESLA organizers.

Accurate, reproducible, and rapid discovery of neoantigen candidates represents a foundation for personalized cancer vaccine development. Paired with a cloud environment, this Neoantigen workflow has the potential to fulfill the demands for rapid turn-around of accurate predictions within a clinically compliant environment for a large number of patients.

## PgmNr 837: Spatial characterization of tumor immune microenvironment by multiplex IF in breast ductal carcinoma in situ.

### Authors:

T. Kumar <sup>1</sup>; F. Yang <sup>2</sup>; M. Almekinders <sup>3</sup>; E. Cuentas <sup>2</sup>; J. Love <sup>4</sup>; E. Lips <sup>3</sup>; I. Wistuba <sup>2</sup>; N. Navin <sup>1</sup>; J. Wesseling <sup>3</sup>; S. Krishnamurthy <sup>5</sup>; A. Thompson <sup>6</sup>; A. Futreal <sup>7</sup>; On behalf of the PRECISION team

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, UT MD Anderson Cancer Center, Houston, Texas.; 2) Translational Molecular Pathology, UT MD Anderson Cancer Center, Houston, Texas.; 3) Netherlands Cancer Institute, Amsterdam, Netherlands; 4) Breast Surgical Oncology, UT MD Anderson Cancer Center, Houston, Texas.; 5) Department of Pathology, UT MD Anderson Cancer Center, Houston, Texas.; 6) Department of Breast Surgery, Baylor College of Medicine, Houston, Texas; 7) Department of Genomic Medicine, UT MD Anderson Cancer Center, Houston, Texas.

---

**Background:** Ductal carcinoma in situ (DCIS) is an abnormal proliferation of epithelial cells within the ducts of the breast, accounts of around 15% of all newly diagnosed breast cancer cases, with an estimated 60,000 patients being diagnosed in the US each year primarily due to the adoption of mammographic screening. However, the risk of invasive breast cancer or death in a patient diagnosed with pure DCIS is less than 1%. Therefore, there is significant overtreatment, with DCIS being characterized as a form of cancer. In this study, we evaluated the tumor microenvironment to quantify the composition and spatial distributions of immune phenotypes in a longitudinal cohort of primary DCIS and matched recurrent DCIS or invasive ductal carcinoma (IDC) tissues.

**Methods:** We performed multiplex immunofluorescence (mIF) using three customized immune marker panels including Checkpoint protein panel (PD-1, PD-L1, CD3, CD8, CD68 and pancytokeratin AE1/AE3), T cell panel (CD3, CD8, Granzyme B, CD45RO, FOXP3, AE1/AE3) and Myeloid panel (Arginase 1, CD68, CD11b, CD33, CD14, CD66b, AE1/AE3). Fourteen primary DCIS with seven matched DCIS (7) or IDC (7) recurrences were examined with Vectra automated quantitative pathology imaging system and inForm software (PerkinElmer).

**Results:** The density of macrophages was higher in primary DCIS samples with invasive recurrence as compared to the ones with DCIS recurrence as well as the ratio of cytotoxic T to regulatory T cells, whereas the density of monocytes was higher in primary DCIS with DCIS recurrence. There was no significant difference in other immune cell types in the primary DCIS time point. All patients had a significant decrease in macrophages from primary DCIS to recurrent DCIS, whereas an increase was observed when they recurred as IDC. Spatial analysis showed that there was a higher probability of occurrence of cytotoxic T and memory T cells around tumor cells in primary DCIS samples that had an invasive recurrence. Higher Cytotoxic T cell density was significantly associated with ER negativity in invasive recurrent samples.

**Conclusions:** Our findings shed light on the composition and distribution of immune cells as a putative predictor of DCIS to IDC progression. A subset of myeloid cells correlated with invasive

recurrence suggesting deeper investigation, including further spatial analyses, of the local immune microenvironment in larger, clinically informative cohorts is warranted.



## PgmNr 838: Longitudinal assessment of the 313-SNP based polygenic risk score for breast cancer risk prediction in a Dutch prospective cohort.

### Authors:

I.M.M. Lakeman <sup>1</sup>; J. van Rooij <sup>2</sup>; R. Ruiter <sup>3</sup>; C.J. van Asperen <sup>4</sup>; A.G. Uitterlinden <sup>2,3</sup>; P. Devilee <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Leiden University Medical Centre, Leiden, Netherlands; 2) Internal Medicine, Erasmus Medical Center Rotterdam, the Netherlands; 3) Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 4) Clinical Genetics, Leiden University Medical Center, the Netherlands; 5) Pathology, Leiden University Medical Center, the Netherlands

### Purpose

Breast cancer associated SNPs, summarized in a Polygenic Risk Score (PRS) are presently not used to guide the appropriate starting age for breast cancer screening. At this moment the most predictive PRS for breast cancer is based on 313 SNPs. The aim of this study is to validate the performance of the 313-SNP PRS in a Dutch prospective cohort and explore the utility of risk stratified screening.

### Materials and Methods

We included women from the Rotterdam study, a Dutch prospective cohort study started in 1989 with >70% response rate. Phenotypic information was available with 25 years of follow-up. Genotyping was performed with the Illumina 550k and 610k array for 6670 women aged 45 years or older. After imputation with the HRC1.1 and 1KG reference panels, the standardized 313-SNP-based PRS (sPRS), per 1 standard deviation (SD) in controls from the Breast Cancer Association Consortium (BCAC) was calculated. Cox regression analysis, adjusted for birth year and age at inclusion, was performed to estimate the association between the sPRS and incident breast cancer risk.

### Results

Of all included women with an average age of 66, 145 had prevalent breast cancer and 323 developed incident breast cancer during follow-up. These women had a higher sPRS, mean 0.45 (SD=1.06) for cases with in situ carcinoma (N=37) and 0.51 (SD=1.06) for cases with invasive carcinoma (N=431) compared to 0 (SD=1.01) in all 6202 controls, and 0.05 (SD=0.98) in 842 of these controls with other tumors than breast cancer. Breast cancer risk increased by sPRS with an Hazard Ratio (HR) per SD=1.56, 95%CI [1.40-1.73], and compared with the middle sPRS quintile, the lowest quintile had an HR of 0.57, 95%CI[0.38-0.87] and the highest quintile an HR of 1.81, 95%CI[1.32-2.50]. We noticed a slightly smaller effect of the sPRS for in situ compared to invasive breast cancer risk only (HR=1.45, 95%CI[1.02-2.01], and 1.58, 95%CI[1.41-1.76], respectively).

### Discussion and conclusion

The 313-SNP PRS can be used for breast cancer risk stratification in the Dutch population and seems specific for breast tumors and not for other tumor types. Further analyses are being performed that include other breast cancer risk factors (i.e., reproductive and lifestyle factors), which are available for the majority of the included women. Furthermore, the added value of the 313-SNP PRS on long- and short-term breast cancer risk prediction and screening stratification will be explored.

## PgmNr 839: Genomic elucidation on the roles of HPK1 in inflammation and immune microenvironment during pancreatic cancer development.

### Authors:

T. Liu <sup>1</sup>; R. Abbasgholizadeh <sup>2</sup>; W. Chen <sup>1</sup>; H. Wang <sup>2</sup>; H.M. Wang <sup>3</sup>; C. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Basic Sciences, School of Medicine, Loma Linda University, Loma Linda, California.; 2) Gastrointestinal Medical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas.; 3) Department of Pathology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas.

---

Pancreatic ductal adenocarcinoma (PDAC) is one of the most common pancreatic malignancies and it is the third most lethal cancer among all cancers. It was believed that genetic factors (e.g., Kras mutation) combined with nongenetic events (e.g., inflammation and pancreatitis) were responsible for PDAC development. Heterogeneity and immunosuppressive nature of tumor microenvironment are among the leading contributors to the lack of effective treatment for PDAC. Recent studies demonstrated that a loss function of hematopoietic progenitor kinase 1 (HPK1) is correlated with PDAC progression and tumor immunosuppressive status. To elucidate the role of HPK1 in inflammation and immune microenvironment during pancreatic cancer development, we carried out comprehensive genomic and epigenomic studies using: **(1) mouse pancreatic cancer cell lines bearing the mutations of K-RAS only, K-RAS/p53 or K-RAS/p53 plus HPK1 non-functional mutation, respectively; (2) transgenic (TG) mice bearing HPK1 overexpression in three different genetic constructs identical to the above cell lines; (3) a human PDAC tumor tissue.** Our RNA-seq results from mouse cancer cell lines identified 1,459 significant differentially expressed genes (DEGs) in response to recovered active HPK1 level. KEGG pathway analysis on the DEGs indicated an enrichment in cell cycle, cell division and DNA repair. TG mouse inflammation models were utilized to study the role of HPK1 in different stages of chronic pancreatitis induced pancreatic cancer development. RNA-seq results based on TG mouse models identified 392 DEGs in response to HPK1 overexpression in the early stage, and a KEGG pathway analysis indicated that these DEGs were enriched in fatty acid metabolism, p53 pathway and IL-17 pathway. We are currently performing 10X Genomics scRNA-seq, scATAC-seq and scVDJ-seq on human PDAC sample in order to characterize the patient pancreatic tumor cell population heterogeneity and the immunosuppressive microenvironment, as well as a comprehensive bioinformatics data analysis integrating the genomic data derived from all cell lines, TG model and patient sample. The outcome of this study will allow us to gain a deep insight into the roles of HPK1 in pancreatic cancer development. It may eventually lead to identification of novel diagnostic and prognostic biomarkers, development of new therapeutic targets and better treatment strategy for pancreatic cancer.

## PgmNr 840: Prevalence of germline mutations in expanded cancer predisposing genes in patients with sporadic renal cell cancer.

### Authors:

R. Metpally <sup>1</sup>; S... Krishnamurthy <sup>1</sup>; D.J. Carey <sup>1</sup>; H. Williams <sup>2</sup>; Regeneron Genetics Center

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Weis Center for Research, Geisinger Clinic, Danville, Pennsylvania, USA; 2) Department of Urology, Geisinger Clinic, Danville, PA, USA

---

**Purpose:** Hereditary renal cancer accounts for up to 5% of all renal cell carcinoma (RCC) cases. Germline testing is recommended in patient with early onset (age  $\leq 46$  y) and family history of the disease. Recent data from Nordic twin studies and Iceland reported that familial contribution is between 38 to 58% suggesting that many more RCC predisposing genes are yet to be discovered. To address this, we evaluated the prevalence of germline mutations in 450 plausible cancer predisposing genes (CPGs) in RCC. **Methods:** We reviewed the records of all patients (PT) with sporadic RCC (no family history, n=511) within our institutional MyCode-DiscovEHR cohort (n=92,455). Corresponding Whole Exome Sequencing data was analyzed, and variant pathogenicity accessed employing ACMG guidelines for all 450 CPGs. **Results:** In our cohort, we had 387 patients with clear cell (ccRCC), 74 papillary (pRCC), 27 chromophobe (chRCC), 11 mixed and 12 other RCCs. The mean age of onset of RCC was 61.65 years (SD, 11.24). One hundred and forty-seven (**29%**) of 511 patients with sporadic RCC cancer had either predicted loss-of-function variants (pLOF) or ClinVar pathogenic or likely pathogenic (CP/LP;  $\geq 1$  gold star review status) germline mutation. Germline alterations were found in 99 of 450 cancer predisposing genes tested. Germline mutations accounted for 27.4%, 35.1%, 25.9%, 27.3%, 41.7% in ccRCC, pRCC, chRCC, mixed, and other RCC respectively. The top altered genes include *CHEK2* (5.4%), *MUTYH* (5.4%), *SERPINA1* (4%), *ATM* (3.4%), *PDE11A* (3.4%), *PYGM* (3.4%), *TYR* (2.7%), *TYRP1* (2.7%), *BRCA2* (2%) and *COL4A3* (2%). 28 out of 147 patients (19%) had germline alteration in more than one CPG. 5.2% of patients also had homozygous or compound heterozygous mutations (Cys282Tyr, His63Asp, Ser65Cys) in homeostatic iron regulator (*HFE*) gene associated with hemochromatosis known to increase cancer risk. We observed 6% of patients with copy number variation in these CPGs. We found no difference in the age of onset among germline positive vs negative sporadic RCC patients. **Conclusion:** In our cohort, germline mutations were identified in 99 CPGs, well beyond the 18 genes currently used for RCC germline testing. Overall 37% (29% pLOF + CP/LP; 5.2% *HFE*; 6% CNVs) of sporadic RCCs patients have germline variations. Pending functional and clinical validation, these genes may represent a more updated landscape of RCC predisposing genes

## PgmNr 841: From biobanking to precision medicine.

### Authors:

A. Metspalu<sup>1</sup>; M. Alver<sup>1</sup>; K. Lall<sup>1,2</sup>; K. Krebs<sup>1</sup>; L. Leitsalu<sup>1</sup>; N. Tonisson<sup>1</sup>; T. Esko<sup>1</sup>; K. Fischer<sup>1,2</sup>; L. Milani<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Estonian Biobank, Univ Tartu, Tartu, Estonia; 2) 1. 2. Institute of Mathematics other and Statistics, University of Tartu, Estonia

---

The Estonian Biobank was founded in 2000 as a volunteer-based biobank, and now, 19 years later, it contains a collection of health and genetics data of close to 200 000 individuals, approximately 20% of the adult population of the country. The Human Genes Research Act (passed in 2000) allows regular updating of data through linking to national registries, enabling long-term follow-up of the cohort and re-contacting biobank participants. A nationwide technical infrastructure (X-road) for the secure electronic exchange of medical data has also been established and is maintained by the state. This allows creating disease trajectories on all participants from birth, based on information stored in the medical systems, including ICD-10 diagnoses, prescriptions, lab tests and electronic medical records. To date, 152 000 individuals have been genotyped with Illumina's Global Screening Array, and the whole genomes of 3,000 individuals and whole exomes of 2500 individuals have been sequenced. This serves as a population-based imputation reference. Using these data, I will report 3 pilot projects where personalized medicine is implemented in Estonia: familiar hypercholesterolemia (FH), breast cancer and drug response. Using a **"genetics first approach"**, we discovered many new FH patients in the biobank based on LoF variants only not detected by the medical system and for over 50% of the cases the treatment was changed after cardiology checkup. We are conducting several pilot projects in order to work out the best ways to return health related research data back to the individuals at the biobank. For that purpose, we have developed decision support tools for several major diseases, including risk estimates based on **polygenic risks scores (PRS)** for breast cancer and pharmacogenomics based recommendations based on CPIC. Today, over 2,000 participants of the Estonian Biobank have received genetics-based counselling and the feedback has been very positive based on questionnaires collected before, right after and 6 months post-counselling. Validation studies using PRS for coronary heart disease and breast cancer are ongoing in the two largest hospitals in Estonia. As all these data will be transferred to the medical system within the next few years, personalized medicine as 4P medicine (personalized, predictive, preventive and participatory) has reached the point of no return in Estonia.

# PgmNr 842: Enrichment of rare germline variants in DNA repair and cell cycle genes in chronic lymphocytic leukemia.

## Authors:

N.S. Moore<sup>1,2,3</sup>; S.H. AlDubayan<sup>1,2,3,8</sup>; E. Kofman<sup>2,3</sup>; S. Han<sup>1,2,3</sup>; A. Taylor-Weiner<sup>1,2,3</sup>; S. Stilgenbauer<sup>4</sup>; G. Getz<sup>3,5,6,7</sup>; C.J. Wu<sup>1,2,3</sup>; E.M. Van Allen<sup>1,2,3</sup>; J.R. Brown<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Harvard Medical School, Boston, MA; 2) Dana-Farber Cancer Institute, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Internal Medicine III, University of Ulm, Ulm, Germany; 5) Department of Pathology, Massachusetts General Hospital, Boston, MA; 6) Massachusetts General Hospital Cancer Center, Boston, MA; 7) Department of Pathology, Harvard Medical School, Boston, MA; 8) Brigham and Women's Hospital, Boston, MA

## Background

Chronic lymphocytic leukemia (CLL) is among the most heritable cancers, with 60% of disease risk genetically determined. However, 75% of the genetic heritability of CLL remains unexplained by GWAS and linkage studies. Previously, we identified ATM as the first CLL risk gene. Here we leverage a deep-learning-based germline variant calling algorithm to explore germline mutational enrichment in DNA repair and cell cycle genes in CLL.

## Methods

A case-control analysis was conducted using gene-based mutational enrichment analysis of 50 established cancer predisposition DNA repair and cell cycle genes (DRG). Germline variants in cases and controls were called and filtered using identical methods. In the discovery phase, a total of 569 non-Finnish European CLL cases and 8,608 ancestry-matched, unselected controls were compared for enrichment in pathogenic variants. An FDR correction was applied and genes with a q-value < 0.1 were considered significant. Variants were also called in an independent cohort of 314 Spanish CLL patients for which ancestry-matched controls are in process.

## Results

Our analysis identified CHEK2 pathogenic variants as significantly enriched in CLL cases compared to controls. Enrichment was observed in both CHEK2 LOF alterations (cases: 2.6%, controls: 0.7%) and the separately assessed low-penetrance (LP) allele c.599T>C (cases: 2.1%, controls 0.3%). Enrichment remained significant after removal of the known founder variant 1100delC. Overall, CLL patients were 2.1-15.6 times more likely to carry CHEK2 pathogenic variants compared to controls (LOF: OR = 3.8, 95%CI = 2.1-6.8, q-value<0.01; LP: OR = 7.7, 95%CI = 3.6-15.6, q-value < 0.001). In addition, 4 CHEK2 LOF variants were observed in the Spanish cohort (1.3%), although no LP variants.

## Conclusion

Our analysis of a large CLL discovery cohort using a high-sensitivity variant calling algorithm supports CHEK2 as a potentially novel CLL predisposition gene that may explain a portion of the missing monogenic heritability of CLL. In addition, this study highlights the DNA repair and cell cycle regulation pathways as potential drivers of CLL susceptibility.

## PgmNr 843: Identifying differential gene induction responsible for varying cisplatin response by bridging non-cancer and cancer cell models.

### Authors:

L. Mortensen; A.L. Stark

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of Notre Dame, Notre Dame, IN.

---

Cisplatin is a widely available chemotherapeutic drug, shown to be successful in combatting various types of cancer through the formation of platinum adducts, contorting DNA. However, chemotherapeutic resistance is a major complication in platinum-based chemotherapies and causes many patients to not be effectively treated. It remains unclear why certain patients are intrinsically resistant and non-responsive to the same cisplatin treatment that other patients are sensitive to. We propose that differences in gene regulation cause the observed differential response to cisplatin treatment. This project compares protein expression patterns observed in cisplatin resistant and sensitive HapMap lymphoblastoid cell lines (LCLs) to the expression signatures generated in multiple cancer models. Beginning with the LCL protein expression at 2, 6, and 12 hours post 5 $\mu$ M cisplatin treatment, genes of interest were selected by analyzing the sign of the slope observed in expression profiles of 482 transcription factors and cell signaling molecules from 6-12 hours after treatment. The 95 genes identified with differential expression slopes were then classified based on their functional role in the cell, where eight proteins involved in cisplatin's canonical mechanisms of action (apoptosis and DNA damage) were selected for study in cancer models. Expression for selected genes (*APTX*, *ATF2*, *BCL2*, *CORO1A*, *HDAC2*, *NF- $\kappa$ B1*, *SMC1A*, and *STAT3*) was quantified at 6, 12, 24, and 48 hours post treatment with 5 $\mu$ M cisplatin in cancer cell models. Comparing the expression induction patterns observed in LCLs to the cancer cell models will yield both global and cancer type-specific molecular insights into contributors to cisplatin resistance mechanisms. Utilizing the existing LCL data as an intrinsic genetic marker for cisplatin sensitivity or resistance, we have been able to classify selected genes as contributing to cisplatin resistance or sensitivity mechanisms. This project's identification of resistance-contributing factors among different cancer models will enhance understanding of the molecular and genomic complexities of cisplatin resistance. This allows for the unique translational potential to modulate those candidate genes contributing to resistance in order to sensitize cancer cells to cisplatin. Because cisplatin resistance is a major problem, being able to sensitize resistant patients would be enormously impactful in bettering patient outcomes and the lives of those affected.



## PgmNr 844: Aggressive prostate cancer: Towards population-based precision prevention.

### Authors:

T. Nguyen-Dumont<sup>1,2</sup>; R.J. MacInnis<sup>3,4</sup>; J.A. Steen<sup>1</sup>; D. Theys<sup>1</sup>; H. Tsimiklis<sup>1,2,5</sup>; F. Hammet<sup>1,2,5</sup>; M. Mahmoodi<sup>1,2,5</sup>; B.J. Pope<sup>1,5,6,7</sup>; D.J. Park<sup>2,7</sup>; K. Mahmood<sup>6,7</sup>; G. Severi<sup>8,9</sup>; J. Pedersen<sup>10</sup>; D. Bolton<sup>11</sup>; R.L. Milne<sup>1,3,4</sup>; G.G. Giles<sup>3,4</sup>; M.C. Southey<sup>1,2,3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton, Victoria, Australia; 2) Genetic Epidemiology Laboratory, Department of Clinical Pathology, The University of Melbourne, Victoria, 3010, Australia; 3) Cancer Epidemiology and Intelligence Division, Cancer Council Victoria, Victoria, 3004, Australia; 4) Centre for Epidemiology and Biostatistics, The University of Melbourne, Victoria, 3010, Australia; 5) The University of Melbourne Centre for Cancer Research, Victoria Comprehensive Cancer Centre, Victoria, 3010, Australia; 6) Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Victoria, 3010, Australia; 7) Melbourne Bioinformatics, The University of Melbourne, Victoria, 3010, Australia; 8) CESP Inserm U1018, Faculté de médecine - Univ. Paris-Sud, Faculté de médecine - UVSQ, Université Paris-Saclay, 94805, Villejuif, France; 9) Gustave Roussy, F-94805, Villejuif, France; 10) TissuPath, Mount Waverley, Victoria, 3149, Australia; 11) Department of Surgery, The University of Melbourne, Austin Health, Victoria, 3084, Australia

The identification of heritable factors that predict clinical outcome of prostate cancer (PrCa) can both improve early diagnosis of potentially aggressive and lethal cases, and reduce overtreatment of indolent disease. However, to date, few heritable genetic risk factors for PrCa have been demonstrated to be specifically associated with aggressive disease.

Here, we report a case-case study that compared the prevalence of germline pathogenic variants in 787 men with aggressive PrCa and 769 men with non-aggressive PrCa. Since most of the genes currently identified as PrCa susceptibility genes are also involved in, or are putatively associated with, predisposition to breast and/or ovarian cancer and Lynch syndrome, we designed a panel targeting *ATM*, *BARD1*, *BRCA1*, *BRCA2*, *BRIP1*, *CDH1*, *CHEK2*, *FANCM*, *HOXB13*, *MLH1*, *MRE11A*, *MSH2*, *MSH6*, *MUTYH*, *NBN*, *NF1*, *PALB2*, *PMS2*, *PTEN*, *RAD50*, *RAD51C*, *RAD51D*, *RECQL*, *RNASEL*, *STK11* and *TP53*. Our analysis focused on rare genetic variants predicted to lead to loss of function and/or classified as pathogenic/likely pathogenic (P/LP) in ClinVar.

We observed 94 P/LP variants in 88/787 (11.2%) men with aggressive PrCa and 80 P/LP variants in 73/769 (9.5%) men with non-aggressive PrCa (Fisher's exact test  $P=0.28$ ). The prevalence of *BRCA2* P/LP variants in men with aggressive PrCa (20/787, 2.5%) exceeded that observed in men with non-aggressive PrCa (4/769, 0.5%) ( $P=0.001$ ). One man in each group carried two P/LP *BRCA2* variants. Other genes in which the prevalence of P/LP variants was higher in men with aggressive disease included *ATM* (11 (1.4%) men with aggressive PrCa, and five (0.7%) men with non-aggressive PrCa, respectively,  $P=0.21$ ), *CHEK2* (10 (1.3%) and five (0.7%) men, respectively,  $P=0.30$ ), and *BRCA1* (five (0.6%) and two (0.3%) men, respectively,  $P=0.45$ ).

Our results support previous reports that rare P/LP germline variants are enriched in men with aggressive PrCa compared with men with non-aggressive disease. Further and extensive internationally coordinated studies are required to confidently classify rare pathogenic variants

identified in panel-testing. Evidence is accumulating in relation to genetic information about *ATM* in the clinical management of men with prostate cancer but more data is required to interpret variation in the many other genes currently included on commercial tests for which there is currently an insufficient evidence-base for clinical translation.



## PgmNr 845: Targeting DDI2 to potentiate proteasome inhibitor-induced cell death in triple negative breast cancer cells.

### Authors:

A. Northrop <sup>1,2</sup>; S. Radhakrishnan <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human and Molecular Genetics, VCU, Richmond, VA; 2) Department of Pathology, VCU, Richmond, VA

---

Despite widespread proteasome addiction in cancer cells, proteasome inhibitors (PIs) have failed to effectively treat most types of cancer. One explanation for this failure is utilization of the NRF1-mediated proteasome bounce-back response, a compensatory cellular mechanism that occurs during proteasome inhibition to evade PI-induced cell death. The response requires the activation and nuclear translocation of a transcription factor, NRF1 (alias for NFE2L1), which is proteolytically-cleaved into its active form by a protease DDI2. During proteasome inhibition, NRF1 promotes the transcriptional synthesis of new, uninhibited proteasomes via the activation of proteasome-related genes to relieve PI-induced cell stress and prevent apoptosis. The goal of this study was to evaluate DDI2 as a therapeutic target to inhibit the NRF1-mediated proteasome bounce-back response to potentiate PI-induced cell death of a triple negative breast cancer (TNBC) cell line *in vitro*. Pulse-chase experiments using a p97 inhibitor, cycloheximide, and an FDA-approved PI, Carfilzomib (CFZ), were followed by subcellular fractionation and western blot to compare NRF1 retrotranslocation from the ER membrane to the nucleus and processing between wild-type and *DDI2* knock-out NIH 3T3 fibroblasts. Cell viability and caspase-3 assays were used to compare the sensitivity of DDI2-deficient and DDI2-competent TNBC cells to CFZ. CFZ treatment of TNBC cells co-expressing wild-type and protease-dead DDI2 (p.D252A) was followed by western blot and cell death assays to analyze NRF1 processing and CFZ sensitivity, respectively. *DDI2* knock-out in NIH 3T3 fibroblasts impairs NRF1 processing to the active form, but not NRF1 translocation to the nucleus. DDI2-deficient TNBC cells are more sensitive to CFZ-induced cell death. Protease-dead DDI2 behaves in a dominant-negative fashion in the TNBC cells to impair NRF1 processing and potentiate CFZ-induced cell death. Our results show that depletion of DDI2 or loss of its protease activity increases the sensitivity of TNBC cells to PI-induced cell death *in vitro*. Previous attempts at using PIs alone or in combination with other cancer therapeutics have failed to effectively treat most types of cancer. A combinational approach using a DDI2 protease inhibitor with CFZ or other FDA-approved PIs may expand the repertoire of cancer types in which PIs can effectively kill cancer cells to improve patient outcomes.

## PgmNr 846: Prostate cancer risks for male *BRCA1* and *BRCA2* mutation carriers: Results from the prospective EMBRACE cohort study.

### Authors:

T. Nyberg <sup>1</sup>; D. Frost <sup>1</sup>; D. Barrowdale <sup>1</sup>; R. Eeles <sup>2,3</sup>; D.G. Evans <sup>4</sup>; M. Tischkowitz <sup>5,6</sup>; D.F. Easton <sup>1</sup>; A.C. Antoniou <sup>1</sup>; EMBRACE collaborators

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 2) Oncogenetics Team, Division of Genetics and Epidemiology, The Institute of Cancer Research, London, United Kingdom; 3) Cancer Genetics Unit, Royal Marsden NHS Foundation Trust, London, United Kingdom; 4) Manchester Regional Genetics Service, Central Manchester University Hospitals NHS Foundation Trust, Manchester, United Kingdom; 5) Department of Medical Genetics, University of Cambridge, Cambridge, United Kingdom; 6) East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Trust, Cambridge, United Kingdom

---

**Background:** *BRCA1* and *BRCA2* mutations have been associated with prostate cancer (PCa) risk but a wide range of risk estimates has been reported, based mostly on retrospective studies. Previous absolute risk estimates from retrospective studies may not be widely applicable today due to the rapid rise in population PCa incidence with the increased use of the PSA test.

**Aim:** To provide prospective estimates of the relative and absolute risks of PCa associated with *BRCA1/2* mutations.

**Material and Methods:** EMBRACE is a prospective cohort study of *BRCA1* and *BRCA2* mutation carriers identified through clinical genetics centres in the UK and Republic of Ireland since 1998. We followed the unaffected male *BRCA1* (N=376) and *BRCA2* (N=447) mutation carriers to estimate age-specific incidences, standardised incidence ratios (SIRs) relative to population incidences, and absolute risks of developing PCa. We also assessed risk modification by family history and mutation location, and the risks of high- and low-grade PCa.

**Results:** Sixteen *BRCA1* and 26 *BRCA2* mutation carriers were diagnosed with PCa during a median follow-up of 5.6 years. *BRCA2* mutation carriers had a SIR of 4.45 (95% confidence interval [CI] 2.99-6.61), and absolute risk of PCa of 60% (95% CI 43%-78%) by age 85. For *BRCA1* mutation carriers, the overall SIR was 2.35 (95% CI 1.43-3.88), with higher SIR at ages <65 (SIR=3.57, 95% CI 1.68-7.58), and absolute risk of 29% (95% CI 17%-45%) by age 85. However, the *BRCA1* SIR was not consistently statistically significant in sensitivity analyses that assessed potential screening effects. The PCa risks for *BRCA2* mutation carriers increased with the number of first or second degree relatives diagnosed with PCa (hazard ratio [HR]=1.68, 95% CI 0.99-2.85). Mutations in the ovarian cancer cluster region of *BRCA2* (c.2831-c.6401) showed weaker association with PCa risk (HR=0.37, 95% CI 0.14-0.96) compared to mutations outside this region. For *BRCA2* mutation carriers, there was

a stronger association with Gleason score  $\geq 7$  (SIR=5.07, 95% CI 3.20-8.02) than Gleason score  $\leq 6$  PCa (SIR=3.03, 95% CI 1.24-7.44).

**Conclusion:** The results confirm the high risk of aggressive PCa for *BRCA2* mutation carriers, and give some support for a weaker association in *BRCA1* mutation carriers, particularly at younger ages. The absolute risk estimates will be informative for the development of screening programs to facilitate early detection of PCa for male *BRCA1* and *BRCA2* mutation carriers.

## PgmNr 847: Identification of drug-interacted non-coding somatic mutations by pan-cancer analysis.

### Authors:

Y. Rong; Y. Guo; S. Dong; H. Chen; T. Yang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, P. R. China

---

It's a valuable topic and challenge to access a detailed treatment guideline for precision medicine, a detailed biomarker panel to seek optimal treatment for individuals. Great efforts from genomics to proteomics have developed lots of biomarkers, but there are still few projects talking about the non-coding somatic mutations, the important part deciding tumor heterogeneity, may cause drug resistance or sensitivity in indirect ways. Here we introduce a strategy to identify drug-interacted non-coding somatic mutations with clinical information and genomics data supplied by TCGA. For each mutation-drug pairs, patients are divided into four groups by they used drug or not and with or without this mutation, then log-rank test are calculated in pairs. we mainly focus on these mutation-drug pairs like: mutation has no effect in non-therapy groups but is significant prognostic in therapied groups while durg is significant effective in patients without mutation but no effect in patients with mutation groups. We calculate trans-eQTL effect of these mutations with corresponding drugs' related genes supplied by drugbank database. These drug-interacted mutations are much more likely to cause abnormal expression of drug related genes than those negative results. These mutations mainly cause drugs' enzymes and transporter overexpression and lead to drug resistance instead of drug's target genes. Cellline data from GDSC also proved the expression of enzymes and transporter genes are more likely to be correlated with IC50 levels than drug's targets. PPI network analysis results suggests the more distant interaction genes number may cause these drugs' enzymes and transporter genes are more sensitive to the global change in tumor cells. Doxorubicin's related enzyme *NDUFS7* suffer the most number of mutations affected. Most of these mutations which cause *NDUFS7* overexpression will lead to drug resistance. But only half of these mutations which cause *NDUFS7* low expression will lead to drug sensitivity. It's supposed to be some compensation mechanism exists. Meanwhile, we found the expression balance between *CYP2D6* and *CYP2C9* will influence the drug sensitivity of Tamoxifen while *CYP2C9* is just the one target of Anastrozole, thus the combination of these two drug is proved to have the greast effect than single one. Our work is the first to explore the relations between non-coding somatic mutation and cancer drug. Our findings may provide some guidance to clinical medication.

## **PgmNr 848: *Ex vivo* drug screening and genomic profiling to guide individualized treatments for children with relapsed or refractory solid tumors and leukemias.**

### **Authors:**

D. Salyakina <sup>1</sup>; I. Sotto <sup>1</sup>; V. Berrios <sup>2</sup>; C. Saghira <sup>3</sup>; M. Fader <sup>1</sup>; M. Janvier <sup>1</sup>; Z. Khatib <sup>1</sup>; H. Abdela <sup>1</sup>; D.J. Azzam <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Research Institute, Nicklaus Children's Hospital, Miami, Florida.; 2) Department of Environmental Health Sciences, Robert Stempel College of Public Health & Social Work, Florida International University, Miami, Florida; 3) Miller School of Medicine, University of Miami, Miami, Florida

---

### **Background**

The implementation of precision medicine in pediatric cancers faces unique challenges. Pediatric cancers are fundamentally different from adults with lower frequencies of genetic mutations and fewer options for targeted therapies. The ultimate goal of this study was to evaluate feasibility of personalized treatment options based on *ex vivo* drug screening and genetic testing for pediatric patients with recurrent and/or refractory cancers.

### **Methods**

In this pilot study, four patients with recurrent cancers have been enrolled to date. The functional drug screening test (DST) panel encompassed 40 formulary drugs frequently used at Nicklaus Children's hospital and 47 non-formulary drugs approved by FDA for cancer treatment as well as drugs from phase III and IV clinical trials. Drug sensitivity score (DSS) was calculated for each drug based on cancer cells' response. DST results were then combined with results from the genetic screen in an attempt to match actionable mutations with selective targeted therapies.

### **Results and discussion**

Fresh tumor samples from four patients with recurrent acute myeloid leukemia, recurrent osteosarcoma and recurrent sclerosing spindle cell rhabdomyosarcoma were obtained and tested. Genetic testing alone did not provide enough information to guide treatment because of the lack of drug resistance or sensitivity information in publicly available databases and scientific literature. In addition, turnaround time on genetic testing did not allow results to be used in a timely manner. In contrast, *ex vivo* DST returned between 10 to 30 treatment options for each patient. Two patients treated on protocols guided by DST responded or currently show stable disease. Remarkably, DST provided valuable information to the oncologists on drug dosing and treatments that may not be effective, allowing them to avoid these selections. Despite the limitations of *ex vivo* DST and challenges in implementation, it remains a very promising strategy for patients with refractory and recurrent cancers for which physicians are often guessing what the most appropriate treatment should be.

## PgmNr 849: Longitudinal tracing of TNBC tumors reveals distinct patterns of response: Molecular profiling results from randomized Phase II clinical trial (ARTEMIS; NCT02276443).

### Authors:

S. Seth <sup>1</sup>; L. Guo <sup>1</sup>; G. Rauch <sup>1</sup>; B. Adrada <sup>1</sup>; H. Piwnica-Worms <sup>1</sup>; A. Thompson <sup>2</sup>; E. Mittendorf <sup>3</sup>; J. Litton <sup>1</sup>; W. Symmans <sup>1</sup>; G. Draetta <sup>1</sup>; A. Futreal <sup>1</sup>; S. Moulder <sup>1</sup>; J. Chang <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The University of Texas MD Anderson Cancer Center Houston, TX; 2) Baylor College of Medicine Houston, TX; 3) Dana-Farber/Brigham and Women's Cancer Center Boston, MA; 4) The University of Texas Health Science Center, Houston, TX

**Background:** The heterogeneity of TNBC results in varied responses to NAST: 30-40% of patients (pts) have pathologic complete response (pCR) with excellent prognosis. Those with residual disease, have a much higher risk of recurrence. Longitudinal profiling assesses biologic response to NAST and mechanisms of resistance. **Methods:** Pts with stage I-III TNBC began a planned 4 cycles of Adriamycin-based chemo (AC). Biopsies were performed pre (mandatory) and post (optional) AC. Volumetric change by ultrasound (VUS) at completion of AC (or progression) was calculated. Pts with sensitive disease received subsequent taxane-based (T) therapy. Pts with insensitive disease were offered phase II trials. Pathologic response was assessed at surgical resection in 47 pts. Matched samples, pre and post AC (N = 48 pts) underwent transcriptomic and genomic profiling. Samples were classified into six previously identified ARTEMIS subtypes of TNBC (ART-Type). Immune deconvolution and estimation was performed using RNA-Seq profiles. Differential pathway-level analysis was performed comparing pre and post AC samples. **Results:** There was heterogeneity in response to AC with 4 predominate patterns of biologic response (Table). In 48% of cases the ART-Type of the tumor switched after AC, with androgen receptor like (LAR) and immune modulatory (IM) showing greatest stability. Tumors with enrichment in EMT or those with no significant dysregulation after AC (Groups C + D) were associated with less immune modulation and lower rates of pCR compared to those with depleted EMT (A and B) (8.7% vs 45.8%,  $p = 0.0078$ ). **Conclusions:** Molecular profiling of longitudinal TNBC samples reveals distinct response patterns in tumors and their micro-environments upon treatment with AC. These patterns were indicative of pathologic response in this cohort; however, they require validation in a separate cohort.

Biologic Response Class (n = 48)	Enriched Hallmark Pathways	Depleted Hallmark Pathways	Pathologic Response pCR/total (n = 47)	% pCR (p-value)
A	Heme metabolism Basophil/Neutrophil lineage (immune deconv.)	EMT, MYC, G2M checkpoint	7/18	38.9% (0.19)
B		EMT	4/6	66.7% (0.04)

C	No significant change	No significant change	2/17	11.8% (0.09)
D	EMT, Inflammatory response, IL6_JAK_STAT3, Allograft rejection, Angiogenesis, Coagulation	MYC, G2M checkpoint, E2F	0/6	0.0% (0.16)
total			13/47	27.7%

## PgmNr 850: The ORIEN Avatar Program: Building a massive oncology health data and genomics resource through collaborative partnership.

### Authors:

S. Szelinger <sup>1</sup>; M. Poi <sup>2</sup>; O. Hampton <sup>2</sup>; T.K. McDaniel <sup>1</sup>; J.M. Trent <sup>1</sup>; W. Dalton <sup>2</sup>; The ORIEN Alliance

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Translational Genomics Research Institute, Phoenix, Arizona.; 2) M2Gen, Tampa, Florida

---

Genomic oncology has a growing but largely unrealized capacity to transform research, drug and diagnostics development, and patient care. To fully realize this potential there is need to develop a data resource containing tumor and germline genomic information, medical histories, and outcomes for large numbers of past and present cancer patients. Such resource would foster basic research, facilitate drug trials for agents targeting rare markers, and usher in the era of artificial-intelligence-enabled predictive medicine. Building such a resource at scale has been hindered by a variety of operational, institutional, and contractual barriers, which we have overcome through ORIEN, the Oncology Research Information Exchange Network. ORIEN is an alliance of 19 U.S.-based cancer centers based on the principles of Collaboration, Inclusiveness, Data accessibility, and Partnerships. Under a universal protocol, patients are consented for a longitudinal, prospective study, providing clinical information and tissue specimen, consenting to data sharing, and enabling proactive matching to clinical trials based on molecular profile during their lifetime in the ORIEN Avatar program. To date, over 215,000 patients have been consented. Molecular profiling funded by a precompetitive consortium of pharmaceutical companies is performed using next generation sequencing to generate deep exome data of each patient's tumor and matched constitutional DNA as well as whole transcriptome tumor RNA data. Molecular data, as well as health and demographic information, are collected and stored uniformly across the network. As of 2019, over 9,600 patients have been profiled by NGS. M2Gen, the coordinating center integrates clinical, molecular and patient reported data. All data storage and management follow IRB, CAP, CLIA, and HIPAA regulatory guidelines. This data warehouse is proactively queried as emerging therapies or clinical trials become available, and to provide evidence-based decision support at point-of-care. De-identified data is made available for ORIEN members and for participating pharmaceutical companies for collaborative research studies. In addition, ORIEN Avatar can create *in silico* patient communities to identify patients with similar molecular and clinical profile for study enrollment and to permit recontact for potential trial enrollment. Thus, the ORIEN Alliance constitute a vast infrastructure and resource for genomic oncology, to the benefit of all stakeholders.



## PgmNr 851: Pan-cancer pharmacogenetics: Targeted sequencing panels or exome sequencing?

### Authors:

L. Tilleman<sup>1</sup>; B. Heindryckx<sup>2</sup>; D. Deforce<sup>1</sup>; F. Van Nieuwerburgh<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Pharmaceutical Biotechnology, Ghent University, Ghent, Belgium; 2) Department for Reproductive Medicine, Ghent University Hospital, Ghent, Belgium

---

Understanding the genetic background of cancer is essential to enhance personalized cancer therapy. Various knowledgebases are available, describing the potential therapeutic consequences of genomic mutations. Pharmacogenetic information in these knowledgebases shows limited overlap and thus seems complementary to each other. In this study, five of these knowledgebases: The Cancer Genome Interpreter, Database of Evidence for Precision Oncology, the Jackson Laboratory Clinical Knowledgebase, the Precision Oncology Knowledgebase, and the Clinical Interpretations of Variants in Cancer, were aggregated in one meta-knowledgebase containing 8326 pharmacogenetic interactions with 3132 genomic mutations. Approximately 86% of the variants are described in just one of the five knowledgebases. Only 355 of these mutations were also discovered as driver mutations in the Pan-Cancer Atlas. Nine contemporary pan-cancer targeted sequencing panels: The Ion AmpliSeq Comprehensive Cancer Panel, the xGen Pan-Cancer Panel v2.4, the TrueSight Oncology 500 panel, the QIAseq Targeted Human Comprehensive Cancer Panel, the AVENIO ctDNA Targeted Kit, the AVENIO ctDNA Expanded Kit, the AVENIO ctDNA Surveillance Kit, the FoundationOne CDx panel and FoundationOne Liquid panel, were compared with the meta-knowledgebase and the genomic mutations in the Pan-Cancer Atlas. The xGen Exome Research Panel, representative for exome sequencing, was also included in the analysis. Overall, the xGen Exome Research Panel outperforms cancer-specific targeted sequencing panels in terms of analyzed genes, clinically relevant cancer mutations, cancer fusion genes, and pharmacogenetic interactions. From the cancer-specific targeted sequencing panels, the xGen Pan-Cancer Panel v2.4, the TrueSight Oncology 500 panel, the Ion AmpliSeq Comprehensive Cancer Panel, the QIAseq Targeted Human Comprehensive Cancer Panel, and the FoundationOne CDx panel cover about the same number of clinical cancer variants and determine a consensus set of pharmacogenetic interactions. Unfortunately, these panels determine virtually none of the known cancer gene fusions. The other panels, optimized for liquid biopsies, target only a subset of pharmacogenetic interactions. These results provide insight into the relevance of the panels considering current pharmacogenetic knowledge on cancer.

## PgmNr 852: Differential responsiveness to BRAF inhibitors in melanoma.

### Authors:

S. Tomei <sup>1</sup>; M. Al Hashmi <sup>1</sup>; S. K. Sastry <sup>1</sup>; V. Mattei <sup>1</sup>; R. Mathew <sup>1</sup>; E. Wang <sup>1</sup>; F. M. Marincola <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sidra Medicine, Doha, Qatar; 2) Refuge Biotechnologies, Menlo Park, California, United States

---

Cutaneous melanoma is an aggressive disease representing one of the leading causes of mortality related to human cancers worldwide. In the recent years, major advances have been made in the clinical management of melanoma patients carrying BRAF mutations with the adoption of two BRAF specific inhibitors. The correct identification of cancer driving mutations is of paramount importance in cancer diagnostics as this allows the appropriate selection of targeted treatments and implementation of personalized therapies. To date, several methods are used in diagnostics to identify clinically relevant mutations. These methods include Sanger sequencing, immunohistochemistry, mutation-specific real-time polymerase chain reaction (q-PCR) and next generation sequencing (NGS) technologies. Of note, BRAF mutation is usually detected at the DNA level, while BRAF expression is generally not assessed prior to target treatment selection. When performing RNA-seq and Sanger sequencing on 15 melanoma cell lines, we found discordance between the results obtained by the two methods for three of the cell lines analyzed, specifically, these three cell lines showed a BRAF V600E mutation by Sanger sequencing but the mutation was not detected by RNA-seq. This prompted us to investigate whether this was due to a differential expression of BRAF WT and V600E alleles and whether these cell lines showed a different responsiveness to BRAF inhibitors as compared to the concordant WT and V600E mutant cell lines. Here we aimed at assessing whether a lower overall BRAF mRNA expression may correlate with a lower responsiveness to BRAF inhibitors. To our knowledge this question has not been previously investigated.

We found that three cell lines whose mutation assignment was discordant between the DNA and RNA level and that express very low level of BRAF mRNA respond less to BRAF V600E specific inhibitors. These finding should be validated in additional studies with cohort of a bigger size and employing tumor tissues rather than cell lines.

## PgmNr 853: Clinical relevance of *TP53* hotspot mutations in high-grade serous ovarian cancers.

### Authors:

M. Tuna <sup>1,2</sup>; Z. Ju <sup>3</sup>; K. Yoshihara <sup>4</sup>; C.I. Amos <sup>2</sup>; J.L. Tanyi <sup>5</sup>; G.B. Mills <sup>6,7,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Department of Medicine, Baylor College of Medicine, Houston, TX; 3) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; 4) Department of Obstetrics and Gynecology, Niigata University, Japan; 5) Department of Obstetrics and Gynecology, University of Pennsylvania, PA; 6) Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; 7) Precision Oncology, Knight Cancer Institute, Oregon Health & Science University, Portland, OR; 8) Department of Cell, Developmental & Cancer Biology, Oregon Health & Science University, Portland, OR

---

**Background:** Mutation of *TP53* is the most frequent genetic alteration in high-grade serous ovarian cancers (HGSOC). The impact of hot spot mutations of *TP53* and protein levels on patient outcomes in HGSOC has not been fully elucidated. In this study we aimed to assess clinical relevance of hot spot mutations of *TP53*. **Methods:** The study population (n= 791) comprised of HGSOC samples with *TP53* mutation from The Cancer Genome Atlas and other publicly available data. RPPA data was retrieved from MD Anderson TCPA portal, and clinical data from XENA. Univariate cox proportional hazards regression analysis was used to select variables that were correlated with survival time, and t-test was used to select differentially expressed proteins between selected groups. **Results:** We assessed the effects of different types of *TP53* mutations based on location, oncogenic function, type of mutation (missense, nonsense, frame shift, splice site, and indel) and individual hotspot mutations on patient outcomes in HGSOC. Only hotspot mutations were associated with patient characteristics and outcomes. Three hotspot mutations; G266, Y163C and R282 both alone and in aggregate were associated with a worsened overall (OS) and recurrence-free survival (RFS) compared with other hotspot mutations (HR = 0.21;  $p < 0.0001$  and HR = 0.41;  $p = 0.001$ ), other non-hotspot missense mutations (HR = 0.57;  $p < 0.0001$  and HR = 0.70;  $p = 0.008$ ), and all other mutations (HR = 0.53;  $p < 0.0001$  and HR = 0.67;  $p = 0.001$ ). Specific hotspot mutations were associated with different protein expression patterns consistent with different functions. **Conclusions:** This study provides evidence that individual *TP53* hotspot mutations have different impact on HGSOC patient outcomes and potentially *TP53* function. Thus assessment of the status of particular *TP53* aberrations could influence response to therapy, and selection of therapeutic agents.

## PgmNr 854: Functional multi-omic profiling identifies Aurora kinase inhibition as a therapeutic strategy in *RIT1*-mutant lung adenocarcinoma.

### Authors:

A. Vichas<sup>1</sup>; N.T. Nkinsi<sup>1</sup>; S. Moorthi<sup>1</sup>; F. Piccioni<sup>4</sup>; F. Mundt<sup>5</sup>; P. Mertins<sup>6</sup>; A.H. Berger<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Biology Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Broad Institute of Harvard and MIT, Cambridge, MA; 5) Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 6) Proteomics Platform, Berlin Institute of Health and Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany

---

Treatment of lung cancer is evolving from standard cytotoxic to personalized treatment based on the molecular alterations unique to each patient's tumor. In our recent work with TCGA and others, we identified somatic mutations in the gene encoding the RAS-related small GTPase, *RIT1*, in lung adenocarcinomas. In addition, somatic *RIT1* mutations have also been identified in myeloid malignancies and germline *RIT1* mutations are found in the developmental "RAS-opathy," Noonan syndrome. In all these diseases, *RIT1* mutations are mutually-exclusive with other RAS/MAPK pathway mutations, including mutations in *EGFR*, *KRAS*, and *ALK*. Our previous analysis demonstrated that gain-of-function *RIT1* mutations induce potent transformation of rodent fibroblasts, and preliminary data indicates that the mechanism of *RIT1* signal transduction is likely to be distinct from that of other RAS GTPases, such as *KRAS*. To genetically dissect the signaling pathways downstream of *RIT1*, *KRAS*, and *EGFR*, we performed genome-wide CRISPR/Cas9 screens in three isogenic lung adenocarcinoma cell lines in which cell survival is dependent on the expression of the oncogene. Through integrative analysis of essential genes across the three cell lines, we identified synthetic lethal relationships unique and shared in each isogenic background. Among the dependencies identified was the requirement of aurora kinase pathway genes for survival of *RIT1*-mutant cells, raising the possibility of therapeutic aurora kinase inhibition in *RIT1*-positive tumors. To evaluate whether the dependence on Aurora kinases extended to small molecule inhibition, we performed a screen of 160 small molecules in clinical development or approved by the FDA. Consistent with our CRISPR/Cas9 results, we find that *RIT1*-mutant cells depend on Aurora kinase activity for survival. Interestingly, global proteomic and phosphoproteomic profiling identified *RIT1* regulation of mitotic spindle assembly genes. Taken together these results identify a new arm of *RIT1* signaling and nominate Aurora kinase inhibition as a potential therapeutic strategy in *RIT1*-mutated tumors.

## PgmNr 855: Cancer care and a tale of three molecular 'genomic' tests.

### Authors:

J. Wallace; B. Zimmermann; N. Liang; R. Swenerton; U. Do; S. Sawyer; J. Saucier; S. Parmar; A. Aleshin; N. Hook

[View Session](#) [Add to Schedule](#)

**Affiliation:** Natera, Inc., San Carlos, California.

---

**Introduction:** The practice of oncology has largely evolved to treating cancer, based on tumor genomics. Genetic testing for hereditary cancer predisposition syndromes and somatic tumor testing for targeted therapy selection have advanced patient care. More recently, cell-free DNA analysis has been leveraged to develop patient-specific, bespoke assays for the detection of circulating tumor DNA (ctDNA) in cancer patients, to allow monitoring for molecular residual disease, recurrence, and treatment response. Here we present a patient diagnosed with cancer where three individual molecular tests are utilized in the care continuum.

**Case Presentation:** A 56-year old Caucasian female diagnosed with stage IIIC ovarian clear cell carcinoma, treated with total abdominal hysterectomy with bilateral salpingo-oophorectomy, omentectomy, appendectomy, lymph node sampling, peritonectomy, and optimal tumor debulking in June 2018. At that time, the patient was assessed with tumor molecular comprehensive genomic profiling and hereditary cancer predisposition syndrome diagnostic testing to identify potential inherited predisposition risk and inform treatment selection, available clinical trials, and potential use of PARP inhibitor therapy. Treatment with first-line adjuvant chemotherapy concluded in Oct 2018. In Jan 2019, the patient relapsed and enrolled in clinical trial GYO16, receiving pembrolizumab infusions with oral epacadostat. In Feb 2019, the patient was assessed for ctDNA using a bespoke assay designed from matched-normal whole exome sequencing. ctDNA was detected at 4.61 MTM/mL. Positive ctDNA results allow the opportunity to monitor treatment response and trending disease burden through longitudinal screening of ctDNA levels.

**Discussion:** Here we present an oncology patient where molecular testing was utilized in three unique ways to inform and monitor cancer treatment. It is important that genetic professionals become familiar with the increasing number of molecular genomic tests being incorporated into oncology, each with a unique purpose in the patient continuum of care. Patients and physicians may seek consultation from clinical genetic professionals to differentiate between molecular genomic tests incorporated into practice.

## PgmNr 856: Genome sequencing analysis identifies high-risk Epstein-Barr virus subtypes for nasopharyngeal carcinoma.

### Authors:

M. Xu <sup>1</sup>; H. Chen <sup>2</sup>; S. Zhang <sup>1</sup>; X. Zhou <sup>1</sup>; W. Zhai <sup>3</sup>; J. Liu <sup>2</sup>; Y.X. Zeng <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, China; 2) Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; 3) Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, China

---

Epstein-Barr virus (EBV) infection is ubiquitous worldwide and is associated with multiple cancers, including nasopharyngeal carcinoma (NPC). The importance of EBV viral genomic variation in NPC development and its striking epidemic in southern China has been poorly explored. Through a large-scale genome sequencing of 270 EBV isolates and association study (discovery phase: 156 patients with NPC and 47 healthy controls; validation phase: 483 patients and 605 controls) of EBV isolates from China, we identified two non-synonymous EBV variants within *BALF2* strongly associated with the risk of NPC (odds ratio (OR) = 8.69,  $P = 9.69 \times 10^{-25}$  for SNP162476\_C and OR = 6.14,  $P = 2.40 \times 10^{-32}$  for SNP163364\_T). The cumulative effects of these variants contributed to 83% of the overall risk of NPC in southern China. Phylogenetic analysis of the risk variants revealed a unique origin in Asia, followed by clonal expansion in NPC-endemic southern China. We further explored a) the performance of NPC-associated EBV risk variants combined with the human susceptibility variants in risk prediction for NPC; b) the role of these risk variants in other EBV-associated malignancies; c) the role of the risk variants in EBV virus replication and immunogenicity. Our results provide novel insights into NPC tumorigenesis and its endemic in southern China. They also pave the way for identifying high-risk individuals and implementing effective intervention programs to reduce the disease burden in southern China.

## PgmNr 857: Copper (II) increases anti-proliferative activity of thymoquinone in colon cancer cells by increasing genotoxic, apoptotic and reactive oxygen species generating effects.

### Authors:

V.B. Yenigun; A. Kocyigit; E. Durmus; A.Z. Gul

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medical Biochemistry, Bezmialem Vakif University, Istanbul, Turkey

---

Colon cancer is one of the most common types of cancer in the world. It is the third most common cancer diagnosed in both men and women in the United States according to American Cancer society. Thymoquinone is the main active phenolic compound derived from the essential oil of the *Nigella sativa* plant seed, also known as black seed. While thymoquinone is an antioxidant, it has been reported in several studies that thymoquinone has dose-dependent pro-oxidant activity with Fenton reaction in the presence of transition elements such as iron and copper. Several studies showed that cellular copper is increased in different cancers and accumulated in the nucleus of the cells. The aim of this study was to investigate genotoxic, apoptotic, cytotoxic and reactive oxygen species (ROS) generating effects of thymoquinone together with copper in colon cancer cells. In the study, the pro-oxidant-acting doses of thymoquinone (10-60  $\mu\text{M}$ ) were administered to colon cancer cells (HT-29) alone and together with non-toxic dose (150  $\mu\text{M}$ ) of Copper (II) Sulfate ( $\text{CuSO}_4$ ) for 24 hours and the cytotoxic, apoptotic, genotoxic and ROS production activities were analyzed by MTT viability test, Acridin Orange / Etidium Bromide (AO/EB) double staining, alkaline single cell gel electrophoresis (comet assay) and  $\text{H}_2\text{DCF-DA}$  assay, respectively. MTT viability test showed that thymoquinone and copper have synergistic effect on colon cancer cells by decreasing  $\text{IC}_{50}$  value of thymoquinone from 50  $\mu\text{M}$  to ? 25  $\mu\text{M}$ . AO/EB double staining results were also supporting MTT results. Moreover, it was shown by comet assay that DNA damage was increased with synergic effect of thymoquinone and copper, and the tail intensity was increased from 25% to 67% with 40  $\mu\text{M}$  thymoquinone. Finally, the intracellular ROS was increased when thymoquinone and copper applied together comparing to thymoquinone alone. Our studies show that the application of redox active copper (II) with thymoquinone increases DNA damage, apoptosis and cell death by increasing the amount of intracellular ROS through pro-oxidant activity. Treatments targeting copper related pathway may open new therapeutic avenues in cancer treatment.



## PgmNr 858: Genomic alterations predict clinical response to systemic chemotherapy and immune checkpoint blockade in biliary tract cancer.

### Authors:

J. Yoon <sup>1</sup>; M.H. Kim <sup>2,3</sup>; M. Jang <sup>4</sup>; H. Kim <sup>4</sup>; Y.N. Park <sup>4</sup>; M.G. Lee <sup>1</sup>; H.K. Hwang <sup>5</sup>; C.M. Kang <sup>3,5</sup>; W.J. Lee <sup>5</sup>; B. Kang <sup>2,3</sup>; H.C. Chung <sup>2</sup>; H.J. Choi <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Research Center for Human Genetics, Department of Pharmacology, Yonsei University College of Medicine, Seoul, Republic of Korea; 2) Division of Oncology, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea; 3) Pancreaticobiliary Cancer Clinic, Yonsei Cancer Center, Severance Hospital, Seoul, Republic of Korea; 4) Department of Pathology, Yonsei University College of Medicine, Seoul, Republic of Korea; 5) Division of Hepatobiliary and Pancreatic Surgery, Department of Surgery, Yonsei University College of Medicine, Seoul, Republic of Korea

**Background:** Previous studies have identified several targetable oncogenic mutations in biliary tract cancer (BTC). However, reliable predictors of clinical response to therapeutic agents have not been established. This study aimed to identify predictors of clinical response to chemotherapy and immune checkpoint blockade in advanced BTC patients.

**Methods:** Genomic alterations in 54 patients with recurrent or metastatic BTC were investigated by targeted deep-sequencing for 524 key cancer genes using pre-treatment tumor tissues. The predictive value of genomic alterations, including copy-number alteration and mutational signatures, on response to chemotherapy and anti-PD-1/PD-L1 therapy, was assessed.

**Results:** Frequently altered genes included *TP53* (50.0%), *NOTCH1* (31.5%), *KRAS* (24.1%), *KMT2C* (22.2%), *BAP1* (22.2%), and *ERBB2* (22.2%) exhibiting highly heterogeneous distribution. Alterations in *BAP1*, *PBRM1*, *NKX2-1*, *CREBBP*, and *RAD51* were enriched in metastatic cases, while *FLT4* and *FGFR3* amplification were enriched in recurrent cases. Among 52 patients with first-line chemotherapy, *BAP1* deletion, *NOTCH1* alteration, and the homologous recombination deficiency (HRD) signature were associated with an early tumor shrinkage, while no benefit in progression-free survival (PFS) was observed. *FLT4* amplification was a favorable prognostic factor, while *ATM*, *CHEK2*, and *NF1* mutation were poor prognostic factors of PFS on chemotherapy. Furthermore, 21 patients received anti-PD-1/PD-L1 therapy after failure of chemotherapy, and the mismatch repair (MMR) signatures were significantly higher in the responding patients. By multivariate Cox analysis, old age, metastasis, and alterations in *ERBB2*, *CHEK2*, *MDM2*, and *NF1* gene were poor prognostic factors of overall survival.

**Conclusions:** This study identified molecular predictors associated with clinical response and outcome of therapeutic agents in advanced BTC. Particularly, mutational signatures associated with HRD and MMR correlated with the response to chemotherapy and anti-PD-1/PD-L1 therapy, respectively. These results provide general framework to select patients benefitting from therapeutic modalities based on molecular profile in advanced BTC.



## PgmNr 859: *HERC1* knockout cell model study provides insights on growth defect phenotypes.

### Authors:

J. Zhao; M. Kane; M. Provenzano; X. Liu; T. Vilboux

[View Session](#) [Add to Schedule](#)

**Affiliation:** Inova Health System, Falls Church, Virginia.

---

The E3 ubiquitin-protein ligase encoded by *HERC1* (HECT And RLD Domain Containing E3 Ubiquitin Protein Ligase Family Member 1) plays a crucial role in regulating cell proliferation, energy sensing and cell membrane trafficking. Loss of function *HERC1* mutations have been associated to a variety of rare neurodevelopmental disorders (macrocephaly, dysmorphic facies, and psychomotor retardation) characterized by large head and overgrowth. In addition, altered *HERC1* expression level has been identified in a variety of cancer types. Increasing number of researches indicate that higher *HERC1* expression facilitate cancer patients' survival. Despite of significant role of *HERC1* in human disease, the underlying mechanism is still not clear.

To better understand the impact of *HERC1* mutations in growth defects and cancer, we studied the molecular and cellular phenotype of a *HERC1* knockout cell model. Our cell model showed that the loss of *HERC1* induces a significant upregulation of the mTOR signaling pathway and an increase in cell proliferation. Interestingly, the knockout cells were able to maintain these phenotypes even in unfavorable growth conditions by recycling cellular contents to maintain cell homeostasis and viability during nutrient starvation in a more efficient manner, i.e. increase of autophagy. These results provide us potential insights on the impact of *HERC1* in normal and abnormal cell proliferation. This will open the door of potential therapeutic measures for patients with impaired growth profile and *HERC1* mutated cancer types using *HERC1* as a novel biomarker and as a target for therapy.

## **PgmNr 860: Gene signature-based prediction of patient response to neoadjuvant chemotherapy with applications to triple-negative breast cancer.**

### **Authors:**

Y. Zhao <sup>1</sup>; E. Schaafsma <sup>1</sup>; C. Cheng <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Molecular and systems biology, Dartmouth College, West Lebanon, NH.; 2) Department of Medicine, Baylor College of Medicine, Houston, TX; 3) Institute for Clinical and Translational Research, Baylor College of Medicine, Houston TX

---

Neoadjuvant chemotherapy (NCT) is the current standard of care for Triple Negative Breast cancer (TNBC). While the clinical benefits of NCT have been illustrated through numerous clinical trials, more than half of the patients do not respond to NCT and yet suffer from side effects. Currently, no clinically applicable biomarkers are available for predicting NCT response in TNBC. Therefore, developing predictive biomarkers for NCT in TNBC is an unmet need. In this study, we introduce a generic computational framework to calculate a response probability score (RPS) which predicts patient response to NCT. This framework identified genes associated with NCT response and captured the pre-treatment gene expression profile difference between NCT responders and non-responders. We first validated this framework in ER+ breast cancer and showed that it predicted NCT response with equal performance as clinically used biomarkers including Oncotype DX and MammaPrint. Then, we applied this framework to TNBC data and for each patient we calculated a response probability score (TNBC-RPS). Our results indicate that the TNBC-RPS achieved the best accuracy for predicting NCT response compared with all previously proposed gene signatures. More importantly, a model including TNBC-RPS and clinical predictors predicted patient response to NCT in TNBC with an accuracy that was comparable to ER+ NCT response prediction. In addition, we demonstrated difference in NCT response mechanisms between ER+ breast cancer and TNBC. In conclusion, TNBC-RPS can accurately predict NCT response in TNBC and has the potential to be clinically used to aid physicians in stratifying patients for more effective NCT treatment.

## PgmNr 861: A reporter mouse model enabling image-based evaluation of CD8<sup>+</sup> T cell dynamics in immunocompetent allografts.

### Authors:

B. Zhao; B. Wang; D. Feng; L. Ci; Y. Han; D. Huang; Y. Lan; J. Wei; J. Wang; R. Sun

[View Session](#) [Add to Schedule](#)

**Affiliation:** Shanghai Model Organisms Center, Inc., Shanghai, China

---

Immune cell recruitment to tumors is indispensable for the success of immunotherapy. In vivo bioluminescence imaging technology to visualize the location and functionality of immune cells is of great value for immuno-oncology research. Here, we describe the generation of a genetically engineered mouse model in which all CD8<sup>+</sup> T cells constitutively express a novel luciferase reporter Akaluc. On the C57/BL6 background, the IRES-Akaluc-EGFP expression cassette was inserted downstream of the stop codon, which allows the bioluminescence imaging of CD8<sup>+</sup> T cell location and function under physiological and pathological conditions. As expected, these engineered mice showed constitutive luciferase expression in lymphoid organs such as thymus, lymph nodes and spleen under normal conditions. Next, we modeled the CD8<sup>+</sup> T cell dynamics in the context of cancer immunotherapy. In the case of subcutaneously implanted mouse cancer cells such as MC38 colon cancer cells and antigenic SIY-expressing B16 melanoma cells (but not the regular B16 cells), the reporter mice showed enhanced T cell recruitment and infiltration into the tumor mass when treated with a mouse anti-PD-1 mAb. This apparent immune response was consistent with the eventual tumor regression. Besides, a similar level of T cell response was visualized when we treated tumor-bearing, PD-1-humanized mice with human anti-PD-1 mAb, suggesting the translatability of this type of reporter models.

Our study underscores the utilities of such reporter mice for uncovering the role of immune cells in tumor responses by revealing dynamics, interactions and distribution of various immune cells in tumor settings.

## PgmNr 862: Two point-NGS analysis of cancer driver genes in cell free-DNA of metastatic cancer patients.

### Authors:

E. Frullanti<sup>1</sup>; M. Palmieri<sup>1</sup>; M. Baldassarri<sup>2</sup>; F. Fava<sup>1,2</sup>; A. Fabbiani<sup>1,2</sup>; E. Gelli<sup>1</sup>; R. Tita<sup>2</sup>; P. Torre<sup>3</sup>; R. Petrioli<sup>3</sup>; T. Hadijstilianou<sup>4</sup>; D. Galimberti<sup>4</sup>; E. Cinotti<sup>5</sup>; C. Bengala<sup>6</sup>; M. Mandalà<sup>7</sup>; S.T. Miano<sup>3</sup>; I. Martellucci<sup>3</sup>; A. Vannini<sup>3</sup>; A.M. Pinto<sup>2</sup>; M.A. Mencarelli<sup>2</sup>; S. Marsili<sup>3</sup>; A. Renieri<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, University of Siena, Siena, Italy; 2) Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 3) Oncology, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 4) Dept. Ophthalmology, Referral Center for Retinoblastoma, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 5) Dept. Medical, Surgical and Neuro-Sciences, Dermatology Unit, University of Siena, Siena, Italy; 6) Medical Oncology, Ospedale Misericordia, Azienda Toscana Sud-Est, Grosseto, Italy; 7) Dept. Otology and Skull Base Surgery, University of Siena, Siena, Italy

---

**Background:** Despite the recognized efficacy of molecularly target agents in vitro, their use in routine setting is still hampered by the apparently low efficacy in vivo. Moreover, core biopsy of a single cancer site may not be representative of the whole expanding clones and cancer molecular profile at relapse may differ with respect to the primary tumor.

**Methods:** We assessed the status of a large panel of cancer driver genes by cell-free DNA (cfDNA) analysis in a cohort of 68 patients with 13 different solid tumors at disease progression. Whenever possible, a second cfDNA analysis was performed after a mean of 2.5 months, in order to confirm the identified clone(s) and to check the correlation with clinical evolution.

**Results:** The approach was able to identify clones plausibly involved in the disease progression mechanism in about 65% of cases. A mean of 1.4 mutated genes (range 1-3) for each tumor was found. Point mutations in *TP53*, *PIK3CA*, and *KRAS* and copy number variations in *FGFR3* were the gene alterations more commonly observed, with a rate of 48%, 20%, 16% and 20%, respectively. Two points NGS analysis demonstrated statistically significant correlation between allele frequency variation and clinical outcome ( $p=0,026$ ).

**Conclusions:** Irrespective of the primary tumor mutational burden, few mutated genes are present at disease progression. Clinical outcome is consistent with variation of allele frequency of specific clones indicating that cfDNA two point-NGS analysis of cancer driver genes could be an efficacy tool for precision oncology.

## PgmNr 863: Immunogenomic profiling for predicting outcome with advanced stage uterine serous carcinoma.

### Authors:

Y. Ni <sup>1</sup>; H. Mahdi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Quantitative Health Science, Cleveland Clinic, Cleveland, Ohio; 2) Section of Gynecologic Oncology, OBGYN and Women's Health Institute, Cleveland Clinic, Cleveland, Ohio

---

**Background:** The lethality of endometrial cancer is primarily due to disease with advanced stage (III or IV) at the time of diagnosis. Typically, a five-year survival rate of 83-97% is achieved for localized disease, in contrast to 43-67% for stage III disease and only 13-25% for stage IV disease. In addition to standard care, targeted therapies specific to individualized tumors, such as immunotherapy, are needed for advanced-stage patients. In this study, we focused on IFN $\gamma$  and TGF $\beta$  as key regulators in the tumor microenvironment (TME) and aim to identify immune-related gene signature in conjunction with IFN $\gamma$  and TGF $\beta$  signaling to predict serous uterine cancer outcome and response to chemotherapy.

**Method:** We conducted whole transcriptome analysis of matched tumor-normal samples from 38 high stage (Stage III and IV) uterine serous cancer patients who received chemotherapy with platinum-based therapy with taxanes. Immune-related genes were analyzed to identify differentially expressed genes separating overall survival (OS) and progression free survival (PFS). Additionally, IFN $\gamma$  18-gene expression score and TGF $\beta$  6-gene expression score were calculated by averaging the normalized and log transformed individual gene read counts. The optimized score cut off was selected to best separating the progression free survival. Then the cut off score was tested in The Cancer Genomic Atlas (TCGA) endometrial and ovarian cancer RNAseq datasets.

**Results:** A 16-gene immune signature was identified separating overall survival (OS) and progression free survival (PFS). The IFN $\gamma$  score of 2.46 was determined based on 18-gene expression derived from 38 high-stage uterine serous cancer samples with average age of 67 years (range: 56-82 years). Patients with score higher than 2.46 showed significantly longer PFS - 57.6 months vs 15months ( $p=0.002$ ) compared to the patients with score lower than 2.46. Then this IFN $\gamma$  based gene signature was then applied to TCGA endometrial cancer samples with RNAseq data. This signature predicted significant improvement in both PFS ( $p=0.001$ ) and OS ( $p=0.005$ ). Interestingly, high TGF $\beta$  score is correlated with worse PFS in these high stage serous patients, but not in TCGA endometrial cancer data.

**Conclusions:** Immune-related genomic signature can predict prognosis and response to chemotherapy in high stage serous endometrial cancer. Additional targeted markers relevant to TME may add more values for future therapeutic options.

## PgmNr 864: Individualized genetic network analysis reveals new therapeutic vulnerabilities.

### Authors:

F. Cheng

[View Session](#) [Add to Schedule](#)

**Affiliation:** 1Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44106, USA; 2Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH 44195, USA

---

Tumor-specific genomic alterations allow systematic identification of genetic interactions that promote tumor evolution and vulnerabilities, offering novel strategies for development of targeted therapies for individual tumors. We develop an Individualized Network-based Co-Mutation (INCM) methodology by inspecting over 2.5 million nonsynonymous somatic mutations derived from 6,789 tumor exomes across 14 cancer types from The Cancer Genome Atlas. Our INCM analysis reveals a higher genetic interaction burden on the significantly mutated genes, experimentally validated cancer genes, chromosome regulatory factors, and DNA repair genes, as compared to human genome-wide essential genes identified by CRISPR-Cas9 screenings on cancer cell lines. We find that genes identified in the cancer type-specific genetic subnetworks by INCM are significantly enriched in established cancer pathways, and the INCM-inferred putative genetic interactions are correlated with patient survival. By analyzing drug pharmacogenomics profiles from the Genomics of Drug Sensitivity in Cancer database, we show that the network-predicted putative genetic interactions (e.g., *BRCA2*-*TP53*) are significantly correlated with sensitivity/resistance of multiple therapeutic agents. We experimentally validated that afatinib has the strongest cytotoxic activity on BT474 ( $IC_{50} = 55.5$  nM, *BRCA2* and *TP53* co-mutant) compared to MCF7 ( $IC_{50} = 7.7$   $\mu$ M, both *BRCA2* and *TP53* wild type) and MDA-MB-231 ( $IC_{50} = 7.9$   $\mu$ M, *BRCA2* wild type but *TP53* mutant). Finally, drug-target network analysis reveals multiple druggable genetic interactions by targeting potentially tumor vulnerabilities. This study offers a powerful network-based methodology for identification of candidate therapeutic pathways that target tumor vulnerabilities and prioritization of potential pharmacogenomics biomarkers for development of personalized cancer medicine.

# PgmNr 865: Genome-wide association study of capecitabine-induced hand-foot-syndrome in colorectal cancer.

## Authors:

S. Dong; T. Wang; J. Zhang; Z. Wu; X. Li; P. Zhang; W. Jia

[View Session](#) [Add to Schedule](#)

**Affiliation:** State Key Laboratory of Oncology in South China, Sun Yat-sen University Cancer Center, Guangzhou, China

## Background:

Capecitabine, an oral prodrug of 5-fluorouracil (5-FU), is approved for first-line treatment of various malignancies, especially in colorectal cancer. Hand-foot-syndrome (HFS) is one of the common adverse events that limits its clinical use. Although HFS is not life-threatening, it always induces dose interruptions and reductions which adversely impacts treatment outcomes. Additionally, it also reduces patients' quality of life and hinder treatment compliance. Previous pharmacogenetics studies of HFS almost use candidate gene approach, but up till now, the pathophysiology and mechanism of HFS is still unclear, and the effective predictive biomarkers of HFS is yet to be found.

## Methods:

We used a two-stage GWAS design including a total of 1104 colorectal cancer patients who received capecitabine-based treatment in Sun Yat-sen University Cancer Center. HFS was graded according to NCI-CTCAE version 3.0. We performed a genome-wide association analysis in 514 patients, including 148 cases with grade 2-3 HFS and 366 controls with grade 0-1 HFS, and validated the top associations in an independent cohort of 590 patients. Sanger sequencing of the coding regions of *TYMS* was performed in 268 patients. All statistical tests were two-sided.

## Results:

In this retrospective study, we found the CRC patients in our cohort treated with capecitabine-based combination treatment had a lower HFS incidence, with a OR of 0.22 (95% CI = 0.22 - 0.50,  $P < 0.01$ ) compared with those treated by single-agent capecitabine, although patients almost received the same dose of capecitabine. Patients who had previously received fluorouracil-based chemotherapy treatment were more likely to develop HFS (OR = 1.89, 95% CI = 1.10 - 3.24,  $P = 0.021$ ). By genome-wide association study, we identified two variants, rs2853741 (OR = 2.01, 95% CI = 1.63-2.49,  $P_{\text{combined}} = 9.72 \times 10^{-11}$ ) in the promoter region of *TYMS*, and an intergenic polymorphism, rs1890775 (OR = 0.62, 95% CI = 0.5-0.76,  $P_{\text{combined}} = 3.91 \times 10^{-6}$ ) were strongly associated with HFS. *TYMS* an essential enzyme for DNA synthesis and repair is a critical target of capecitabine. Additionally, the coding regions sequencing of *TYMS* identified two novel variants in patients with severe HFS.

## Conclusions:

This genome-wide association study identifies two novel SNPs significant association with HFS in colorectal cancer patients in Chinese, which could help to elucidate the underlying mechanism of hand-foot-syndrome.

## PgmNr 866: Genomic analyses of L-asparaginase-induced pancreatitis in pediatric cancer patients.

### Authors:

B.I. Drogemoller <sup>1,2</sup>; G.E.B. Wright <sup>2,3</sup>; S.R. Rassekh <sup>2,3</sup>; S. Ito <sup>4</sup>; B.C. Carleton <sup>2,3,5</sup>; C.J.D. Ross <sup>1,2</sup>; The Canadian Pharmacogenomics Network for Drug Safety Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada; 2) BC Children's Hospital Research Institute, University of British Columbia, Vancouver, BC, Canada; 3) Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; 4) Clinical Pharmacology and Toxicology, The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 5) Pharmaceutical Outcomes Programme, BC Children's Hospital, Vancouver, BC, Canada

---

**Background:** L-asparaginase is highly effective in the treatment of pediatric acute lymphoblastic leukemia. Unfortunately, the use of this treatment is limited by the occurrence of pancreatitis, a severe and potentially lethal adverse drug reaction, which occurs in 2-18% of patients. As previous studies have been unable to identify strong associations between clinical variables and susceptibility to L-asparaginase-induced pancreatitis, genetic factors are expected to play an important role in this adverse drug reaction.

**Objectives:** We sought to explore the role of these genetic susceptibility factors in L-asparaginase-induced pancreatitis in pediatric cancer patients.

**Methods:** Patients who were treated with L-asparaginase were recruited from 13 pediatric oncology units across Canada ( $n=284$ ) and extensive clinical data were collected for all patients. Genotyping was performed using the Illumina HumanOmniExpress and Global Screening Arrays and pancreatic gene expression profiles were imputed in these individuals using GTEx v7 and S-PrediXcan. Genome- and transcriptome-wide associations (GWAS and TWAS) were performed to identify associations with L-asparaginase-induced pancreatitis.

**Results:** GWAS analyses identified significant associations between genetic variants in *HLA-DQA1* and *-DRB1* and pancreatitis, while TWAS revealed that individuals experiencing L-asparaginase-induced pancreatitis exhibited lower expression levels of *HLA-DRB5*. Further interrogation of the TWAS data revealed an enrichment in genes involved in the somatic diversification of immune receptors.

**Conclusions:** These analyses uncovered an association between genetic variation in immune-related genes and the development of L-asparaginase-induced pancreatitis. These associations mirror previous associations with the *HLA* region and (i) pancreatitis induced by other drugs and (ii) L-asparaginase-induced hypersensitivity.



## PgmNr 867: Transcriptome-wide association study of anthracycline-induced cardiotoxicity in pediatric cancer patients.

### Authors:

E.N. Scott <sup>1,2</sup>; G.E.B. Wright <sup>1,2</sup>; B.I. Drogemoller <sup>2,3</sup>; J.S. Hasbullah <sup>1,2</sup>; B.C. Carleton <sup>1,2,3,4,5</sup>; C.J.D. Ross <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; 2) BC Children's Hospital Research Institute, Vancouver, BC, Canada; 3) Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada; 4) Pharmaceutical Outcomes Programme, BC Children's Hospital, Vancouver, BC, Canada; 5) Division of Translational Therapeutics, Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

---

**Background:** Anthracyclines are highly effective chemotherapeutic agents; however, their clinical utility is limited by severe anthracycline-induced cardiotoxicity (ACT). Genome-wide association studies (GWAS) have uncovered several genetic variants associated with ACT, but these findings explain only a portion of its heritability.

**Methods:** To provide an in-depth examination of the genetics underlying ACT, we conducted a transcriptome-wide association study (TWAS) using our previous GWAS summary statistics (n=280 patients) to identify gene expression-related associations with ACT. Z-scores were used to describe the differential expression of genes between cases and controls. Next, we conducted gene set enrichment analyses in which a Wilcoxon rank sum test was used to compare mean  $Z^2$  statistics from each gene set with those of all other genes in the TWAS. Finally, pathway analyses were conducted to investigate biological mechanisms underlying ACT.

**Results:** Three genes expressed in heart/arterial tissues were significantly associated with ACT in the TWAS: *GDF5* (Z-score=-4.30,  $P=1.70 \times 10^{-5}$ ), *FRS2* (Z-score=4.07,  $P=4.67 \times 10^{-5}$ ), and *HDCC2* (Z-score=4.01,  $P=6.08 \times 10^{-5}$ ). Significantly enriched gene sets consisted of genes essential in: mice ( $P=2.0 \times 10^{-7}$ ), culture ( $P=0.036$ ), and humans ( $P=1.36 \times 10^{-5}$ ), as well as genes differentially expressed upon treatment with the cardioprotectant all-trans retinoic acid (ATRA,  $P=1.53 \times 10^{-4}$ ) and genes up-regulated at lower doses of anthracyclines ( $P=3.60 \times 10^{-4}$ ). Pathway analyses revealed an overrepresentation of genes involved in ribosomal, spliceosomal, and cardiomyopathy pathways. Results from functional validation conducted to confirm the roles of *GDF5*, *FRS2*, and *HDCC2* in ACT will also be presented.

**Conclusions:** In summary, we have identified three novel genetic associations with ACT; determined that genes essential for survival, cardioprotection, and response to anthracyclines are important for development of ACT; and uncovered pathways dysregulated in patients with ACT, providing further insight into the mechanism of ACT.

## PgmNr 868: A genome-wide association study identifies five novel genetic markers for trastuzumab-induced cardiotoxicity.

### Authors:

C. Udagawa<sup>1,2</sup>; M. Hara<sup>1,3</sup>; A. Shimo<sup>3</sup>; Y. Kojima<sup>3</sup>; R. Yoshie<sup>3</sup>; H. Zaha<sup>4</sup>; N. Abe<sup>4</sup>; T. Motonari<sup>4</sup>; M. Unesoko<sup>4</sup>; K. Tamura<sup>5</sup>; T. Shimoi<sup>5</sup>; M. Yoshida<sup>6</sup>; T. Yoshida<sup>7</sup>; H. Okamura<sup>2</sup>; T. Mushiroda<sup>8</sup>; K. Tsugawa<sup>3</sup>; H. Zembutsu<sup>1,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Project for Development of Liquid Biopsy Diagnosis, Cancer Precision Medicine Center, Japanese Foundation for Cancer Research, Tokyo, Japan; 2) New Business Development Life Science Group, Toyo Kohan Co., Ltd., Tokyo, Japan; 3) Department of Breast and Endocrine Surgery, St. Marianna University School of Medicine, Kawasaki, Japan; 4) Department of Breast Surgery, Nakagami Hospital, Okinawa, Japan; 5) Department of Breast and Medical Oncology, National Cancer Center Hospital, Tokyo, Japan; 6) Department of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo, Japan; 7) Fundamental Innovative Oncology Core, National Cancer Center Research Institute, Tokyo, Japan; 8) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 9) Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan

**Purpose:** Trastuzumab has been administered to patients with HER2-positive cancer, however, the cardiotoxicity is identified as one of the life-threatening toxicities. Clinically useful biomarker for trastuzumab-induced cardiotoxicity has been expected to be developed. To identify novel genetic markers determining the risk of trastuzumab-induced cardiotoxicity, we performed a first genome-wide association study (GWAS) in Japanese population.

**Experimental Design:** We enrolled 481 patients who had been treated with trastuzumab and carried out a GWAS using 11 cases (with cardiotoxicity) and 257 controls (without cardiotoxicity). Top 100 single nucleotide polymorphisms (SNPs) which revealed the smallest  $P$  values in GWAS ( $P = 7.60 \times 10^{-7}$  -  $2.01 \times 10^{-4}$ ) were further examined using replication samples consisted of 14 cases and 199 controls.

**Results:** The combined analysis of the GWAS and replication study indicated possible association of five loci with trastuzumab-induced cardiotoxicity (HERADR1 on chromosome 13q14.3, HERADR2 on chromosome 15q26.3, HERADR3 on chromosome 17q25.3, HERADR4 on chromosome 4q25, and HERADR5 on chromosome 15q26.3,  $P_{combined} = 6.00 \times 10^{-6}$ ,  $8.88 \times 10^{-5}$ ,  $1.07 \times 10^{-4}$ ,  $1.42 \times 10^{-4}$ ,  $1.60 \times 10^{-4}$ , respectively). Furthermore, we developed a risk prediction model for trastuzumab-induced cardiotoxicity using five marker SNPs. The incidence of trastuzumab-induced cardiotoxicity in patients with risk score  $\geq 5$  was significantly higher (42.5%) compared to that in patients with score  $\leq 4$  (1.8%) ( $P = 7.82 \times 10^{-15}$ , odds ratio = 40.0).

**Conclusion:** We identified five novel genetic loci associated with trastuzumab-induced cardiotoxicity. These findings provide new insights into precision therapy for patients with HER2-positive cancer.

## PgmNr 869: Whole genome sequencing to identify predictive marker for the risk of drug-induced interstitial lung disease.

### Authors:

H. Zembutsu <sup>1</sup>; C. Udagawa <sup>1</sup>; H. Horinouchi <sup>2</sup>; K. Shiraishi <sup>3</sup>; T. Kohno <sup>3</sup>; T. Okusaka <sup>4</sup>; H. Ueno <sup>4</sup>; K. Tamura <sup>5</sup>; Y. Ohe <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Project for Development of Liquid Biopsy Diagnosis, Japanese Foundation for Cancer Research, Research Institute, Tokyo, Japan.; 2) Department of Thoracic Oncology, National Cancer Center Hospital, Tokyo, Japan.; 3) Division of Genome Biology, National Cancer Center Research Institute, Tokyo, Japan.; 4) Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital, Tokyo, Japan.; 5) Department of Breast and Medical Oncology, National Cancer Center Hospital, Tokyo, Japan.

---

Drug-induced interstitial lung disease (DIILD) is one of the critical issues of chemotherapy for patients with cancer because the mortality of this adverse event is extremely high. To identify genetic variants determining the risk of DIILD, we carried out whole genome sequencing of germline DNA samples from 26 patients who developed DIILD, and conducted a case-control association study between these 26 cases and general Japanese population controls registered in the integrative Japanese Genome Variation Database (ijGVD) as a screening study. The associations of 42 single nucleotide variants (SNVs) showing  $P < 0.0001$  were further validated using an independent cohort consisted of 18 DIILD cases as a replication study. Combined analysis of the screening and replication studies showed a possible association of two SNVs with DIILD ( $P_{combined} = 1.87 \times 10^{-5}$  and  $5.16 \times 10^{-5}$ , respectively). Furthermore, in subgroup analysis of epidermal growth factor-tyrosine kinase inhibitor-induced ILD, we observed 7 candidate SNVs which were possibly associated with the ILD ( $P < 0.00001$ ). These SNVs might be applicable in predicting the risk of DIILD in patients receiving chemotherapy.

## PgmNr 870: Intelligent biomarker selection and individualized therapy.

### Authors:

Q. Ge <sup>1,2</sup>; Y. Liu <sup>2</sup>; X. Huang <sup>3</sup>; M. Xiong <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Fudan University, Shanghai, CHINA; 2) University of Texas School of Public Health; 3) M. D. Anderson Cancer Center

---

The “precision medicine” approach with electronic health records (EHRs), lab systems, imaging, pathology and genomics attempts to match the most accurate and effective treatments with the individual patient, rather than one-size-fits-all monotherapy. Treatment response is heterogeneous. Therefore, individualized treatment effect (ITE) estimation and therapy selection from observational data are a central talk in the precision medicine. We want to select the optimal treatment for each individual to ensure that the right therapy is offered to “the right patient at the right time”. However, if it were to use a standard estimation platform for each treatment separately, the estimated treatment effects are the average effects in the population, which cannot be taken as individual effect. Unfortunately, estimating individualized treatment effects and design individualized therapy is beyond the state of the art of current medical knowledge. The true barrier we face is the challenge of developing intelligent algorithms for estimating individualized treatment effects and designing optimal individualized therapy via integrating imaging, HER, physiological and omics data. To meet the challenge, we develop a novel analytic platform that integrate auto-encoding, conditional generative adversarial networks (CGAN) and sparse techniques for individualized treatment effect estimation, biomarker selection and optimal therapy decision. We use auto-encoder for high dimensional feature data reduction, take counterfactual estimation as a missing value problem and adapt the well-known CGAN framework for imputing the counterfactual outcomes and estimating the individual treatment effects. After the individualized treatment effects are estimated, sparse regression methods will be used to select features (biomarkers), to predict treatment response and to select the optimal treatment for each individual. Due to the lack of ground truth, we conduct large scale simulations where imaging, omics and pathology features, treatment assignment and potential outcomes are generated, to evaluate the performance of the proposed algorithms. The algorithms are also applied to newly diagnosed acute myeloid leukemia (AML) of 256 patients, treated with multiple therapy at M. D. Anderson Cancer Center. Both simulation and real data analysis results show that the intelligent algorithms substantially outperform the state-of-the-art methods.

## PgmNr 871: Prospective modeling of genetic risk of thromboembolism in acute lymphoblastic leukemia in a competing risks setting.

### Authors:

M. LeBlanc <sup>1</sup>; K. Jarvis <sup>1</sup>; M. Tulstrup <sup>2</sup>; R. Nielsen <sup>3</sup>; R. Gupta <sup>3</sup>; P. Huttunen <sup>4</sup>; O. Jonsson <sup>5</sup>; C. Rank <sup>2</sup>; S. Ranta <sup>6</sup>; E. Ruud <sup>1</sup>; K. Saks <sup>7</sup>; R. Tuckuviene <sup>8</sup>; S. Trakymiene <sup>9</sup>; K. Schimiegelow <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oslo University Hospital, Oslo, Norway; 2) Rigshospitalet, University Hospital of Copenhagen, Denmark; 3) Technical University of Denmark, Denmark; 4) Helsinki University Hospital, Finland; 5) Landspítali University Hospital, Iceland; 6) Karolinska Institutet, Sweden; 7) Tallinn Children's Hospital, Estonia; 8) Aalborg University Hospital, Denmark; 9) Vilnius University Children's Hospital, Lithuania

---

Thromboembolism (TE), while an extremely rare event in healthy children, is a relatively common serious toxicity of acute lymphoblastic leukemia (ALL) treatment and contributes to substantial morbidity and mortality. Several SNPs, including functional SNPs, have been associated with TE in the general adult population. However, the impact of these SNPs in patients with ALL, particularly in children, remains uncertain. Since ALL patients are also at risk for other events (including non-TE related mortality and relapse), these so-called competing risks need to be taken into account in the prospective statistical modeling of the genetics of TE in ALL patients. We prospectively registered TE, mortality and other competing events in a cohort of 1,252 patients aged 1–45 years (91% <18 years) with ALL included in the Nordic Society of Pediatric Hematology and Oncology (NOPHO) ALL2008 protocol in the Nordic and Baltic countries (7/2008–7/2016). Based on previously published non-cancer data, we selected four SNPs: *F5* rs6025, *F11* rs2036914, *FGG* rs2066865, and *ABO* rs8176719. We constructed genetic risk scores using a combination of these SNPs and then modeled risk of TE in a competing risks setting. The 2.5 year cumulative incidence of TE was 7.1% (95% confidence interval (CI) 5.6–8.5). *F11* rs2036914 was associated with TE development (HR 1.52, 95% CI 1.11–2.07) and there was a borderline significant association for *FGG* rs2066865 (HR 1.37, 95% CI 0.99–1.91), but no significant association for *ABO* rs8176719 or *F5* rs6025. A genetic risk score based on *F11* rs2036914 and *FGG* rs2066865 was associated with TE development (HR 1.45 per risk allele, 95% CI 1.15–1.81), the association was strongest in adolescents 10.0–17.9 years (HR 1.64). If validated, a *F11* rs2036914/*FGG* rs2066865 risk prediction model should be tested as a stratification tool for TE prevention measures in patients with ALL. Current work is focusing on polygenic risk scores based on TE-GWAS Contortia summary statistics.

## PgmNr 872: Precision medicine: Predicting lung cancer drug responses.

**Authors:**

C.H. Wu; E.S.C. Shih; M.J. Hwang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, Taiwan

---

Lung cancer ([MIM 211980]) causes severe healthcare problems worldwide. Here, we identified the drug response genes of tyrosine kinase inhibitors Erlotinib and Sorafenib, and showed that these genes can perfectly classify a small set of patients from the BATTLE clinical trial who were without biomarker mutations into responding and non-responding using machine learning and regression approaches. Using pathway and protein-protein interaction network analysis, we further showed that these genes are associated with kinase signaling pathways. Specifically, Erlotinib-associated genes are related to protein ubiquitination and DNA damage, and Sorafenib-associated genes are related to autophagy, apoptosis and proliferation, shedding light on causes for the drug resistance. Our analysis also suggests that ADORA3 could be is a good biomarker and target to counter Erlotinib resistance. Interestingly, our study suggests that some of these Erlotinib/Sorafenib-associated genes not only are drug-response biomarkers, but also are prognostic biomarkers for *EGFR* wild-type lung cancer patients. Our AI (artificial intelligence)-based drug response model has a potential to contribute significantly to precision medicine treatment of lung cancer patients.

## PgmNr 873: Effect of cisplatin on MAPK pathway gene expression is reversed by naringenin in HCT116 colorectal cancer cell lines.

### Authors:

A.G. Thornburg <sup>1</sup>; R. Georgiadis <sup>2</sup>; A.L. Stark <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Case Western Reserve University, Cleveland, Ohio; 2) University of Notre Dame, Notre Dame, Indiana

---

Platinum agents are widely used to treat cancer inducing apoptosis and cell death by contorting the DNA with platinum adducts. Despite multiple platinum based therapies (cisplatin, carboplatin, oxaliplatin), resistance, poor response, and side effects still plague their use. The MAPK pathway is a critical pathway to study in cancer treatment due to its dysregulation in many cancers. In this project, we aim to understand the role of platinum-induced MAPK pathway expression. We began by assessing 5 genes in the MAPK pathway in HCT116 cells (colorectal cell line) after treatment with cisplatin and oxaliplatin. Cisplatin and oxaliplatin were found to induce MAPK pathway genes differently, suggesting that while platinum is the active component for both, their effect on MAPK signaling is distinct. Oxaliplatin did not induce a clear pattern across the pathway genes evaluated, while cisplatin did. All assayed genes demonstrated no change or downregulation until 48hr, where all genes showed significant upregulation by cisplatin. We then queried both DLD1 (another colorectal cancer cell line) and A549 (lung cancer cell line) treated with cisplatin to assess if the same MAPK pathway induction was seen. Interestingly, A549 cells revealed the opposite effect. Cisplatin treated A549 cells were observed to have minimal induced expression changes until 48 hr where significant downregulation of MAPK genes was observed. DLD1 did not have an observable pattern of induction across the pathway, suggesting that the effect of cisplatin on HCT116 is specific to that cancer subtype. To further understand the effects of cisplatin on the MAPK pathways in HCT116 cells, naringenin was identified as a broad inhibitor of the MAPK pathway. HCT116 cells were treated with cisplatin, naringenin, and a combination of cisplatin and naringenin. Surprisingly, at 48 hr post treatment, 4 out of the 5 genes assayed (CREB1, RAF1, MAPK1, and MAP2K1) demonstrated a significant reversal of the induction observed by cisplatin. Individual treatment of cisplatin and naringenin resulted in significant upregulation at 48 hr. However, combination treatment resulted in significant downregulation at 48 hr. These data suggest an intriguing interaction between cisplatin and naringenin affecting the MAPK pathway. The next steps include understanding the effect this interaction has on apoptosis and cell death with the ultimate goal of improving the effectiveness of cisplatin treatment in colorectal cancer.



## PgmNr 874: The Genetic Risk Estimate (GENRE) trial: Influence of a breast cancer (BC) polygenic risk score (PRS) on personalized approach to understanding BC risk and decision to take preventive medication.

### Authors:

S. Pruthi <sup>1</sup>; J.O. Kim <sup>2</sup>; D. Schaid <sup>5</sup>; A. Cooke <sup>2</sup>; C. Kim <sup>2</sup>; B. Goldenberg <sup>2</sup>; J. Sinnwell <sup>5</sup>; D. Grenier <sup>2</sup>; A. Degnim <sup>1</sup>; F. Couch <sup>4</sup>; C. Vachon <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of General Internal Medicine- Breast Diagnostic Clinic, Mayo Clinic, Rochester, Minnesota.; 2) Cancer Care Manitoba, Winnipeg, Manitoba Canada; 3) Health Science Research, Mayo Clinic; 4) Lab Medicine and Pathology, Mayo Clinic; 5) Statistics and Informatics, Mayo Clinic

**Background:** Studies demonstrate breast cancer risk reduction of 50-65% with the use of preventive medications. Multiple barriers exist that negatively influence uptake including attitudes toward medications, risk perception and drug side-effects. A PRS comprised of 77 BC genetic susceptibility loci (Single Nucleotide Polymorphisms (SNP)) can be used to personalize breast cancer risk. We studied the influence of the addition of a PRS to standard risk calculator estimates on the decision to receive prevention medication.

**Methods:** Eligible women required either a 5 year Gail Model risk of  $\geq 3\%$  or 10 year Tyrer-Cuzick risk (IBIS) of  $\geq 5\%$ . Women with a history of BC or hereditary BC syndrome were excluded. High risk women were counseled at baseline using their Gail and IBIS risk scores and preventive medication options were discussed including benefits and risks. Participants completed a self-reported questionnaire at baseline to assess understanding of BC risk and decision to take preventive medication. Blood samples were obtained and genotyped for 77 SNPs, and an updated BC -PRS risk report was shared with study participants that reflected the IBIS and Gail risk predictions for 5 yr, 10 yr, & lifetime BC risk with and without the PRS. Post visit questionnaires assessed influence of PRS on decision to take preventive medication.

**Results:** 151 women in Canada & USA were enrolled on the study from 2016 to 2017. The median age was 56.1 (range 36-76.4), 35.6% were premenopausal, 98.7% were Caucasian and 64.7% had  $\geq 1$  family member with BC. Median 5yr, 10yr, & lifetime IBIS risk estimates were 3.8% (2.0-11.5), 7.9% (5.0-23.1), and 25.3% (5.5 to 92.2). After the BC-PRS risk estimate report was shared with the participant, the influence on their decision to take preventive medication significantly changed ( $p < 0.001$ ): 41.9 % were more likely to take preventive medication if there was an increase in the PRS risk estimate and 46.7 % were less like to take preventive medication if they had a decrease in the PRS risk estimate.

**Conclusion:** In high risk women, the personalized approach to BC-PRS risk estimates in addition to standard BC risk calculators significantly influenced the decision to take preventive medication. Future study is underway on how the PRS impacts uptake and adherence of preventive medication.



## PgmNr 875: Systematic establishment of robustness and standards in patient-derived xenograft experiments and analysis.

### Authors:

M.W. Lloyd <sup>3</sup>; .The NCI PDXNet Consortium <sup>1,2,3,4,5,6,7,8,9, . . 10,11,12</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Leidos Biomedical Research, Inc, Frederick National Laboratory for Cancer Research, Frederick, MD; 2) Seven Bridges Genomics, Cambridge, MA; 3) The Jackson Laboratory, Bar Harbor, ME.; 4) The Jackson Laboratory for Genomic Medicine, Farmington, CT.; 5) Hunstman Cancer Institute, Salt Lake City, UT; 6) Baylor College of Medicine, Houston, TX; 7) The University of Texas M.D. Anderson Cancer Center, Houston, TX.; 8) The Wistar Institute, Philadelphia, PA; 9) Washington University School of Medicine, St. Louis, MO; 10) Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD; 11) Investigational Drug Branch, National Cancer Institute, Bethesda, MD; 12) University of Connecticut Health Center, Farmington, CT

---

Patient-Derived Xenograft (PDX) models present valuable systems for demonstrating clinical translatability in preclinical testing of therapeutic agents. Large PDX collections, such as those in active development in the NCI PDXNet Consortium, provide a critical resource for assays of therapeutic efficacy. However, variation in experimental and assessment procedures among research groups often limit conclusions regarding the robustness of PDX results with regard to patient outcomes. To assess the sensitivity of PDX assays with regard to different institutional procedures, PDXNet conducted a study of temozolomide drug response, in replicate, on PDX models from three patients. Four PDX Development and Trial Centers (PDTCs) used internally developed SOPs that varied in procedures for establishment of xenografts, treatment dosing and scheduling, time period of study, and response criteria, with all sites blinded to the procedures and results of other groups. The three PDX models were chosen based on prior treatment experiments conducted by the NCI Patient Derived Models Repository, which showed one model was sensitive, one was resistant, and one had intermediate sensitivity. Drug responses measured by PDTCs were consistent across all testing groups, with each group correctly classifying the sensitive, resistant, and intermediate models. Additionally, to assess consistency in PDX genomic data, all PDTCs followed internal sequencing procedures to generate one exome-seq and one RNA-seq dataset per PDX model. All sequence data were analyzed on the Seven Bridges Genomics Cancer Genomics Cloud platform using benchmark optimized workflows for whole exome variant calling and RNA expression estimation. Across PDTCs variant calls derived from exome-seq data showed strong overlap within PDX models, but were affected by differing panels and variable read depth. RNA expression results showed strong batch effects, but following correction clustered by PDX model. Overall, our results show that drug responses and genomic data of PDX models from across different research institutions are consistent and robust. We report standardized SOPs for experimental procedures, response measurement and statistical assessment, and sequence analysis workflows based on these results. We expect these PDXNet standards to improve the use of PDX and other in vivo models in advancing cancer precision medicine.

## PgmNr 876: *TNS1*: A candidate gene for predisposition to and/or development of pheochromocytoma.

### Authors:

F. Faucz<sup>1</sup>; A.G. Maria<sup>1</sup>; N. Settas<sup>1</sup>; A. Berthon<sup>1</sup>; S.K. Flores<sup>2</sup>; P.L. Dahia<sup>2</sup>; C.A. Stratakis<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 2) Dept. Medicine and Mays Cancer Center, University of Texas Health Science Center at San Antonio, TX.

---

*TNS1* gene encodes the actin-binding protein Tensin 1 (TNS1). Tensin1 is an adaptor phosphoprotein linking integrin to the actin cytoskeleton; it is also involved in fibrillar adhesion formation. *Tns1* knockout (KO) mice develop kidney abnormalities; female mice have in addition reduced fertility. A *drosophila melanogaster Tns1* KO model is known as “*blistery*” due to defective wing unfolding and the consequent blister formation. Genome-wide association studies have linked *TNS1* to asthma, lung disease and colorectal cancer, and *TNS1* expression is reduced or absent in prostate and breast cancer cells. Recently our group identified a high frequency of *TNS1* mutations in patients with pheochromocytomas (PHEO). Among 27 patients where a known causative mutation was not found (including genes such as SDHx, NF1, RET, HRAS, EPAS1, VHL, and FH) we identified a total of 7 different *TNS1* coding sequence alterations, which were predicted to be damaging through in silico analysis (33.33% - 9/27). In order to access signaling pathways of *TNS1* and its functionality in PHEO, shRNA and siRNA approaches were used to knockdown *TNS1* in a human (h)PHEO cell line. Both gene and protein expression were decreased by quantitative real-time PCR and western blot, respectively. Quantitative real-time PCR was performed to access gene expression and cell viability was accessed by MTT assay. Our results show that decrease of *TNS1* expression in hPHEO cells led to decreased expression of collagen I and collagen II by 6-and 4-fold, respectively. On the other hand, *MMP3* expression was increased by 3-fold in these cells. hPHEO cells treated with *TNS1*s iRNA showed increased cell viability after 72 hours of treatment (p=0.02). Significant differences in cell viability after 24 and 48 hours of *TNS1*s iRNA treatment were not observed in this study. Taken together, our results suggest that *TNS1* dysfunction leads to higher cell viability and increased *MMP3* expression. These data are complementary to our previously reported finding of mutations in this gene both at the germline and tumor level in PHEOs (and other tumors). We speculate that *TNS1* defects may play an important role in PHEO predisposition and/or development and progression; *TNS1* could be yet another molecular target of new therapies for this and related tumors.

## PgmNr 877: A case of preimplantation genetic diagnosis for a female patient with a Peutz-Jeghers syndrome caused by a *de novo* deletion in the *STK11* gene: Next-generation sequencing contribution.

### Authors:

M. Willems <sup>1</sup>; V. Viart <sup>2</sup>; S. Baert-Desurmont <sup>3</sup>; T. Anahory <sup>4,5</sup>; S. Hamamah <sup>5</sup>; P. Pujol <sup>1</sup>; D. Geneviève <sup>1</sup>; C. Coubes <sup>1</sup>; A. Girardet <sup>2</sup>; C. Corsini <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, Hôpital Arnaud de Villeneuve, CHU de MONTPELLIER, MONTPELLIER, Occitanie, France; 2) Molecular genetics of Preimplantatory Diagnosis laboratory, Montpellier hospital, France; 3) Molecular genetics laboratory, Rouen hospital, France; 4) Gynecology department, Montpellier hospital, France; 5) Reproductive biology department, Montpellier hospital, France

### Aim

The aim was to propose a preimplantation genetic diagnosis (PGD) to a female patient suffering from a Peutz-Jeghers syndrome (hamartomatous polyps and mucocutaneous melanin hyperpigmentation) due to a *de novo* deletion (g.1222449\_1225004del) including exon 8 and its flanking intronic regions in the *STK11* gene.

### Methods

The deletion was initially identified by QMPFS and next-generation sequencing (NGS). The identification of the breakpoints by NGS has allowed to define a 2555 nucleotides deletion. The PGD protocol was based on the direct diagnosis of the deletion in addition to the study of polymorphic-linked microsatellite markers defining the 2 parental haplotypes. As PGD criteria did not allow analysis of fragment > 400 bp, we designed primers each sides of the breakpoints and in intron 7 in order to amplify simultaneously the wild-type (122 bp) and the mutated fragment (142 bp).

### Results

A prePGD workup was performed on DNA from the couple and each parents. At this stage, the *de novo* inheritance induced that the deletion could not be associated to a parental haplotype in this patient. However, familial segregation showed that the patient was fully-informative for 4 microsatellite markers and semi-informative for 2 markers, whatever the haplotype associated to the deletion. In our centre we did not perform polar body analysis. Thus, this PGD was accepted in condition of the analysis of at least 6 embryos (2 cells per embryo) with at least 2 affected embryos, in order to definitively associate the deletion to the correct haplotype. The first PGD cycle led to the freeze of 2 embryos at day 3, and the second one to the freeze of 4 embryos at day 2. Unfortunately, these 4 embryos did not develop correctly for a further transfer. Also we decided to analyze them to perform familial segregation and found the deletion in only one embryo. The 2 remaining frozen embryos were analyzed, one was affected and the second one displayed only one haplotype instead of 2, resulting in no transfer.

### Conclusion

Thanks to the identification of the breakpoints by NGS, the deletion was found and associated to the correct haplotype by the study of the embryos. This PGD for Peutz-Jeghers syndrome did not conduce to a transfer but a spontaneous pregnancy ensued for this couple. At our knowledge, this is a first

case of PGD for a *de novo* deletion. A similar request is performed in another centre in France.

## **PgmNr 878: Non-invasive screening using cell-free DNA as first line of screen at the first trimester in the Israeli population.**

### **Authors:**

D. Bercovich <sup>1</sup>; S. Horn-Saban <sup>2</sup>; R. Gershoni <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Human Molec Gen & Pharm, Tel Hai College, Tel Hai, Israel; 2) Galile Genetic Center, Kazarin, Israel

---

Non-invasive screening using cell-free DNA (cfDNA) from maternal blood has greatly improved the reliable detection of autosomal aneuploidies in recent years. To implementation this method, as a first line of screening test (first trimester screening), we used (for the first time in Israel), the Illumina NIPS platform, on 1444 cfDNA samples from the Israeli population. Maternal average age was 34.3 (min: 18.9; max: 48.6), gestational average age was 16.9 (min: 10.1; max: 34.7), average fetal fraction was 11% (min: >1%; max: 31%). Results were obtained for 1375 (95.2%) samples, of which 1255 (91.3%) derived from singleton and 120 from twins (8.7%) pregnancies. Singleton pregnancies resulted in 659 XY and 596 XX pregnancies. Normal karyotype was reported in 1350 (98%) and abnormal results were derived from the analysis of 28 samples of which 18 were clinically confirmed and 5 samples were normal (0.36% false positive). As far as we know, no false negative results were obtained. Abnormal karyotypes found included T21 (n=15), T18 (n=3), T13 (n=4) and one XXY. Out of the 69 samples which were not reported, 47 had low fetal fraction (less than 1%), 13 were at borderline values, in 6 samples the data was out of range and 3 were rejected because of technical parameters (0.21%). Overall we found that NIPS is a reliable and replaceable method to be used as first line prenatal screening. Other than this NIPS will continue to increase our insight into early gestational placental findings not previously well recognized.

## PgmNr 879: Cancer gene mutations in congenital pulmonary airway malformation patients.

### Authors:

J. Hsu <sup>1,2,3</sup>; R. Zhang <sup>4</sup>; F. Yeung <sup>5</sup>; C. Tang <sup>5</sup>; J. Wong <sup>2</sup>; M. So <sup>5</sup>; H. Xia <sup>4</sup>; P. Sham <sup>2,3</sup>; P. Tan <sup>5</sup>; M. Li <sup>2,3</sup>; K. Wong <sup>5</sup>; M. Garcia-Barcelo <sup>3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Molecular and Genomic Medicine, National Health Research Institute, Miaoli County 350, Taiwan; 2) Dept of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; 3) Centre for Genomics Science, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China; 4) Dept of Pediatric Surgery, Guangzhou Women and Children's Medical Center, Guangzhou, China; 5) Dept of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

---

Newborns affected with congenital pulmonary airway malformations (CPAMs) may present with severe respiratory distress. While surgical resection is the definitive treatment for symptomatic CPAMs, prophylactic elective surgery may be recommended for asymptomatic CPAMs owing to the risk of tumor development. However, the implementation of prophylactic surgery is quite controversial on the grounds that more evidence linking CPAMs and cancer is needed. The large gap in knowledge of CPAM pathogenesis results in uncertainties and controversies in disease management. As many known developmental genes control postnatal cell growth and contribute to cancer development as well, we hypothesized that CPAMs may be underlain by germline mutations in genes governing airways development. Sequencing of the exome of 19 patients and their unaffected parents. A more than expected number of mutations in cancer genes (false discovery rate  $q$ -value  $< 5.01 \times 10^{-5}$ ) was observed. The co-occurrence, in the same patient, of damaging variants in genes encoding interacting proteins is intriguing, the most striking being thyroglobulin (*TG*) and its receptor, megalin (*LRP2*). Both genes are highly relevant in lung development and cancer. The overall excess of mutations in cancer genes may account for the reported association of CPAMs with carcinomas and provide some evidence to argue for prophylactic surgery by some surgeons.

## PgmNr 880: Splicing-QTL analysis identified regulatory variants and their target genes at prostate cancer risk loci.

### Authors:

Y. Tian; L. Wang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pathology, Medical College of Wisconsin, MILWAUKEE, WI.

---

Genome-wide association studies (GWAS) have identified hundreds of single nucleotide polymorphisms (SNPs) associated with an increased risk to prostate cancer (PrCa). To build links between risk SNPs and potential genes, expression quantitative trait loci (eQTL) analysis has been widely used. To complement the eQTL analysis, we present splicing QTL (sQTL) analysis using RNA-seq data from benign prostate tissues (n=467). From our 2.5M Illumina genotype data, we first performed population-based imputation by Markov Chain Haplotyping (MaCH1) and selected risk variants by LDproxy ( $R^2 \geq 0.5$ ). We then applied RNA-Seq by Expectation-Maximization (RSEM) package to estimate transcript-level profiles. Integrating the SNP genotypes and RNA profiles, we applied sQTLseeker package to perform a distance-based splicing ratio analysis. With the input of transcript abundance (n=114,670) and risk SNPs (n=21,986) at PrCa loci, we identified significant associations in 1,533 SNP-transcript pairs ( $FDR \leq 0.1$ ). Among those were two sQTLs [rs4074037-*MUC1* (CA15-3) ( $FDR=9.9E-5$ ), rs10424178-*BABAM1* ( $FDR=0.056$ )], which were previously reported in other cancers. We also found new sQTLs including a common SNP (rs7247241) and its susceptible gene (*PPP1R14A*) ( $FDR=0.089$ ), with allele C being linked to a short isoform (ENST00000587515), leading to 50% shortened protein product. To functional characterize this new sQTL, we examined ENCODE ChIP-seq tags and found preferred CTCF binding for rs7247241-T allele in both prostate cell lines and immortalized lymphoblastoid cell lines. Bisulfite sequencing from ENCODE cell lines demonstrated a negative correlation between CTCF binding at this locus ( $\pm 75$ bp flanking region) and averaged methylation percentage ( $r=-0.512$ ). Interestingly, allele-specific methylation analysis showed a strong association of rs7247241-C allele and hypomethylation at this locus. To evaluate the role of the *PPP1R14A* short isoform, we examined the percent splice-in value (PSI) for the short isoform and observed significant downregulation in TCGA primary prostate tumors ( $p=6.21E-14$ ). Further analysis showed a positive association between the short isoform and relapse-free survival ( $p=3E-6$ ). Overall, by analyzing a large prostate-specific SNP genotype-gene expression database, we identified a set of sQTLs at PrCa risk loci. We characterized regulatory SNP rs7247241 and its role in controlling DNA methylation of flanking sequence, splicing event of target gene *PPP1R14A*.

## PgmNr 881: Mechanistic insights of inactivation of chitinase 3-like-1.

**Authors:**

N. Kishigami; K. Okawa; F. Oyama

[View Session](#) [Add to Schedule](#)

**Affiliation:** Kogakuin University, Hachioji, Tokyo, Japan

---

Chitinase-like proteins are structurally homologous to chitinase but lack the ability to degrade chitin. It is generally assumed that the lack of chitinase activity in chitinase-like proteins is due to the mutation of crucial residues within the conserved catalytic sequence. Chitinase 3-like-1 (Chi3l1), one of the chitinase-like proteins, is expressed in mice and human. Chi3l1 levels are increased in individuals with asthma, chronic obstructive pulmonary disease (COPD), cystic fibrosis, rheumatoid arthritis, inflammatory bowel disease, alcoholic cirrhosis and many types of malignant tumors. Thus, Chi3l1 may play important roles in the pathophysiological conditions. However, the contribution of the protein to the pathophysiology of the related diseases remains to be determined. Despite Chi3l1 has a relatively high sequence homology to chitotriosidase (Chit1), it lacks chitinase activity. Although we attempted to activate Chi3l1 by introducing two substitutions of amino acid (A139D and L141E) in the active center of the wild type Chi3l1, it remained inactive. We have shown that M at the position 61 is highly conserved in the mammalian chitinases. Although M61 is conserved in both mouse Chit1 and AMCase, Chi3l1 has I61. Thus, we predicted that the amino acid at the position 61 is involved in chitinase activity, and we attempted to activate Chi3l1 by introducing substitution of I61M. The Chi3l1 with amino acid substitution of I61M remains inactive. This result suggests that it is necessary to substitute multiple amino acids in addition to I61M to activate Chi3l1. Further studies are needed to activate Chi3l1, and it is currently under investigation.



## **PgmNr 883: Longitudinal immune and genomic monitoring reveals signatures of response and immune-related adverse events in cancer patients receiving checkpoint inhibitor therapy.**

### **Authors:**

S. Khan; M. von Itzstein; V. Popat; D. Hsieh; M. Waters; S.A. Khan; S. Barnes; R. Lu; F. Fattah; A. Moses; C. Arana; Y. Gloria-McCutchen; J. Saltaraski; Q. Li; Y. Xie; E. Wakeland; D.E. Gerber

[View Session](#) [Add to Schedule](#)

**Affiliation:** UT Southwestern Medical Center, Dallas, Texas.

---

Despite the remarkable success of immune checkpoint inhibitor (ICI) therapy, a significant number of patients develop severe and unpredictable immune-related adverse events (irAEs) affecting a wide variety of organs. Concerns over irAE have led to the exclusion of patients with autoimmune disease from ICI clinical trials. Role of host genetic and immune factors in mediating irAEs remain unclear and it is not clear if the manifestations of irAEs is associated with response to therapy. Here, we use multi-faceted approach to characterize changes in host immune system in 250 patients receiving ICI therapy at baseline and post immunotherapy. In addition, we assessed genetic predisposition to autoimmunity in these patients using the Illumina GSA SNP array and via targeted resequencing of over 150 immunoregulatory loci including the HLA region. In this meeting, we will present our initial genetic data, analysis of serum cytokine and autoantibody profiles and RNA sequencing on peripheral blood mononuclear cells (PBMCs) at baseline and post immunotherapy in patients with and without irAEs. Our preliminary findings suggest that patients who developed irAEs have lower baseline levels and greater post-treatment increases in key interferon gamma inducible cytokine/chemokine levels. Our preliminary autoantibody profiling data demonstrates that pre-existing quiescent humoral autoimmunity correlates with irAEs. We will present immune and genetic correlates of irAEs and response to therapy identified in our preliminary analysis. We are currently generating and integrating cytokine, autoantibody, RNA-seq data and genetic data and intersecting with clinical profiles of 250 patients to gain insight in to the mechanisms underlying irAE and to identify biomarker signatures predictive of irAEs and/or response. These findings may ultimately help identify high-risk patients, customize therapy, tailor monitoring, expand use of immunotherapy and prevent toxicities.

## PgmNr 884: Genetic polymorphism studies in alcoholic pancreatitis.

### Authors:

V. Aaren; G. Sudhakar

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Human Genetics, Andhra University, Visakhapatnam, Andhra Pradesh, India

---

Chronic alcohol consumption is considered as a major cause of acute and chronic pancreatitis in many populations. To determine the genetic susceptibility to alcoholic pancreatitis (AP) in South Indian populations, 250 cases with alcoholic pancreatitis (AP), 250 alcoholic controls (AC) and an additional 320 healthy ethnic and sex-matched controls were recruited. Five of the previously associated genetic variations of exon 3, Arg47His G/A of *ADH2/ADH1B*; exon 8, Ile349Val G/A of *ADH3/ADH1C*; intron 6, T7678A of *CYP2E1*; *GSTM1*; and *GSTT1* genes were genotyped by PCR based Restriction Fragment Length Polymorphism (PCR-RFLP) analysis of blood DNA samples using gene and locus specific oligos. A detailed history of each case including the age, alcohol consumption (quantity and type of beverage), smoking, tobacco intake and other lifestyle factors were recorded. Gene polymorphisms were compared between alcoholic pancreatitis subjects (AP) vs alcoholic controls (AC) and alcoholic pancreatitis (AP) vs healthy controls (HC). Genotypic frequencies were calculated and odds ratios (OR) were determined using 2 × 2 contingency tables with 95% confidence intervals. Statistical significance for Odds Ratio (OR) was calculated by chi-square test using SPSS Version 16 software. Fisher's exact test was used when expected cell frequency was less than five. The association of the gene polymorphisms in patients and controls was examined using OR and Chi-square tests with CI 65%,  $P < 0.05$  was considered as significant. A significant association between *ADH3 Ile349Val* and AP when compared to the AC controls ( $P = 0.005$ ) and HC controls ( $P = 0.01$ ). No association was observed between markers *CYP2E1* T7678A compared to AP and controls ( $P = 0.42736$ ). Moreover, *GSTM1* ( $P = 0.7208$  and  $P = 0.42034$ ), and *GSTT1* ( $P = 0.3962$  and  $P = 0.20326$ ) genotypes were also not statistically associated with AP when compared to AC and HC controls. Similarly, lack of association between AP and *ADH2 / ADH1B* Arg47His gene polymorphisms were observed. The present data on the causes of pancreatitis may support more effective disease management, provide hope for a potential cure and aid in developing strategies for the prevention or treatment of alcoholic pancreatitis. We further conclude that a strong association with *ADH3* gene that might have a major influence on susceptibility to pancreatic cancer in south Indians, however future studies with other Indian populations are needed.

## PgmNr 885: Identify breast cancer risk-associated lncRNAs that play a role in DNA damage repair.

### Authors:

N. Abdou <sup>1,2</sup>; M. Moradi Marjaneh <sup>1</sup>; H. Sivakumaran <sup>1</sup>; K. Hillman <sup>1</sup>; S. Kaufmann <sup>1</sup>; L. Fachal <sup>3</sup>; A. Dunning <sup>3</sup>; S. Edwards <sup>1</sup>; J. French <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Queensland Institute of Medical Research Berghofer, Herston, QLD, Australia; 2) Faculty of Medicine, University of Queensland, Brisbane, Australia.; 3) Department of Public Health and Primary Care, University of Cambridge, UK

---

Genome-wide association studies (GWAS) have identified 196 independent signals that predispose to breast cancer. The vast majority of risk variants within these signals are located in intronic or intergenic regions. This suggests that distal regulatory elements such as transcriptional enhancers and/or noncoding RNAs contribute significantly to the risk of developing breast cancer. Many breast cancer susceptibility genes have been linked to DNA double-strand break (DSB) repair, where DSBs lesions evoke genomic instability and cellular lethality if not repaired. Furthermore, long noncoding RNA (lncRNA)-directed regulation of DNA damage repair pathways has been recently reported for different cancer types. Here, we use target RNA-sequencing to identify and catalogue all lncRNAs expressed from breast cancer risk signals that are induced in response to DNA damage. MCF7 breast cancer cells were treated with single ionizing irradiation (IR) dose to induce DSBs, and then RNA collected at 1, 6 and 24hr post-IR. De novo assembly of the sequencing reads identified 280 lncRNAs significantly induced at 1hr and 6hr post-IR timepoints (RPKM  $\geq$  0.5; FDR < 0.05), 166 of which are not currently annotated in existing databases. One of these lncRNAs called XLOC156282 is transcribed from the 4q21 breast cancer risk signal, which is associated with risk of developing estrogen receptor (ER)-positive. XLOC156282 is a two-exon, 4.6 kb nuclear transcript that is induced at 1 hr post-IR in normal breast cells and ER-positive breast cancer cells. XLOC156282 is antisense to ABRAXAS1, which encodes the BRCA1-A Complex Subunit Abraxas 1, a known DSB repair protein. To determine the function of XLOC156282, we used locked nucleic acid (LNA<sup>™</sup>) technology to silence XLOC156282 in MCF7 cells. Quantitative PCR confirmed that XLOC156282 knockdown resulted in reduced MRPS18C expression, but does not impact ABRAXAS1 mRNA levels or alternative splicing. MRPS18C encodes the Mitochondrial Ribosomal Protein S18C protein, which assists in protein synthesis within the mitochondrion, but has no reported role in breast cancer. Further studies are required to determine the mechanism(s) by which this XLOC156282 alters MRPS18C expression and how this potentially contributes to breast cancer development.

## **PgmNr 886: Heritable and somatic variation drive gene expression changes associated with survival in multiple myeloma.**

### **Authors:**

L. Bui; H. Natri; B. Argente; J. Keats; N. Banovich

[View Session](#) [Add to Schedule](#)

**Affiliation:** Translational Genomics Research Institute, Phoenix, AZ.

---

Multiple myeloma (MM), the second most common hematological cancer, has a poor prognosis and a 5-year survival of 50.7%. While new therapies have nearly doubled the survival rate, most patients still experience a relapse. The CoMMpass study, developed by The Multiple Myeloma Research Foundation, is a longitudinal clinical trial aimed at accelerating the discovery of more targeted treatments for MM. Clinical parameters and tumor specimens are collected from each of 1,147 patients at baseline and through out the 8-year period. Whole Genome, Exome and Transcriptome sequencing were performed on each tumor specimen to identify genetic determinants of clinical outcomes. This study has uncovered several somatic mutations and copy number alterations associated with tumor progression and response to therapy, and a novel classification of MM into 12 distinct subtypes based on gene expression. However, the contribution of germline genetic variation to gene expression changes and MM outcome remains poorly understood.

Genome-wide association studies have identified germline variants associated with MM risk, indicating inherited genetic susceptibility. MM exhibits a disparity in occurrence and mortality between sexes and ethnicities, with men and African Americans being at a higher risk than women or those of European ancestry. To better understand the genetic and biological basis of MM predisposition, we utilize a functional genomics approach to examine non-coding germline and somatic effects on gene expression and MM outcome. Data from 607 CoMMpass participants with available WGS and RNAseq from normal blood and baseline tumor specimens were used to map eQTLs, resulting in 3,929 genes with at least one cis-acting eQTL. We identified eQTLs modulating survival in MM patients and several regulatory variants with differential effects on tumor gene expression between sexes. Several eQTLs overlap MM GWAS risk loci, providing regulatory mechanisms connecting these loci to MM. To identify eQTLs overlapping open chromatin areas we performed ATACseq on male (3) and female (3) MM cell lines, and identified 371 eQTLs on accessible DNA regions. To functionally validate the eQTLs we performed CRISPR activation/interference in MM cell lines. Using integrated functional genomics approaches combined with clinical outcomes we have identified new regulatory regions associated with survival in MM, supporting the development of personalized medicine approaches for better treatments.

## PgmNr 887: Network pathways leading to discover biomarker candidates based on genomic profiles of racial differences in triple negative breast cancer.

### Authors:

C.-S. Chang <sup>1,2</sup>; E. Kitamura <sup>2</sup>; L. Hawthorn <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics Resource, Georgia Cancer Center, Augusta University, GA; 2) Georgia Cancer Center, Augusta University, Augusta, GA

---

### ABSTRACT

**Objective:** Triple negative breast cancer (TNBC) is more prevalent in African Americans (AAs), has a more aggressive clinical course including a higher mortality rate and increased occurrences of metastases. This study was designed to determine if racial differences at the molecular level might explain the more aggressive phenotype in AAs.

**Method:** Mutation profiling was performed on 51 AA and 77 CA tumor/normal pairs using Exome-Seq analysis and the GATK pipeline. Transcript expression analysis was performed on 35AA and 37CA samples and analyzed using the Tuxedo Suite tools. The importance coefficient ranking of genes was performed using the Machine Learning model, Support Vector Machine (SVM) to detect differences between the two racial groups. The novel network analysis was conducted for further Hub gene candidates applicable for biomarker selection.

**Results:** Genes with high frequency mutation rates such as MUC4 and TP53 were common to both racial populations, however genes that were less frequently mutated differed between the races suggesting that those cause the more aggressive nature of TNBC in AA women. JAK-Stat signaling was unique to the AA and PTEN and mTOR were unique to the CA profiles. SVM and co-expressed and interacted network analysis identified Interferon and Cytokine signaling pathways as more highly expressed in AA patients.

**Conclusion:** Many pathways identified by the mutational profiles were predicted to be down-regulated by the transcript expression profiles. Genes in ILK signaling pathway commonly found in both races can be examined as genetic targets to treat TNBC. Genes in the respective pathways, either unique to AA or unique to CA can be tested as genetic targets by race. The discovered Hub genes in the networks can be associated with a biomarker identification to differentiate the phenotypes of AA and CA.

## PgmNr 888: Investigating the role of ATRX in osteosarcoma progression.

### Authors:

Y. Chen <sup>1</sup>; Y. Bae <sup>2</sup>; B. Lee <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Integrative Molecular and Biomedical Sciences Program; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

Osteosarcoma (OS) is the third most prevalent pediatric cancer. Its high propensity of pulmonary metastasis makes the overall survival rate less than 30%. Despite advances in surgery and chemotherapy, the poor understanding of the genomic features has hampered the development of effective and specific treatments over the decades. OS is characterized by a high level of genomic instability including large chromosomal structure variations and epigenetic alterations. However, molecular mechanisms driving the genomic instability during OS development and progression is not well understood.

Previously, we and others analyzed whole exome/genome sequencing datasets of human OS primary tumor with matched blood samples and identified somatic inactivating mutations in *ATRX* (alpha-thalassemia / mental retardation, X-linked). *ATRX* belongs to the SWI/SNF family and plays fundamental roles in chromatin remodeling. As an epigenetic regulator, *ATRX* modulates histone H3.3 deposition, maintains chromatin configuration, resolves DNA superstructure such as G-quadruplex, and secures genomic integrity. Its role in genome maintenance and the tendency of recurrent mutations in OS patients suggest potential participation of *ATRX* loss-of-function in inducing genomic instability and reshaping the chromatin landscape during the cancer progression.

To understand the cellular functions of *ATRX*, we knocked down *ATRX* in OS cell lines using siRNA approach. *ATRX* knockdown in 143B and MNNG cells showed increased proliferation and migration potential, and the cellular morphology became more mesenchymal-like. N-cadherin protein level was also upregulated after *ATRX* knockdown, which indicates mesenchymal-like cell property. These results show that loss of *ATRX* leads to increase proliferation and prone to become mesenchyme with elevated migratory potential, which suggests a tumor suppressive role of *ATRX* in OS pathogenesis. Currently, we are developing *ATRX* knockout OS cell lines to examine the impacts of *ATRX* deficiency toward transcriptome alteration and epigenetic modification in OS by using CRISPR-Cas9 mutagenesis. Furthermore, to elucidate the tumor suppressive role of *ATRX* in OS development, conditional *Atrx* knockout mouse models are being established.

## PgmNr 889: Identifying functional variants at chr7p14.1 affecting pancreatic cancer risk.

### Authors:

K.E. Connelly<sup>1</sup>; J. Zhong<sup>1</sup>; E. Abdolalizadeh<sup>1</sup>; D. Eiser<sup>1</sup>; J.W. Hoskins<sup>1</sup>; A. Jermusyk<sup>1</sup>; I. Collins<sup>1</sup>; B.M. Wolpin<sup>2</sup>; R.Z. Stolzenberg-Solomon<sup>3</sup>; A.P. Klein<sup>4</sup>; J. Shi<sup>3</sup>; J.P. Smith<sup>5</sup>; G.M. Petersen<sup>6</sup>; L.T. Amundadottir<sup>1</sup>; Pancreatic Cancer Cohort Consortium, Pancreatic Cancer Case-Control Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Gaithersburg, MD; 2) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA; 3) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD; 4) Johns Hopkins University, Baltimore, MD; 5) Department of Medicine, Georgetown University, Washington; 6) Mayo Clinic College of Medicine, Rochester, MN

---

Pancreatic ductal adenocarcinoma (PDAC), the third leading cause of cancer-related deaths, is often difficult to diagnose. An effort to identify common susceptibility loci for PDAC using genome-wide association studies (GWAS) has led to the identification of twenty loci for pancreatic cancer risk. One of these is at chr7p14.1 (rs12701838,  $P = 3.9 \times 10^{-9}$ ) which maps to the final intron of *SUGCT*. We used linkage disequilibrium (LD) analysis and a likelihood ratio of 1:1000 to identify 18 candidate functional variants. Expression quantitative trait locus (eQTL) analysis of the GTEx v7 pancreas data set demonstrated an association between 7p14.1 and *INHBA* (rs10279715, normalized effect score=0.29,  $P = 1.9 \times 10^{-5}$ ) with lower *INHBA* expression correlated to the risk genotype. *INHBA*, which is approximately 1 Mb downstream of rs10279715, is important for endocrine cell differentiation and insulin secretion. Furthermore, our transcriptome-wide analysis study (TWAS) identified *INHBA* as a gene-trait association signal for pancreatic cancer ( $Z = -5.11$ ,  $P = 3.2 \times 10^{-7}$ ). After conditioning on the lead GWAS SNP rs12701838, the *INHBA* TWAS signal is no longer significant ( $P = 0.81$ ). Co-localization analysis with the Coloc R package suggests the GWAS and eQTL signals share a single causal variant ( $PP = 0.74$ ). To identify functional variants lying within regulatory regions identified in pancreatic samples, we are using CRISPR in pancreatic cancer cell lines to delete small genomic regions (~250 bp) each harboring a single variant and assessing the effects on *INHBA* and *INHBA-AS1* gene expression. Candidate regions will subsequently be examined for allele-specific effects with electrophoresis mobility shift assays, luciferase assays, and proteomics. Future work aims to elucidate the role of *INHBA* and the functional variant(s) in PDAC risk and pathogenesis.



## PgmNr 890: Evidence for an ancient *BRCA1* pathogenic variant and a novel *BRCA2* pathogenic variant in inherited breast cancer patients from Senegal.

### Authors:

J.P.D. Diop <sup>1</sup>; V. Bourdon-Huguenin <sup>4</sup>; A. Dem <sup>5</sup>; D. Diouf <sup>5</sup>; S.M.K. Gueye <sup>2</sup>; S.A. Ba <sup>1</sup>; Y. Dia <sup>1</sup>; R. Ndiaye <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Cytology, Cytogenetics and Reproductive Biology, Aristide Le Dantec Hospital, Dakar, Senegal; 2) Faculty of Medicine, Pharmacy and Odontology, University Cheikh Anta Diop, Dakar, Senegal; 3) African Center of Excellence for Mother and Child Health (CEA-SAMEF); 4) Laboratory of Molecular Oncogenetics, Paoli-Calmette Institute, Marseille, France; 5) Joliot Curie Institute, Aristide Le Dantec Hospital, Dakar, Senegal

**Background:** *BRCA1* and *BRCA2* are the most incriminated genes in inherited breast/ovarian cancers. Several pathogenic variants of these genes conferring genetic predisposition have been described in different populations but rarely in subsaharan Africa. The objectives of this study were to identify pathogenic variants of the *BRCA* genes involved in hereditary breast cancer in Senegal and to search for a founder effect.

**Methods:** We recruited after free informed consent, 27 unrelated index cases diagnosed with breast cancer and each having a family history, and a control population of 48 healthy women and 48 women diagnosed with sporadic breast cancer. All *BRCA1* and *BRCA2* exons were Sanger sequenced for the first 15 index cases recruited. For the following 12, healthy relatives and controls, the recurrent pathogenic variant of *BRCA1* identified in the first group was genotyped by PCR. To search for a founder effect of the recurrent pathogenic variant of *BRCA1*, haplotype analysis of 7 microsatellites encompassing the gene was performed in 10 selected index cases with the pathogenic variant and 15 unrelated healthy controls. Mutation age was estimated using the following equation:  $G = \log/\log (1 - q)$ .

**Results:** Mutation screening identified a recurrent duplication of 10 nucleotides c.815\_824dupAGCCATGTGG, (p.Thr276Alafs) of *BRCA1*, in 15 index cases (detection rate 55.5%). Haplotype analysis showed a shared haplotype encompassing approximately 400 kb. Estimation of the age of the pathogenic variant suggested that it occurred approximately 1400 years ago and confirm a West African origin. The variant was also detected in healthy relatives of selected families. The genotyping method is now available as a first screening test in women at risk. For *BRCA2* gene, a novel pathogenic variant, c.5219 T >133 G; p.(Leu1740Ter) was identified in a proband from a consanguineous family along with three sisters and one daughter of the index case.

**Conclusion:** Our study identified a founder pathogenic variant of *BRCA1* predisposing to breast cancer and enabled the establishment of an affordable genetic test as a mean of prevention for Senegalese women at risk. Oncogenetic counselling and mutation screening for other causal genes should now be implemented in Senegal.

**Keywords:** Hereditary Breast Cancer, *BRCA1*, *BRCA2*, pathogenic variant, Senegal.



## PgmNr 891: Associations of *BRCA1/2* mutations with Chinese breast/ovarian cancer patients.

### Authors:

M. He <sup>1,2</sup>; M. Yan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Simcere Diagnostics Co., Ltd., Nanjing, 210042, China; 2) State Key Laboratory of Translational Medicine and Innovative Drug Development, Nanjing, 210042, China

**Background:** *BRCA1* and *BRCA2* are well-characterized predisposition genes associated with breast and/or ovarian cancers. However, the majority of the studies has been conducted in the western populations. The genomic impacts on Chinese breast and/or ovarian cancer patients require further assessments of the *BRCA* genes and identifications of potential novel ones.

**Patients and methods:** In this study, 276 patients with clinical diagnosis consist of 139 breast cancer (BC; including 1 male; age (mean  $\pm$  standard deviation):  $48 \pm 11$ ) and 139 ovarian cancer (OC; including 1 male; age:  $54 \pm 10$ ) while two patients are both BCs and OCs. In addition, 402 individuals (male to female ratio: 1:1, age:  $43 \pm 13$ ) are in-house controls with no BC/OC or other cancers examined at the time that the study was initiated. Whole-exome sequencing (WES) was conducted and the association between BCs/OCs and the in-house controls was assessed. The other East Asians (total 7,209) in Genome Aggregation Database (gnomAD) was used as an additional set of controls in this study. Cohort allelic sums test (CAST) was used for the comparisons between cases and controls at the gene level. Fisher's exact test was used to estimate the significances of the single variants between cases and controls. All P values were adjusted by the Benjamini-Hochberg false-discovery rate (FDR) multiple testing correction.

**Results:** Mutations in both *BRCA1* and *BRCA2* were dramatically enriched in the cases. Compared to the in-house controls, *BRCA1* and *BRCA2* showed gene-based significance, with the FDR-adjusted p-values being 0.006 and 0.003 in BC, 0 and 0.008 in OC, and 0 and 0.0003 in both BC and OC cohorts, respectively. While compared to the other East Asians in the gnomAD, the adjusted p-values of *BRCA1* and *BRCA2* were  $5.61\text{e-}5$  and  $1.89\text{e-}5$  in BC,  $4.18\text{e-}15$  and  $2.77\text{e-}4$  in OC, and  $5.83\text{e-}15$  and  $7.65\text{e-}9$  in both BC and OC. Additionally, *PRKRA* also showed high significance while compared to the other East Asians in the gnomAD, with adjusted p-values of  $6.16\text{e-}15$ ,  $4.18\text{e-}15$  and  $5.83\text{e-}15$  in the BC, OC, BC and OC subsets, respectively.

**Conclusions:** *BRCA1/2* were confirmed to be associated with the occurrence of BC/OC in the Chinese patients in contrast to either in-house controls or the other East Asians in the gnomAD. Furthermore, *PRKRA* might play a role in the onset of BC/OC in Chinese descent. Further work is required to expand the patient size to provide more solid confirmations.

## PgmNr 892: Classifying variants of unknown significance in the BRCA1-PALB2 binding interface.

**Authors:**

C.A. Hurd; M.D. Stewart

[View Session](#) [Add to Schedule](#)

**Affiliation:** Biology, Texas Christian University, Fort Worth, Texas.

---

BRCA1 (Breast Cancer associated protein 1) and PALB2 (Partner and Localizer of BRCA2) are two tumor-suppressing proteins that have been shown to associate with each other in the DNA damage response to repair double-stranded breaks through homologous recombination. Mutations in both *BRCA1* and *PALB2* have been shown to be associated with increased breast and ovarian cancer risk in accordance with their role as tumor suppressors. Genetic databases currently have lists of variants of unknown significance (VUS) for these two corresponding genes gathered from previous studies. Our research aims to establish an *in vitro* biochemical model to predict the effect of VUS in the BRCA1 and PALB2 interaction domains. NMR titrations and isothermal titration calorimetry show that our minimized wild-type interaction domain constructs interact *in vitro*. We are using site-directed mutagenesis to recreate known VUS in these minimized constructs. Through comparison with wild-type binding affinities and affinity of known detrimental mutant proteins, we can establish the effect of the VUS on the crucial interaction. We will present our *in vitro* model system and ongoing data collection from *BRCA1* and *PALB2* variants of unknown significance.

## PgmNr 893: Mutational analysis of chromosome 16q in wilms tumors.

**Authors:**

E. Kitamura; C.-S. Chang; J.K. Cowell; L. Hawthorn

[View Session](#) [Add to Schedule](#)

**Affiliation:** Georgia Cancer Center, Augusta University, Augusta, GA

---

**Objective:** Wilms tumor (WT) is the most common childhood renal tumor and affects approximately 1 in 10,000 children in Europe and North America. Chromosome 16q deletion (del) or loss of heterozygosity (LOH) has been associated with recurrence and adverse prognosis in WT. Our study aims to identify genes implicated in the adverse event on chromosome 16q.

**Method:** We performed exome-sequencing on 8 WTs with chr16q del/LOH from our legacy collection. In addition, we analyzed 42 exome-seq data of WT patients from the TARGET database. The fastq data were aligned by BWA v.0.6.1 software and variants were detected using haplotypcaller. The detected variants were annotated by VEP and filtered using custom cut-offs to search for somatic mutations. We compared the mutational profiles of total 50 WTs between with and without chr16q del/LOH. Also we searched signaling pathways in association with the mutational profiles in WT using Ingenuity Pathway Analysis.

**Results:** In this study, a total of 24,271 variants predicted as damaging and/or deleterious were found in 10,789 genes. By comparing these variants between WT with and without chr16q del/LOH, we identified 131 homozygous somatic variants with putative loss-of-function in 91 genes on chr16q. Pathway analysis showed that cilia-related signaling pathways such as sonic Hedgehog (SHh), Wnt signaling and Notch signaling are highly mutated in WTs with chr16q del/LOH.

**Conclusion:** The results suggest that the signaling pathways are involved in poor prognosis in WT, providing a possibility of using key genes in the pathways as biomarkers for diagnosis and treatment of aggressive WT and investigating mechanism leading to adverse prognosis.

## PgmNr 894: SVFX: A machine learning framework to identify pathogenic structural variations in disease studies.

### Authors:

S. Kumar <sup>1,2</sup>; A. Harmanci <sup>3</sup>; J. Vytheeswaran <sup>4</sup>; M. Gerstein <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut.; 2) Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut; 3) Center for Precision Health, School of Biomedical Informatics, University of Texas Health Sciences Center, Houston, Texas; 4) California Institute of Technology, Pasadena, California

---

A rapid decline in sequencing cost has enabled large-scale genome sequencing studies feasible. One of the fundamental goals of these studies is to catalog all variants of functional consequence. Numerous methods and tools are being developed to interpret point mutations, small insertion, and deletions. However, there is a lack of methods for understanding the functional consequence of genomic structural variations. Structural variations play a crucial role in various diseases by altering the sequence and three-dimensional structure of the genome. In this work, we build a machine learning based workflow that leverages genomics, epigenomics, and conservation metric to assign a pathogenicity score to somatic and germline structural variations in various diseases. We apply this workflow on structural variants present in six different cancer types. Overall, our predictor achieves high accuracy in identifying pathogenic germline and somatic structural variations. In particular, our predicted pathogenic somatic variants are enriched among known cancer genes and many cancer-related pathways, including Wnt signaling, Ras signaling, DNA repair, and ubiquitin-mediated proteolysis. Finally, we note that our framework is highly flexible can be easily extended to identify pathogenic germline structural variations in other disease-specific studies.

## PgmNr 895: The Reactome pathway knowledgebase: Linking human genetic variation to disease.

### Authors:

L. Matthews <sup>1</sup>; G. Wu <sup>2</sup>; H. Hermjakob <sup>3</sup>; L. Stein <sup>4</sup>; P. D'Eustachio <sup>1</sup>; Reactome Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NYU Langone Medical Center, New York, NY USA; 2) Oregon Health & Science University, Portland, OR, USA; 3) EMBL-EBI, Hinxton, UK; 4) OICR, Toronto, ON, Canada

---

Human reaction space is a network of processes whose extensive connectivity undoubtedly contributes to maintenance of homeostasis in the face of environmental stresses. This connectivity also undoubtedly contributes to the difficulty of predicting the effects of a genetic variation on a whole-organism phenotype given its effects on a protein's function at a molecular level.

Pathway databases like Reactome enable the comprehensive, detailed annotation of reaction space in the framework of a data model that supports the organization and analysis of molecular-, cell-, and tissue-level data to investigate effects of changes in the functions of one or more gene products, to provide better genotype-phenotype predictions. Reactome is an open-source, expert-curated, peer-reviewed knowledgebase. Its holdings now comprise 12,505 human reactions organized into 2,272 pathways involving 11,109 proteins encoded by 10,840 different human genes, 2,042 small molecules (of which 185 are drugs), and 12,206 complexes. The database provides the molecular details of a wide range of cellular processes ranging from metabolism and signal transduction to development and neuronal systems. These annotations are supported by 30,027 literature references. The roles of variant forms of 305 proteins, both germline and somatically arising, have been annotated into disease-variant forms of 968 reactions and additional reactions that capture the effects of small molecule drugs on these disease processes. Variants are chosen for annotation based on guidelines and data from ClinGen and Cosmic.

Reactome thus provides a powerful tool for analyzing expression data to identify the specific parts of reaction space affected by a disease process or environmental stress, and is a useful starting point for investigating possible propagating effects of the loss of activity of single genes or gene combinations.

## PgmNr 896: Visualizing the functional landscape of BRCA1 missense variants.

### Authors:

A.N. Monteiro <sup>1</sup>; P.C. Lyra, Jr. <sup>2</sup>; T.C. Nepomuceno <sup>1,3</sup>; M. Vreeswijk <sup>4</sup>; F.J. Couch <sup>5</sup>; M.A. Carvalho <sup>3</sup>; L.B. Rangel <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Moffitt Research Center (MRC3WEST), 12902 Magnolia Drive, Tampa, Florida.; 2) Biotechnology/RENORBIO Program, Federal University of Espírito Santo, Vitória, ES, Brazil; 3) Instituto Nacional de Câncer, Programa de Pesquisa Clínica, Rio de Janeiro, Brazil; 4) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 5) Department of Laboratory Medicine, Mayo Clinic, Rochester, MN, USA

---

The last few years has seen widespread genetic testing for germline variants in *BRCA1* for assessment of breast and ovarian cancer risk. However, this trend led to an increase in findings of *BRCA1* variants of uncertain clinical significance (VUS). Carriers of VUS cannot be told whether they are at increased risk or not. While for those with a strong family history of cancer risk assessment can be based on clinical factors, those without a family history, ~½ of all carriers, have no such alternative. Thus, the determination of pathogenicity associated with a VUS is a critical barrier to the implementation of precision medicine.

Classification of missense variants, the largest class of *BRCA1* VUS, can be based on a multifactorial statistical model that incorporates data on family history, co-segregation, and co-occurrence with a known pathogenic variant in the same gene. Currently, functional data is used as supporting evidence for classification but are not yet integrated in these multifactorial statistical models. We expect that functional data will be critical for the classification of rare variants. Previously we compiled all functional data for missense variants in *BRCA1* into an interactive web tool called *BRCA1* Circos (<https://research.nhgri.nih.gov/bic/circos/>). We have now updated and greatly expanded our comprehensive database of functional assays for *BRCA1* missense variants.

We identified 35 published articles describing 129 sets of functional experiments leading to 3,427 results assessing the functional impact of missense variants on *BRCA1* functions. Altogether, 1,005 unique (out of a possible 11,009) missense variants were tested at least once. To harmonize results we used the thresholds and classifications described by the original authors and converted all results into “functional impact” or “no functional impact”. We used a panel of 298 known reference variants (252 benign or likely benign and 45 pathogenic or likely pathogenic) to calculate sensitivity and specificity for every assay. There were 46 and 242 variants (out of 463 tested more than once) that scored as “functional impact” or “no functional impact” consistently in every assay tested, respectively. This work is the first step to harnessed functional data on missense variants of *BRCA1* for clinical annotation. Generating a comprehensive picture of *BRCA1* variants using validated assays can accelerate the classification of VUS.

## PgmNr 897: Population-wide screening for germline variants of hereditary cancer genes in 12K unselected colorectal cancer and 27K controls in Japan.

### Authors:

H. Nakagawa <sup>1</sup>; X. Liu <sup>1</sup>; Y. Iwasaki <sup>1</sup>; M. Fujita <sup>1</sup>; K. Akagi <sup>2</sup>; M. Hirata <sup>3</sup>; Y. Murakami <sup>3</sup>; K. Matsuda <sup>4</sup>; M. Kubo <sup>1</sup>; Y. Momozawa <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) RIKEN IMS, Yokohama, Japan; 2) Saitama Cancer Center, Saitama, Japan; 3) Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 4) Graduate School of Frontier Science, The University of Tokyo, Tokyo, Japan

---

**AIM.** Colorectal cancer (CRC) is one of the most common cancers in Western countries and Japan. A few % of CRCs can be attributed to recognizable hereditary germline variants of susceptibility genes, predominantly the DNA mismatch repair genes (Lynch syndrome), and universal genetic screening has been proposed for CRCs. Now there are a number of variants of these genes aggregated and annotated in several databases, but still most are VUS and sometimes ethnic-specific. Hence it is require to have a large-scale variant dataset in a specific ethnic group.

**METHODS.** We here sequenced coding regions of eleven genes (*MSH2*, *MSH6*, *MLH1*, *PMS2*, *EPCAM*, *APC*, *MUTYH*, *SMAD4*, *PTEN*, *STK11*, and *TP53*) in 12,347 unselected CRC patients and 27,706 controls (60~yo) without any cancer and cancer family history by our high-throughput method (*Nat Commun* 9:4083, 2018). The samples were collected by Biobank Japan (BBJ) from locations all over Japan. We assigned their clinical significances using ACMG and Genomics and the Association for Molecular Pathology guidelines as well as ClinVar.

**RESULTS.** We identified 1,900 variants in the eleven genes and annotated them as 210 pathogenic variants, 261 benign ones, and 1,429 VUS, of which 30% were not registered in ClinVar and novel. Some variants annotated as pathogenic or VUS in ClinVar were observed frequently in the Japanese control population, such as p.Leu582Val and p.Thr151Thr of *MLH1*, p.Lys1358fs of *MSH6*, and p.Lys845Glu of *MSH2*. In total, 1.5% of the unselected CRC patients and 0.014% of the controls had a pathogenic variant. The pathogenic variants of *MSH2* (OR =10), *MLH1* (OR =18), *MSH6* (OR =6), and *APC* (OR=47) were significantly associated with CRC development in Japanese population ( $P$ -value  $<10^{-4}$ ). One recurrent mutation of *MSH6* (p.Phe1088fs) was found in Japanese CRCs. All of 21 truncating or pathogenic variants of *APC* in this CRC cohort were located not in the mutation cluster region, indicating their attenuated phenotype. Furthermore, from the genome-wide SNP chip data, we called CNVs or SVs of the target genes and confirmed SVs of *MSH2/EPCAM*, *MLH1*, and *APC* by MLPA (n=23).

**COCLUSION.** In unselected Japanese CRC cases, pathogenic variants or CNVs of *MSH2*, *MLH1*, *MSH6* and *APC* were observed in 1.7%. This is the largest study, to the best of our knowledge, of CRC heredity and would contribute to development of guidelines for genetic testing and variant interpretation for familial and sporadic CRCs.

## **PgmNr 898: Functional classification of selected exonic splicing variants occurring outside of canonical 5' and 3' exon boundaries in cancer-predisposing genes.**

### **Authors:**

P. Nix; M. Braithwaite; E. Goosen; E. Mundt; S. Manley; K. Bowles; N. Singh

[View Session](#) [Add to Schedule](#)

**Affiliation:** Myriad Genetics Laboratories, Inc. , Salt Lake City , UT.

---

Sequence variants that affect mRNA splicing account for 10-15% of disease-causing mutations in cancer-predisposing genes. A majority of these occur at the 5' and 3' exon-intron boundaries, and the potential impact on splicing often can be predicted based on known sequence conservation at splice acceptors and splice donors. Sequence variants located outside of canonical splice sites also may alter mRNA processing through the use of cryptic splice sites or through the alteration of exonic splicing enhancers and silencers. The effects of these types of variants on mRNA processing are more difficult to predict and require experimental verification to determine splicing outcomes for accurate variant classification. In this IRB-approved study, we assessed the effects of eight variants of uncertain significance (VUS) identified during clinical testing of cancer-predisposing genes. RNA was extracted from blood of VUS carriers and reverse transcribed for analysis of aberrant splicing. Seven of the eight variants analyzed occur in exonic regions, outside of exon-intron boundaries. Of these, five variants were shown to produce a complete splice defect and were upgraded to likely pathogenic mutations. One variant, *BRCA2* c.68-2A>G, occurs at a consensus splice acceptor and could be classified as likely pathogenic based on splicing predictions. Our RNA results raise the possibility that a functional transcript could be produced, and therefore the classification remains VUS. Investment in RNA studies is integral to a robust variant classification program in hereditary cancer risk testing. Our laboratory RNA analyses continue to provide valuable information that may ultimately alter patient medical management, with implications for their family members.



## PgmNr 899: Flow variant assays of the double strand break repair pathway improve risk classification for women receiving cancer panel sequencing.

### Authors:

H. Ostrer<sup>1</sup>; S. Klugman<sup>2</sup>; L. Tesfa<sup>1</sup>; L. Yang<sup>3</sup>; Y. Shao<sup>3</sup>; S. Dolan<sup>2</sup>; E. Suskin<sup>2</sup>; L. Weingarten<sup>2</sup>; K. Aaron<sup>2</sup>; J. Keating<sup>2</sup>; E. Scavone<sup>2</sup>; E. Goldberg<sup>2</sup>; J. Loke<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pathology, Albert Einstein College of Medicine, Bronx, New York.; 2) OB-GYN, Montefiore Medical Center, Bronx, NY; 3) Population Sciences, New York University School of Medicine, New York, NY

---

About 10% of breast cancer is thought to be hereditary and caused by mutations in genes known to increase risk of cancer, especially those in the double strand break (DSB) repair pathway. Hereditary cancer gene sequencing panels can identify mutations in a fraction of high-risk patients and families, i.e., those with multiple affected members and earlier age of diagnosis. The remainder of patients has variants of uncertain significance (VUS) or negative results and, thus, their risk is not properly assessed. This issue of unclear risk can be mitigated using flow variant assays (FVAs), a functional approach for identifying defects in the DSB repair pathway. Defects in the repair process are identified in cells treated with radiomimetic agents using flow cytometry. To account for heterogeneity in individual assays, results are combined using logistic regression to calculate risk probability scores. Based on sensitivity of 0.91 and specificity of 0.99 for risk probability scores, Bayesian analysis showed that all VUS would be reassigned to the pathogenic ( $>0.99$ ) or likely pathogenic categories ( $>0.90$ ), if risk probability scores were high. Conversely if risk probability scores were low, VUS would be considered benign ( $<0.01$ ) or likely benign ( $<0.10$ ). Within a cohort of 119 women seeking genetic counseling and testing based on personal and/or family history that was considered to be high-risk, 16 (13%) were found to have mutations and 33 (28%) were found to have VUS. Among these women, 7 (6%) were reassigned to the population risk category by virtue of not inheriting a familial mutation. Among a subgroup of 43, including 24 with VUS, 15 (34%) were reassigned to the low-risk category based on their low-risk FVA probability scores and 28 (65%) were confirmed as high-risk based on their probability scores. A subject with a mutation in the APC gene also had a high-risk probability score, suggesting crosstalk with the DSB repair pathway or misreporting of the mutation as pathogenic. Thus, FVAs represent a highly accurate adjunct for identifying those with germline DSB repair defects that were missed by panel sequencing and for reassigning risk to population risk for those that had VUS or negative test results.

## PgmNr 900: Partitioned heritability and functional enrichment reveal ovarian cancer risk variants in histotype-specific enhancers that disrupt transcription factor binding sites.

### Authors:

P.-C. Peng<sup>1</sup>; S.G. Coetzee<sup>1</sup>; J. Tyrer<sup>2</sup>; A.L. Reyes<sup>1</sup>; J. Plummer<sup>1</sup>; R.I. Corona<sup>1,3</sup>; K. Lawrenson<sup>3</sup>; J.-H. Seo<sup>4</sup>; M.L. Freedman<sup>4</sup>; B.P. Berman<sup>1</sup>; P. Pharoah<sup>2</sup>; D.J. Hazelett<sup>1</sup>; S.A. Gayther<sup>1</sup>; M.R. Jones<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 2) CR-UK Department of Oncology, University of Cambridge, Strangeways Research Laboratory Cambridge, UK; 3) Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 4) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

---

Genome wide association studies (GWAS) have identified 41 regions associated with risk of epithelial ovarian cancer (EOC). Given that the vast majority of these variants lie in the non-coding genome, and that regulatory landscapes are highly tissue specific, we hypothesized that EOC risk alleles disrupt regulatory regions and function through their interaction with epigenomic features in cell types relevant to EOC.

We applied computational tools to identify functional mechanisms at GWAS risk loci for EOC. We first estimated the heritability explained by common SNPs for each EOC histotype and found that the current known risk loci of high grade serous ovarian cancer (HGSOC) account for approximately 6% of narrow sense heritability ( $h_g^2$ ). Moreover, we partitioned SNP-heritability across non-cell-type-specific functional categories and observed a significant contribution of regulatory elements to EOC heritability, e.g. Promoter ( $p = 0.016$ ), and Super-enhancer ( $p = 0.02$ ). To understand how risk variants residing in regulatory elements affect cancer development through the epigenomic landscape, we collected epigenomic datasets for a total of 121 cell and tissue types, from public resources (Roadmap Epigenomics and ENCODE) and in-house generated ChIP-seq. We identified significant enrichment of EOC risk variants in cell and tissue specific active regulatory elements marked by H3K27Ac. Histotype specific analysis of credible causal SNPs identified histotype-specific patterns of enrichment in regulatory elements. Credible causal SNPs for HGSOC were enriched in H3K27Ac only in HGSOC tumors, and depleted in ovarian surface epithelium, demonstrating that germline risk variants for EOC are located within regulatory elements that are active in tumor, rather than precursor normal tissues, but also add to the existing evidence that ovarian surface epithelium are unlikely the cell of origin for HGSOC. We predicted the effect of credible causal variants within H3K27Ac regions in EOC related cell types on transcription factor binding sites. The most frequently broken motif was REST, which is disrupted by 19 SNPs and has been implicated as both a tumor suppressor and an oncogene.

Overall, these systematic analyses guide comprehensive interpretation of EOC risk variants. Functional annotation with epigenomic data from relevant cell types can identify putative causal

regulatory elements and transcription factor binding sites that are disrupted by risk variants for EOC.

## PgmNr 901: Genome-wide analysis of ovarian super-enhancer associated lncRNAs implicates *WT1-AS* in inherited risk of high-grade serous ovarian cancer.

### Authors:

B.M. Reid <sup>1</sup>; K.C. Vavra <sup>2</sup>; R.I. Corona <sup>2</sup>; Z. Chen <sup>1</sup>; J. Seo <sup>3</sup>; X. Lin <sup>2</sup>; M. Freedman <sup>3,4</sup>; Y.A. Chen <sup>1</sup>; L. Kelemen <sup>5</sup>; A. Monteiro <sup>1</sup>; K. Nakamura <sup>1</sup>; J. Teer <sup>1</sup>; X. Yu <sup>1</sup>; E. Goode <sup>6</sup>; F. Karreth <sup>1</sup>; J.B. Permuth <sup>1</sup>; P.D.P. Pharoah <sup>7</sup>; S.A. Gayther <sup>2</sup>; K. Lawrenson <sup>2</sup>; T.A. Sellers <sup>1</sup>; Ovarian Cancer Association Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Moffitt Cancer Center, Tampa, FL; 2) Cedars-Sinai Medical Center, Los Angeles, CA; 3) Dana-Farber Cancer Institute, Boston, MA; 4) The Eli and Edythe L. Broad Institute, Cambridge, MA; 5) Medical University of South Carolina, Charleston, SC; 6) Mayo Clinic, Rochester, MN; 7) University of Cambridge, Cambridge, UK

Long non-coding RNAs (lncRNAs) are critical regulators of gene expression and have emerged as both targets and effectors of super-enhancers (SEs). SEs are large regions of active chromatin associated with particularly high levels of target gene expression and are major drivers of dysregulated gene expression in cancer. We explored SE-associated lncRNAs in epithelial ovarian cancer (EOC) susceptibility by integrating *in silico* and experimental studies with genome-wide association study (GWAS) data from over 60,000 cases and controls. SEs were mapped using H3K27ac chromatin immunoprecipitation (ChIP-seq) and transcriptomic profiling in ovarian cancer precursor cells—fallopian tube secretory and ovarian surface epithelial cells (FTSECs and OSECs, n=2 per cell type) – and four different tumor histotypes (n=5 each). A total of 5,711 SEs were identified and linked to putative target genes by correlating ChIP-seq signal with gene expression, or using gene-enhancer pairings from the GeneHancer database. 61% of SEs were associated with lncRNA expression, and 126 SE-lncRNA pairs coincided with susceptibility regions identified by GWAS ( $p < 10^{-5}$ ). We prioritized SE-lncRNA pairs at risk loci by identifying lncRNAs which were either (1) exquisitely sensitive to transcriptional inhibitors (and therefore highly likely to be SE-regulated), or (2) exhibited risk-SNP *cis*-eQTL associations in normal and/or tumor tissues. This prioritized 10 SE-lncRNA pairs including *TIPARP-AS1* and *LINC00886* (3q25.31), *CTD-2245E15.3* (5p15.33), *SNHG7* (9q34.2), *MAP3K14-AS1* and *HOXB-AS3* (17q21.32), and *TTC28-AS1* (22q12.1). These analyses also identified 13 putative novel risk regions (GWAS p-value:  $10^{-8}$ -5). Novel high-grade serous EOC risk loci included chromosome 11p13 (index SNP rs910515, OR=0.92,  $p=2.9 \times 10^{-6}$ ), where SE signal is associated with expression of *WT1-AS*, a lncRNA expected to regulate WT1, a key transcription factor in high-grade serous EOC. *WT1-AS* was downregulated after THZ1 treatment (fold-change=-2.2) and *WT1-AS* expression is positively correlated with *WT1* expression in primary tumors ( $\rho=0.67$ ,  $p < 2.2 \times 10^{-6}$ ), suggesting *WT1-AS* positively regulates *WT1* expression *in cis*. In summary, we integrated epigenomic and transcriptomic profiles of ovarian tissues with GWAS data to identify SE-lncRNA axes that may play a role in EOC risk. Functional follow-up studies to elucidate the mechanisms at these regions are ongoing.

## PgmNr 902: Ultra-deep targeted sequencing of the tumor suppressor gene *DEAR1* identifies heterogeneous mutations in ductal carcinoma *in situ*.

### Authors:

J. Reuther <sup>1,2</sup>; F. Yang <sup>2</sup>; T.L. McDowell <sup>2</sup>; A.A. Sahin <sup>3</sup>; J. Rodriguez-Canales <sup>2</sup>; N. Chen <sup>2</sup>; S. Balasenthil <sup>2</sup>; U.Q. Le <sup>2</sup>; H. Kadara <sup>2</sup>; C. Behrens <sup>4</sup>; I.I. Wistuba <sup>2</sup>; A.M. Killary <sup>2</sup>; S.T. Lott <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pathology Department, Baylor College of Medicine, Houston, Texas; 2) Department of Translational Molecular Pathology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas; 3) Division of Pathology and Laboratory Medicine, The University of Texas M. D. Anderson Cancer Center, Houston, Texas; 4) Department of Thoracic/Head and Neck Medical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

---

Little is known regarding the genetic drivers of the transition from ductal carcinoma in situ (DCIS) to invasive ductal carcinoma (IDC) of the breast. *DEAR1* (annotated as *TRIM62*) has previously been found to be an important regulator of cellular polarity and negative regulator of TGF $\beta$ -mediated epithelial-mesenchymal transition (EMT), a process intimately linked to tumor progression and invasion. Using ultra-deep targeted next generation sequencing (NGS), *DEAR1* was found to be frequently mutated in patients with DCIS alone and in those with concurrent DCIS and IDC. *DEAR1* exonic variants previously reported in IDC were identified, as well as novel missense and nonsense variants. Orthogonal validation of select variants was completed via digital PCR. *In-vitro* functional screens of select variants identified in DCIS indicated that these variants led to the abrogation of tumor suppressor functions of *DEAR1*. Moreover, analytical sensitivity of the assay to detect substitutions and indels at the lower limit of detection was determined using custom engineered spike-in constructs. Frequent mutation of *DEAR1* in DCIS suggests its role as a driver in DCIS progression and supports the utility of ultra-deep sequencing for the discovery of rare variants that could have major impact on patient prognosis and stratification for precision medicine approaches.

## PgmNr 903: CRISPR-Cas9 generated allelic series of rat mutations confirms *Tox3* as a breast cancer susceptibility gene.

### Authors:

L. Shunkwiler <sup>1</sup>; R. Pipaliya <sup>1</sup>; C.C. Ashy <sup>1</sup>; B. van Peel <sup>1</sup>; Y. Zhao <sup>1</sup>; E. Ballou <sup>1</sup>; J. Guz <sup>2</sup>; A. Awgulewitsch <sup>2</sup>; M.J. Kern <sup>2</sup>; B.M.G. Smits <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology & Laboratory Medicine, Medical University of South Carolina, Charleston, SC.; 2) Department of Regenerative Medicine and Cell Biology, Medical University of South Carolina, Charleston, SC.

---

Diagnoses of breast cancer in the US now exceed 250,000 cases a year, posing a huge clinical burden. Because risk is influenced by genetics and the environment, studies on susceptibility have largely been neglected outside of high-penetrance alleles. Independent genome-wide association studies (GWAS) have repeatedly identified a group of common, low-penetrance variants in populations of women of European ancestry that exist in high linkage-disequilibrium ( $r^2 > 0.9$ , tag SNP minor allele frequency  $\approx 0.24$ ) on *16q12.1* that strongly, and uniquely associate with risk of developing Estrogen Receptor-positive (ER+) breast cancer, the most common subtype of the disease. These variants are associated with reduced expression of the gene *Tox3*, a poorly-characterized transcription factor that is highly expressed in the luminal epithelium of the mammary gland and also been shown to functionally alter the expression of Estrogen Receptor (ER) responsive genes in breast cancer cells. To model the effect of the risk-associated locus, our lab used CRISPR-Cas9 to generate an allelic series of mutations in highly-conserved, orthologous regions of the rat genome. The deletion mutations were targeted toward the promoter and/or upstream regulatory regions and cause decreased transcript-level expression of *Tox3* in the mammary gland of our novel rat model. Results from 7,12-dimethylbenz[a]anthracene-induced mammary carcinogenesis experiments showed a significant increase in multiplicity and advanced pathological staging of tumors that arise in low-*Tox3*-expressing rats, as compared to wild type. The tumors were found to be ER+ and progesterone receptor positive (PR+) and resemble human adenocarcinoma of the breast. Of importance, tumors that arise in the mutant rats continue to express low levels of the gene and appear more invasive. Analysis of RNA-sequencing data from these tumors identified subsets of differentially-expressed genes that support the phenotype observed in this rat model and the human data. Our findings highlight the importance of this new approach, which used CRISPR-Cas9 genetic engineering technology to model GWAS-identified loci and confirmed *Tox3* is a breast cancer susceptibility gene.

## PgmNr 904: Functional analysis of three PALB2 germline variants of unknown significance seen in breast, skin, and renal cancers.

### Authors:

M.R. Toh <sup>1</sup>; C.E. Low <sup>2</sup>; S.T. Chong <sup>2</sup>; S.H. Chan <sup>2</sup>; N.D. Binte Ishak <sup>2</sup>; E. Courtney <sup>2</sup>; A. Rodrigue <sup>3,4</sup>; J.Y. Masson <sup>3,4</sup>; J. Ngeow <sup>1,2,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) DukeNUS, Singapore, Singapore; 2) National Cancer Center, Singapore, Singapore; 3) CHU de QuÃ©bec Research Center, Quebec City, Canada; 4) Department of Molecular Biology, Quebec City, Canada; 5) Lee Kong Chian School of Medicine, Singapore, Singapore; 6) Institute of Molecular and Cellular Biology, Singapore, Singapore

### Introduction

The PALB2 protein is essential to RAD51-mediated homologous recombination (HR) repair. Germline monoallelic *PALB2* pathogenic variants confer significant risks for breast cancer. However, majority of *PALB2* variants remain classified as variants of unknown significance (VUS). We aim to functionally and mechanistically evaluate three rare *PALB2* VUS.

### Methods

Patient-derived lymphoblastoid cell lines containing the VUS were analyzed for nuclear localization and foci formation of RAD51 as a measure of HR efficiency. To understand the mechanism underlying the HR deficiency, PALB2 nuclear localization was assessed using immunofluorescence studies.

### Results

Among these VUS, c.3251C>T (p.Ser1084Leu) occurred in a patient with metastatic breast cancer while c.1054G>C (p.Glu352Gln) and c.1057A>G (p.Lys353Glu) were seen in patients with squamous cell carcinoma of skin and renal cell carcinoma respectively. Variant c.3251C>T was located within the WD40 domain which normally masked the nuclear export signal sequence responsible for nuclear delocalization of PALB2. Correspondingly, c.3251C>T displayed aberrant cytoplasmic localization of PALB2 which led to an impaired RAD51 nuclear localization and foci formation. On the other hand, both c.1054G>C and c.1057A>G showed intact HR functions and nuclear localization of PALB2, consistent with their locations within non-functional domains. Additionally, the prevalence of c.1054G>C was similar among healthy controls and patients with breast cancer, suggestive of its non-pathogenicity.

### Conclusion

Our studies provided the functional evidence to support the consideration of c.3251C>T as a likely pathogenic variant, and c.1054G>C as a likely benign variant. More epidemiologic evidence is needed for the reclassification of c.1057A>G.

## PgmNr 905: Detection of somatic L1 insertions in single nuclei from cancer tissues.

### Authors:

K. Yamaguchi; A.O. Soares; H.H. Kazazian Jr.

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

---

Long interspersed element-1s (L1s) are transposable elements that proliferate in human genomes. L1 retrotranspositions are endogenous mutagens that cause insertional mutations in chromosomes and have caused over 150 cases of genetic diseases. Recently, our group found somatic insertions in Barrett's esophagus-, pancreatic-, colon- and gastric cancer. These insertions are present in both pre-cancer and cancer, suggesting that L1 insertion events occur before or very early during cancer development. From these findings using bulk gDNA from cancer patients, however, we could not know the number and varieties of L1 retrotransposition events varying among individual cells and how the cells which have de novo somatic insertions are distributed in pre-cancerous and cancer tissues. To further understand somatic L1 retrotransposition which contributes to cancer induction, we established a new method to detect L1 insertions from single nuclei based on capturing circularized L1 fragments containing 3' flanking gDNA. Using this method, we investigated L1 insertions in single nuclei from colon cancer, liver metastasis, and normal liver. We found that our L1-seq technique identifies many candidate somatic L1 sequences in single nuclei at their 3' ends. However, the great majority of these sequences turned out to be either unannotated non-reference L1s or multiple displacement amplification (MDA) chimeras. We now believe that it is imperative that potential somatic L1 insertions in single nuclei be characterized completely at both their 5' and 3' ends, including their TSDs and flanking genomic DNA. By PCR and Sanger sequencing at the 5' and 3' junctions, we found nine new L1HS insertions that were not observed in bulk gDNA study from same patient. From further investigation in single nuclei, we found five other insertions that had been previously observed in bulk gDNA from same patient. In addition, we have found four insertions in the pancreatic cancer case. Three are in cancer and metastasis, one is only in metastasis. Thus, using this new technique we can find new L1 insertions in single nuclei that are not found in bulk gDNA study. Based on these and future data, we should be able to determine how the new L1 insertions spread in cancerous cells.



## PgmNr 906: Characterization of 331g/a polymorphism of rp gene and identification of viral oncogene hmtv virus as genetic markers for the improvement of breast cancer management in CHU Yaounde Cameroon.

### Authors:

N. Nguedia Kaze <sup>1</sup>; M. Mbacham Wilfred <sup>1,2</sup>; C. Chedjou Jean Paul <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Yaounde I, YAOUNDE , CENTRE, Cameroon; 2) Biotechnology centre: Laboratory for public health and biotechnology; 3) Faculty of medicine

Characterization of 331g/a polymorphism of rp gene and identification of viral oncogene hmtv virus as genetic markers for the improvement of breast cancer management in CHU Yaounde Cameroon

**NGUEDIA KAZE N<sup>1</sup>**, CHEDJOU,J.P<sup>1</sup>, MBACHAM W<sup>1,2</sup>.

University of Yaoundé I

Faculty of medicine and Biomedical Sciences

**Background:** Breast cancer is a real public health problem in Cameroon, where more patients with this cancer usually die a year after diagnosis, as it is still based on histological examination, mortality due to cancer is far from decreasing. Since cancer is an accumulation of molecular changes, the +331 G/A polymorphism of PgR gene (progesterone receptor) and viral oncogene HMTV (Human Mammary Tumor Virus) has been recently considered as a molecular markers associated with breast cancer. Due to that we fixed our objectives to characterize these markers.

**Aim and objectives:** characterization of +331 G/A polymorphism of PgR gene (progesterone receptor) and viral oncogene HMTV (Human Mammary Tumor Virus) by semi-nested PCR to understand etiological factor of that cancer in Cameroon.

**Method:** We carried out a case control study, in which 26 cases diagnosed positive for breast cancer at the CHU of Yaounde were recruited through the identification of archived biopsies. Blood samples were also collected from 20 women recruited using a questionnaire and a inform concern sign by each of them. +331 G/A polymorphism in the PgR gene was identified using NlaIV endonuclease by PCR-RFLP, and HMTV viral oncogene by hemi-nested PCR. The data were analyzed using Microsoft Excel and SPSS v20.

**Results:** We got a mean age of 57,73 +/- 9,87 in our cancerous group with the predominance of infiltrant duct carcinoma at grade II of SBR. An Odd Ratio of 1.268 with Confident Interval of 95% 1.004-1.664 proving that there is a significant association between 331G/A mutation and breast cancer with P-value of 0.026, obtained by comparing the mutant group (AA) 28,5% and wild genotype (GG). In addition, 3 cases were detected with the HMTV virus, one was found in the cancer group and two in the control group.

**Conclusion:** These results indicate that, HMTV is considered as viral cause and can predispose to breast cancer, beside 331 G/A polymorphism is an associated risk factor of that cancer.

**Key words:** Breast cancer, progesterone receptor, +331G/A PgR gene polymorphism, Human Mammary Tumor Virus.



## PgmNr 907: Contribution of PALB2 germline pathogenic variants in breast and ovarian cancer in Greece.

### Authors:

F. Fostira; I. Konstantopoulou; P. Apostolou; M.S. Papamentzelopoulou; A. Delimitsou; V. Dellatola; A. Vagena; D. Kalfakakou; D. Yannoukakos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Molecular Diagnostics Laboratory, NSCR Demokritos, Athens, Greece

### Introduction:

Germline pathogenic variants in *PALB2* (Partner And Localizer of BRCA2) are rare, attributing for approximately 1-2% of familial breast cancer cases, while weaker but growing evidence suggests association with other homologous-recombination-deficiency associated malignancies, such as ovarian, prostate and pancreatic cancer, but also diffuse gastric cancer. We sought to assess the frequency and spectrum of *PALB2* mutations in a large cohort of Greek patients.

**Materials and Methods:** In total, 1958 (1301 breast and 657 ovarian) cancer patients fulfilling the NCCN criteria for genetic testing have been analyzed by Trusight Cancer panel, which targets 94 genes associated with cancer predisposition. All participants have signed informed consent prior to genetic testing.

**Results:** In total, ten distinct loss-of-function *PALB2* variants have been identified in 30 unrelated patients, of which 28 and 3 were diagnosed with breast and ovarian cancer (one patient had both diagnoses), respectively, resulting in a frequency of 2.15% and 0.46% of the breast and ovarian cancer group, respectively. Of note is that all ovarian cancer cases displayed high-grade serous histology, while 6/28 of breast cancer cases were triple-negative and 3/28 developed metachronous contralateral tumors. Mean age at breast cancer diagnosis was 43.7 years (range 32-61 years). Cascade testing was offered in 23 family relatives.

Interestingly, three recurrent variants were identified, i.e. found in >3 unrelated patients; one of those, c.2257C>T / p.(Arg753Ter) was recently shown to be a Greek founder. In addition, three recurrent missense variants [p.(Leu931Arg), p.(Leu1143His), and p.(Ser1165Leu)], which are predicted deleterious by *in silico* tools and have been shown to segregate with disease in our families, are rare in public databases and are therefore, classified as likely pathogenic, following ACMG guidelines. When also taking into account likely pathogenic variants, frequency of *PALB2* causative/likely causative alterations would calculate to 2.77% (36/1301) of the breast and 0.91% (6/657) of the ovarian cancer patients.

**Conclusions:** *PALB2* pathogenic/likely pathogenic variants account for a proportion of hereditary breast and ovarian cancer burden in individuals of Greek descent, with a spectrum exhibiting heterogeneity, but also strong founder effects. These patients can potentially benefit from targeted therapies, such as PARP inhibitors, as well as tailored clinical management.

## PgmNr 908: ANO7 expression shows association with the clinical outcome of prostate cancer.

### Authors:

C. Cieza-Borrella <sup>1,2</sup>; S. Wakerell <sup>2</sup>; Q. Karlsson <sup>2</sup>; S. Merson <sup>2</sup>; R. Rageevakumar <sup>2</sup>; T. Dadaev <sup>2</sup>; A.S. Ahmad <sup>3</sup>; J. Cuzick <sup>3</sup>; D. Berney <sup>3</sup>; C. Cooper <sup>2,4</sup>; E. Anokian <sup>2</sup>; E. Saunders <sup>2</sup>; R.A. Eeles <sup>2,5</sup>; Z. Kote-Jarai <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Medical and Biomedical Education, St.George's-University of London, UK; 2) The Institute of Cancer Research, London, UK; 3) Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, UK; 4) Norwich Medical School, University of East Anglia, UK; 5) The Royal Marsden Hospital NHS Trust, London, UK

---

Prostate Cancer (PCa) is the most frequently diagnosed cancer among men in developed countries. The effectiveness of PSA screening remains controversial and new molecular biomarkers are needed to improve screening and clinical management of the disease.

*ANO7* gene (*NGEP-L*) encodes for a member of the TMEM16 superfamily of transmembrane proteins predominantly expressed in prostate epithelial cells. Studies have suggested a role for ANO7 as calcium activated chloride channel and scramblase protein potentially involved in cell structure, adhesion and motility processes. Genome Wide Association studies (GWAS) and fine-mapping analyses have identified three independent PCa susceptibility signals with the credible set of candidate causal variants centred within the *ANO7* gene including nonsynonymous variants as top candidates.

In order to further investigate this association and the potential role of *ANO7* in PCa, we have performed a meta-analysis of germline mutations from 877 PCa samples from whole exome and targeted gene panel sequencing studies. We have also carried out immunochemistry staining of ANO7 protein in tissue microarrays (TMAs) including 475 PCa cases from the Transatlantic Prostate Group (TAPG). Finally we have overexpressed *ANO7* gene in the prostate epithelial normal cell line RWPE-1 and carried out proliferation, viability and migration analysis in 3D cultures along with RNA sequencing pathway analysis.

Next Generation Sequencing (NGS) meta-analysis identified a total of 55 cases carrying deleterious or likely-deleterious *ANO7* mutations. Of these mutations, 11 were classified as protein truncating variants (PTV). PTV carriers status showed significant association with Gleason score  $\geq 8$  ( $p=0.0485$ ). Analyses of the immunohistochemistry results showed significant association between low ANO7 expression and poor survival ( $p<2.2\times 10^{-16}$ ). This remained significant in a multivariable analysis including grade group, serum PSA and extent of disease ( $p=0.006$ ). In addition, functional assays in 3D cell culture showed greater viability in the RWPE-1 overexpressed clones compared to *ANO7*-ve control ( $p<0.0001$ ).

In conclusion, these results show evidence that *ANO7* expression associates with clinical outcome of prostate cancer and further functional analyses will provide insights into the molecular mechanisms underlying this association.



## PgmNr 909: Candidate gene polymorphism study in South Indian men with prostate cancer.

### Authors:

S. Daram <sup>1,2</sup>; S. Zubeda <sup>1</sup>; S. Poornima <sup>1,2</sup>; A. Hari <sup>2</sup>; V. Ayyala <sup>3</sup>; B. Srinadh <sup>4</sup>; P. Shailaja <sup>5</sup>; A. Sandhya <sup>6</sup>; D. Ramakrishna <sup>7</sup>; Q. Hasan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Molecular Medicine, Kamineni Hospitals, Hyderabad, Telangana, India; 2) Genetics and Molecular Medicine, Kamineni Life Sciences, Hyderabad, India; 3) Biochemistry, Kamineni Hospitals, Hyderabad, India; 4) Radiology, Kamineni Hospitals, Hyderabad, India; 5) Histopathology, Kamineni Hospitals, Hyderabad, India; 6) Genetics, Osmania University, Hyderabad, India; 7) Biochemistry, Kamineni Academy of Medical Sciences and Research Centre, Hyderabad, India.

---

Prostate Cancer (PCA) is the second most common cancer diagnosed in men and fifth leading cause of cancer related death in men globally. Prostate Specific Antigen (PSA) is the clinically used marker available for PCA screening however, it lacks specificity and sensitivity. To develop a non-invasive molecular marker. Androgen receptor (AR) and Cytochrome P450 3A5\*3 (CYP3A5\*3) gene polymorphisms were evaluated which play an important role in the metabolism of testosterone.

Aim of the present study was to evaluate AR CAG repeats and CYP3A5\*3 gene polymorphism in men with PCA, benign prostate hyperplasia (BPH) and controls from South Indian men.

Genomic DNA was isolated and a three step PCR procedure was carried out using specific primers for AR and CYP3A5 genes. CAG repeats in AR gene and CYP3A5 genotype was analyzed by agarose gel electrophoresis (AGE). Odds ratios (OR) and chi-square were calculated using MedCalc<sup>R</sup> version 18.2.1.

AR with short (<22) CAG repeats and 'GG' genotype of CYP3A5 gene was identified in **61%** of PCA patients, whereas BPH and controls had these genotypes in 10% and 13% respectively. There is a significant difference of AR <22 CAG repeats with CYP3A5\*3 GG genotype with PCA ( $X^2 = 50.105$ , 95%CI- 35.4494 to 58.3715, **P<0.0001**) when compared with controls.

This is the first case-control study exhibiting an association of AR and CYP3A5 gene polymorphisms in combination, with PCA indicating that these molecular markers can differentiate PCA from benign hyperplasia and also identifying men who are at a risk of developing PCA.

## PgmNr 910: Circulating tumor DNA content varies in different fractions of plasma.

### Authors:

L. Sun <sup>1,2,3</sup>; M.J. Du <sup>2</sup>; Y.F. Zhu <sup>1</sup>; S.K. Wang <sup>3</sup>; L. Wang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Laboratory Medicine, the Second Affiliated Hospital, Nanjing Medical University, Nanjing, Jiangsu, China; 2) Department of Pathology and MCW Cancer Center, Medical College of Wisconsin, Milwaukee, WI, USA; 3) Department of General Clinical Research Center, Nanjing First Hospital, Nanjing Medical University, Nanjing, Jiangsu, China

---

Circulating tumor DNA (ctDNA) in plasma has been used as a biomarker for cancer detection and outcome prediction. It is well known that ctDNA accounts for a small fraction of plasma cell-free DNA. To sensitively detect such low level ctDNA, multiple experimental and computational approaches have been evaluated, including size selection and molecular barcoding. However, few studies have systematically examined the effect of variable processing of plasma to generate fractionated plasma samples for ctDNA content comparison. To address this issue, we performed a five consecutive physical and chemical precipitations of plasma samples from 9 pre-selected small cell lung cancer (SCLC) patients with high tumor burden. We collected five precipitate fractions (fraction 1: cells and large debris, fraction 2: debris and large vesicles, fraction 3: large microvesicles, fraction 4: fibrin, fraction 5: exosomes) and one leftover supernatant (fraction 6) for a total of 54 fractionated specimens. Since all 6 fractions were derived from the same 1ml of plasma by sequential precipitations, we were able to compare fragment size and ctDNA contents in different fractions from the same patients. Fragment size analysis showed dramatic differences from a peak of 7-10kb in the first precipitation fraction 1 (500g centrifugation) to 140-160bp in final supernatant fraction 6 after 5 precipitations. After low-pass whole genome sequencing in each of the 54 samples, we calculated ctDNA content using a novel copy number-based algorithm. This analysis showed relatively high ctDNA content in the fractions of large vesicles (median=20.18%) and small vesicles (exosomes) (median=20.45%), but the highest ctDNA content in the fraction 6 of "leftover supernatant" (median=27.04%) after multiple precipitations. Comparatively, median ctDNA content in whole plasma from the same 9 patients was 23.34%. Our study provided a new insight into fragment size and ctDNA content variations in different fractions of plasma. DNA from supernatant after sequential centrifugation contains more ctDNA than total plasma cfDNA. It is worth mentioning that the plasma fractionation may allow maximum use of valuable plasma samples for a wide variety of studies. Our result will provide timely guidance in the emerging field of liquid biopsy.

## PgmNr 911: Nanopore ultra-long read sequencing provides insights into structural variation in cancer genome.

### Authors:

L. Gong <sup>1</sup>; C.H. Wong <sup>1</sup>; Y. Zhu <sup>1</sup>; C.Y. Ngan <sup>1</sup>; A. Gujar <sup>2</sup>; R. Verhaak <sup>2</sup>; C.L. Wei <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Technologies, The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut.; 2) The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut.

---

Genomic structural variants (SVs), the major hallmarks of cancer genomes, can promote tumor progression by perturbing gene structures and expression control, which contribute to the tumor heterogeneity and evolution. Despite their importance, the ability to detect and characterize SVs at the molecular level is limited by short-read sequencing approaches: large and complex SVs thwart efforts to detect them and correctly define their structures. Single-molecule DNA sequencing technologies offer significantly longer read lengths which can facilitate the detection and analysis of highly rearranged cancer genomes. In this study, we generated deep coverage of ultra-long reads from three glioblastoma (GBM) patient-derived neurospheres and characterized the full-spectrum of SVs with our customized SV caller Picky. From over 100 Gb of nanopore ultra-long reads with N50 length over 50 Kb, we presented the unbiased survey of full-scale SV landscapes in GBM cancer genomes. Leveraged by the long read lengths, we detected complex SVs with low allele frequency and provided SV with phasing information. Particularly, we uncovered SVs affecting genes with known oncogenic activities. Comparing the SVs in the primary and recurrent tumors from the same patients, we revealed rearrangement events associated with tumor relapse and characterized their genomic features associated with chromatin organization and transcriptional regulation. A complete understanding of the structure and distribution of SVs in cancer genome will empower the cancer research community to reveal mechanisms that induce genome instability, identify prognostic signatures of tumor progression and suggest targets for novel treatment strategies.



## PgmNr 912: Evaluation of copy-number variants as modifiers of breast and ovarian cancer risk for *BRCA1* and *BRCA2* pathogenic variant carriers.

### Authors:

C. Hakkaart <sup>1</sup>; L. Marquart <sup>2</sup>; J. Dennis <sup>3</sup>; J.F. Pearson <sup>1</sup>; G. Wiggins <sup>1</sup>; A.B. Spurdle <sup>4</sup>; G. Chenevix-Trench <sup>4</sup>; A. Antoniou <sup>2</sup>; L.C. Walker <sup>1</sup>; Consortium of Investigators of Modifiers of *BRCA1/2*

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology and Biomedical Science, University of Otago, Christchurch, Christchurch, New Zealand; 2) Statistics Unit, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 3) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 4) Genetics and Computational Biology Division, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

---

Germline pathogenic variants in *BRCA1* and *BRCA2* (*BRCA1/2*) confer a high risk of breast and ovarian cancer. These risks are modified by environmental and genetic factors. Genome-wide studies of *BRCA1/2* pathogenic variant carriers have identified strong associations between single nucleotide polymorphism and breast and ovarian cancer risk. Copy number variants (CNVs) make a significant contribution to our genetic variation but are largely unexplored as modifiers of breast and ovarian risk in *BRCA1/2* pathogenic variant carriers. We have conducted a genome wide association study of germline CNVs in 15,679 *BRCA1* and 10,981 *BRCA2* pathogenic variant carriers from the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA) using genotype data generated by the OncoArray Network. Analyses were based on the modeling the retrospective likelihood of observing the CNV conditional on the observed phenotype. We identified 241 gene loci that were associated (unadjusted  $p < 0.01$ ) with breast and/or ovarian cancer risk. For *BRCA1* pathogenic variant carriers, we identified 16 deletions associated with breast cancer and 28 deletions associated with ovarian cancer risk. For *BRCA2* pathogenic variant carriers, we identified 18 deletions associated with breast cancer and 52 deletions associated with ovarian cancer risk. Deletions in *BRCA1* were associated ( $p = 1.9 \times 10^{-4}$ ) with a 1.8-fold risk of breast cancer in the *BRCA1* pathogenic variant carrier cohort. We also identified risk-associated CNVs in several genes (*LSP1* and *TERT*) identified by previous SNP based analyses as candidate modifier genes. CNV deletions in the gene *SULT1A1* were associated with decreased breast cancer risk ( $RR = 0.5$ ,  $p = 9.8 \times 10^{-3}$ ) but increased ovarian cancer risk ( $RR = 3.3$ ,  $p = 7.3 \times 10^{-3}$ ) in *BRCA1* pathogenic variant carriers. We selected the *SULT1A1* gene for functional analysis as it is important for detoxification of carcinogens, steroid hormones (e.g., estrogen), and medications (e.g., tamoxifen). We showed that expression of *SULT1A1* was positively correlated with copy number status in breast and ovarian tissue. Breast cancer cell lines (MCF7 and MCF10A) containing CRISPR-Cas induced heterozygous pathogenic variants in *BRCA1* are being utilised to assess the functional impact of *SULT1A1* depletion. In summary, our study has provided evidence that multiple CNV loci with overlapping genes modify risk of breast and ovarian cancer in *BRCA1/2* pathogenic variant carriers.

## PgmNr 913: Identification and functional follow up of a locus at 12q24.33 associated with progression-free survival in ovarian cancer.

### Authors:

K. McCue<sup>1</sup>; M. Quinn<sup>1</sup>; S. Johnatty<sup>1</sup>; J. Beesley<sup>1</sup>; S. Armasu<sup>2</sup>; S. Winham<sup>2</sup>; E. Goode<sup>2</sup>; S. MacGregor<sup>1</sup>; G. Chenevix-Trench<sup>1</sup>; Ovarian Cancer Association Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Computational Biology, QIMR Berghofer, Brisbane, Queensland (QLD), Australia; 2) Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA

---

Genome-wide association studies have identified many loci for cancer risk, but very few for cancer outcome. Epithelial ovarian cancer (EOC) is one of the most chemosensitive of the solid tumours and has a high initial response rate to platinum/taxane treatment, but the median progression free survival (PFS) is only 18 months, and most women relapse within 5 years. We conducted a GWAS of PFS in 2352 women with EOC who had undergone cytoreductive surgery followed by standard carboplatin and paclitaxel chemotherapy. We identified seven variants at 12q24.33 associated with PFS at genome-wide significance, tagged by rs10794418 ( $P=1.47e-08$ ). The association with PFS was stronger for endometrioid (HR 2.48; 95% CI 1.57-3.92) than for serous (HR 1.17; 95% CI 1.07-1.27) EOC. The only gene in the region associated with PFS in EOC patients treated with standard chemotherapy was *ULK1* ( $P=0.02$ ), with higher expression associated with shorter PFS. In addition, expression analysis in 310 EOC patients showed that rs10794418 was associated with expression of *ULK1* ( $P=0.009$ ), with the T allele associated with shorter PFS being associated with increased expression. We performed chromatin conformation capture using the *ULK1* promoter as a bait and tiling across the associated variants. This identified consistent looping in both serous and endometrioid EOC cell lines. The variants lie in a putative regulatory element (PRE), marked by H3K27Ac marks in ovarian cells. Luciferase reporter assays show that the PRE acts as a silencer in A2780 and OVCAR8 EOC cell lines but no differential effect was seen with the two variants that could be tested. Electrophoretic mobility shift assays are underway. Inhibition of *ULK1* can inhibit autophagy and sensitises some cancer cell lines to chemotherapy. We therefore treated TOV112D and OVCAR8 EOC cell lines with carboplatin and paclitaxel in addition to an inhibitor of ULK1, SBI0206965. SBI0206965 had some efficacy as a single agent but exhibited no synergy with chemotherapy *in vitro*, but we will also test this combination *in vivo*.

## PgmNr 914: Gene-specific criteria for germline *RUNX1* variant curation: Recommendations from the ClinGen Myeloid Malignancy Variant Curation Expert Panel.

### Authors:

Luo <sup>1</sup>; S. Feurstein <sup>2</sup>; S. Mohan <sup>3</sup>; C. Porter <sup>4</sup>; S.A. Jackson <sup>5</sup>; S. Keel <sup>6</sup>; M. Chicka <sup>7</sup>; A.L. Brown <sup>8</sup>; C. Kesserwan <sup>9</sup>; A. Agarwal <sup>10</sup>; M. Luo <sup>11</sup>; Z. Li <sup>12</sup>; J.E. Ross <sup>3</sup>; P. Baliakas <sup>13</sup>; D. Pineda-Alvarez <sup>14</sup>; C. DiNardo <sup>15</sup>; A. Bertuch <sup>16</sup>; T. Vulliamy <sup>17</sup>; Y. Wang <sup>18</sup>; K.E. Nichols <sup>19</sup>; L. Malcovati <sup>20</sup>; M.F. Walsh <sup>21</sup>; L.H. Rawlings <sup>22</sup>; S.K. McWeeney <sup>23</sup>; J. Soulier <sup>24</sup>; N. Mehta <sup>25</sup>; N.A. Speck <sup>26</sup>; S.E. Plon <sup>1</sup>; D. Wu <sup>27</sup>; L.A. Godley <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics-Oncology, Baylor College of Medicine, Houston, TX, USA.; 2) Section of Hematology/Oncology and Center for Clinical Cancer Genetics, University of Chicago, Chicago, IL, USA.; 3) Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC, USA.; 4) Department of Pediatrics, Emory University School of Medicine, Atlanta, GA, USA.; 5) GeneDx, Gaithersburg, MD, USA.; 6) Department of Medicine, Division of Hematology, University of Washington, Seattle, WA, USA.; 7) Prevention Genetics, Marshfield, WI, USA.; 8) Center for Cancer Biology, SA Pathology & University of South Australia, Adelaide, Australia.; 9) Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA.; 10) Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA.; 11) Division of Genomic Diagnostics at Children's Hospital of Philadelphia, University of Pennsylvania, PA, USA.; 12) Houston Methodist Institute for Academic Medicine, Houston, TX, USA.; 13) Department of Immunology, Genetics, and Pathology, Uppsala University, Uppsala, Sweden.; 14) Invitae, San Francisco, California, USA.; 15) Department of Leukemia, University of Texas MD Anderson Cancer Center, Houston, TX, USA.; 16) Department of Pediatrics, Section of Hematology and Oncology, Baylor College of Medicine, Houston, TX, USA.; 17) Genomics and Child Health, Blizard Institute, Queen Mary University of London, London, United Kingdom.; 18) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.; 19) Department of Oncology, St Jude Children's Research Hospital, Memphis, TN, USA.; 20) Department of Molecular Medicine, University of Pavia, Pavia, Italy.; 21) Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.; 22) Department of Genetics and Molecular, SA Pathology, Adelaide, Australia.; 23) Division of Biostatistics, Department of Public Health and Preventive Medicine, Oregon Health & Science University, Portland, OR, USA.; 24) INSERM U944/CNRS UMR7212, Saint Louis Hospital, Paris, France.; 25) Hematopathology Division of the Department of Laboratory Medicine & Pathology, Mayo Clinic, Rochester, MN, USA.; 26) Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.; 27) Department of Laboratory Medicine, University of Washington, Seattle, WA, USA.

---

An increasing number of myeloid neoplasia is reported in association with germline alterations in genes involved in hematopoiesis and lineage differentiation, and clinicians are becoming more aware of the contribution of germline alterations to myeloid neoplasia. The wider utilization of multiplex germline testing using next-generation sequencing and microarray platforms has led to a deluge of sequence variant alterations for which pathogenicity interpretation continues to be challenging. Despite the implementation of ACMG/AMP guidelines for sequence variant interpretation, the number

of variants of uncertain significance (VUS) and curation discrepancies remain elevated. To address this hurdle, the Clinical Genome Resource (ClinGen) Myeloid Malignancy Variant Curation Expert Panel (MM-VCEP) was initiated as a collaboration between the American Society of Hematology and ClinGen to perform disease-/gene-specific modifications for inherited myeloid malignancies. This team consists of expert clinicians, clinical laboratory diagnosticians, and researchers interested in developing and implementing standardized protocols for expert interpretation of germline variation in inherited myeloid malignancies. The MM-VCEP began optimizing ACMG/AMP rules on *RUNX1* associated familial platelet disorder with predisposition to acute myeloid leukemia (FPD/AML). The optimization of the ACMG/AMP guidelines encompasses disease-/gene-informed specifications or strength adjustments of existing rules, included calculation of minor allele frequency thresholds, functional domains and mutational hot spots identifications, functional assay thresholds, and phenotype-specific guidelines. Preliminary rules were further validated and refined using a pilot set of 52 variants with benign/likely benign, pathogenic/likely pathogenic, VUS and conflicting assertions in ClinVar, covering various types of SNVs (missense, nonsense, frameshift, splice-site, in-frame duplication, synonymous and intronic variants) and CNVs (intragenic deletions). The application of *RUNX1*-specific ACMG/AMP criteria resulted in the resolution of conflicting variants and reduced the number of VUS. In summary, the ClinGen MM-VCEP aims to develop recommendations to optimize ACMG/AMP criteria for standardization of variant interpretation in myeloid leukemia genes and make expert-reviewed and interpreted variants available to the hematology community through ClinVar to support patient care and research.

## PgmNr 915: Distinct microbial communities that differ by race, stage, or breast-tumor subtype in breast tissues of non-Hispanic black and non-Hispanic white women.

### Authors:

A. Starlard-Davenport <sup>1</sup>; J.F. Pierre <sup>2</sup>; L. Makowski <sup>3</sup>; E. Tolley <sup>4</sup>; B. Lyn-Cook <sup>5</sup>; G. Vidal <sup>6</sup>; A. Smith <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee; 2) Department of Pediatrics, University of Tennessee Health Science Center, Memphis, Tennessee; 3) Department of Medicine, Department of Pharmaceutical Sciences, University of Tennessee Health Science Center, Memphis, TN; 4) Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee; 5) Division of Biochemical Toxicology, FDA/National Center for Toxicological Research, Jefferson, AR; 6) Department of Medicine, University of Tennessee, West Cancer Center, Memphis, TN

---

Growing evidence highlights an association between an imbalance in the composition and abundance of bacteria in the breast tissue (referred as microbial dysbiosis) and breast cancer in women. However, studies on the breast tissue microbiome have not been conducted in non-Hispanic Black (NHB) women. We investigated normal and breast cancer tissue microbiota from NHB and non-Hispanic White (NHW) women to identify distinct microbial signatures by race, stage, or tumor subtype. Using 16S rRNA gene sequencing, we observed that phylum Proteobacteria was most abundant in normal (n=8), normal adjacent to tumor (normal pairs, n=11), and breast tumors from NHB and NHW women (n=64), with fewer Firmicutes, Bacteroidetes, and Actinobacteria. Breast tissues from NHB women had a higher abundance of genus *Ralstonia* compared to NHW tumors, which could explain a portion of the breast cancer racial disparities. Analysis of tumor subtype revealed enrichment of family *Streptococcaceae* in TNBC. A higher abundance of genus *Bosea* (phylum Proteobacteria) increased with stage. This is the first study to identify racial differences in the breast tissue microbiota between NHB and NHW women. Further studies on the breast cancer microbiome are necessary to help us understand risk, underlying mechanisms, and identify potential microbial targets.

## PgmNr 916: Effects of 3300 del A-1061 Ter *BRCA1* frameshift mutation on breast carcinogenesis.

**Authors:**

M. Pongsavee

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medical Technology, Faculty of Allied Health Sciences, Thammasat University, Patumthani, Thailand

---

*BRCA1* gene mutation increases risk of hereditary breast/ovarian carcinogenesis. Superoxide dismutase enzyme protects oxidative stress in human. This report describes the effects of 3300 del A-1061 Ter *BRCA1* mutation at exon 11 for oxidative stress induction and may cause breast carcinogenesis. The effects of this mutation was studied by MTT dye reduction technique and superoxide dismutase (SOD) activity assay. The results showed that the percentage of cell viability of mutant cells in different H<sub>2</sub>O<sub>2</sub> concentrations was less than the percentage of cell viability of wild type cells ( $P < 0.05$ ). This mutation causes oxidative stress through superoxide dismutase activity inhibition ( $P < 0.05$ ). 3300 del A-1061 Ter *BRCA1* mutation can effect oxidative stress and lead to breast carcinogenesis.

## PgmNr 917: Characterizing the proteome- and transcriptome-QTL landscape in prostate tumors.

### Authors:

N. Mancuso<sup>1,2</sup>; K.E. Houlahan<sup>3</sup>; J. Livingstone<sup>3</sup>; T. Kislinger<sup>4</sup>; B. Pasaniuc<sup>5,6,7,8,9,12</sup>; P.C. Boutros<sup>3,4,6,7,8,9,10,11,12</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; 2) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA; 3) Ontario Institute for Cancer Research, Toronto, ON M5G 0A3, Canada; 4) Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada; 5) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 6) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 7) Department of Computational Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 8) Jonsson Comprehensive Cancer Center, University of California, Los Angeles, CA 90024, USA; 9) Institute for Precision Health, University of California, Los Angeles, CA 90024, USA; 10) Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada; 11) Department of Urology, University of California, Los Angeles, CA 90024, USA; 12) these authors contributed equally

---

The steady-state abundance of proteins plays a crucial role across broad cellular processes and biological pathways. While considerable effort has been made to characterize the regulatory impact of genetic variation on steady-state mRNA abundance, with few exceptions, the genetic regulation of proteins is still largely unmapped. In this work we integrate protein and mRNA abundance measured in prostate cancer samples ( $n = 56$ ) together with common genetic variation to elucidate the regulatory landscape and trace the causal molecular cascade.

First, we performed a local protein quantitative trait loci (pQTL) analysis using 6,697 protein groups quantified through mass spectrometry (Sinha, Huang, Livingstone et al. Cancer Cell 2019). Overall, we identified 13 (11) genes with pQTLs at an FDR  $< 0.1$  (0.05) localized within 100kbp of transcription start-sites (TSS). To establish the validity of our findings, we compared pQTL results with a recently published plasma protein pQTL analysis in a much larger cohort ( $n=3,301$ ). We found a significant overlap of total genes assayed in both studies (OR=1.4;  $P=8 \times 10^{-8}$ ) indicating that we were well powered to replicate our findings. Focusing on the 13 genes with prostate pQTLs, 4 were assayed in plasma and 3 were plasma pQTLs, which suggests a shared regulatory landscape between plasma and prostate tumors.

Next, we performed a local expression quantitative trait loci (eQTL) analysis using 28,692 genes measured by RNA-seq. We identified 182 (152) genes with eQTLs at an FDR  $< 0.1$  (0.05) with 143/182 localizing within 100kbp of the TSS. To validate our eQTL findings, we recapitulated our analysis in the Cancer Genome Atlas (TCGA) prostate adenocarcinoma (PRAD) cohort. We saw a significant overlap of all genes assayed in both studies (OR=6.8;  $P<2 \times 10^{-16}$ ). Of the 182 genes with prostate eQTLs, 46 were found in TCGA with all 46 eQTLs replicating, demonstrating the generalizability of our results.

To quantify the shared underpinnings of steady-state mRNA and protein abundance in prostate tumors, we next computed the similarity of measured expression levels. We found that the average rank correlation across 6,756 genes was  $r=0.19$  (95CI[0.185, 0.195]). We also estimated the average mediating effect of eQTLs on pQTLs while accounting for shared environmental effects and observed  $\beta=0.45$  (95CI[0.09-0.8]). Our findings shed light on the causal regulatory cascade of genetic variation, mRNA, and protein abundance in prostate cancer tumors.



## PgmNr 918: Human papillomavirus (HPV) 16 genetic variants affect *E7* viral oncogene protein expression and transformation.

### Authors:

H. Lou <sup>1</sup>; H.C. Li <sup>1</sup>; J.F. Boland <sup>1</sup>; M. Steinberg <sup>1</sup>; M. Cullen <sup>1</sup>; S. Bass <sup>1</sup>; S.A. Brodie <sup>1</sup>; R. Burk <sup>2</sup>; M. Yeager <sup>1</sup>; N. Wentzensen <sup>3</sup>; S. Anderson <sup>4</sup>; L. Mirabello <sup>3</sup>; M. Dean <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Leidos Biomedical Research, Inc., National Laboratory for Cancer Research, Frederick, MD USA.; 2) Albert Einstein College of Medicine, Bronx, NY, USA; 3) Division of Cancer Epidemiology and Genetics, NCI, Gaithersburg, MD USA; 4) Cancer and Inflammation Program, CCR, NCI, Frederick, MD USA

**Introduction:** The HPV16 E7 oncoprotein is expressed in cervical cancer cells and serves as a tumor-specific antigen target for immunotherapy. E7 genetic conservation is critical to carcinogenesis. Our previous study of 5,570 HPV16 infected subjects found that cancer-free controls with benign HPV16 infections had significantly more E7 nonsynonymous and nonsense variants compared to cancers from around the world. This work also identified two rare variants in a putative enhancer in the viral upstream regulatory region (URR) that were associated with a reduced risk of cancer. To investigate whether the E7 variants affect protein function, or cell invasion, we expressed individual E7 wild-type and mutant proteins.

**Methods:** The full length HPV16 E7 coding region was cloned and modified by site-directed mutagenesis. Variants found in both controls (H9R, D21N, N29S, E33K, T56I, D62N, S63F, S63P, T64M, E80K, D81N, P92L, P92S); and precancers (P6L, D14E, N29H, H51N, R77S) were generated. Proteins were expressed in HeLa (cervical cancer) and HEK293 (embryonic kidney) cell lines and steady state levels of nuclear and cytoplasmic protein were determined by western blotting. The variants identified in the HPV16 URR were studied by electrophoretic mobility shift assays (EMSA).

**Results:** The H9R, E33K, E80K, D81N, P92L and P92S mutants found in the control group showed significantly lower E7 protein levels, indicating a defect in translation, or stability. In contrast, the P6L and R77S variants found in precancers had E7 levels similar to wild-type E7. Interestingly, the D14E variant, found in pre-cancer, had little effect on colony formation in NIH3T3 mouse fibroblasts soft agar cultures, whereas the E33K mutation from the control group reduced growth in soft agar. The EMSA demonstrated that the protective URR variants were associated with the loss of binding of yet unidentified transcription factors in HEK293, HeLa and MCF7 (human breast cancer cells) nuclear lysates. From the sequence, CEBP/A and RUNX are the most likely candidates.

**Conclusions:** Taken together, these data demonstrate molecular differences between HPV16 E7 coding and regulatory mutants that might explain population-based variation in cancer risk of HPV16 variants.

## PgmNr 919: Variants in *LRRC34* reveal distinct mechanisms for predisposition to papillary thyroid carcinoma.

### Authors:

D.F. Comiskey<sup>1</sup>; H. He<sup>1</sup>; S. Liyanarachchi<sup>1</sup>; M. Sheikh<sup>1</sup>; L.K. Genutis<sup>1</sup>; L. Yu<sup>2</sup>; P. Brock<sup>3</sup>; A. de la Chapelle<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Cancer Genetics Program, Comprehensive Cancer Center, The Ohio State University, Columbus, OH; 2) Center for Biostatistics, The Ohio State University, Columbus, OH; 3) Department of Internal Medicine, Comprehensive Cancer Center, The Ohio State University, Columbus, OH

---

Papillary thyroid carcinoma (PTC) arises from thyroid follicular epithelial cells and accounts for over 90% of all thyroid cancer diagnoses. PTC demonstrates high heritability and a low somatic mutation burden relative to other cancers. Furthermore, heritable high-penetrance pathogenic mutations have been largely absent from familial and sporadic PTC cases. We therefore hypothesized that the genetic risk predisposing to PTC is mainly due to a combination of low penetrance variants. In order to discover these variants we used genome-wide association studies (GWAS). The most recent GWAS revealed the association of thyroid cancer with a missense variant, rs6793295, at 3q26 in a gene called *LRRC34*.

In addition to rs6793295, we observed a second missense variant, rs10936600, in high linkage disequilibrium (LD) with rs6793295. We hypothesized that missense variants in *LRRC34* would result in a change of affinity for its binding partners. Therefore, we performed a mass spectroscopy screen and found that RANBP1 binds less to missense *LRRC34* as compared to the wild-type isoform. Furthermore, overexpression of the missense *LRRC34* caused a significant reduction in RanGTP levels and an increase of cells undergoing apoptosis.

We then characterized a second LD block containing several regulatory variants upstream of *LRRC34*. We conducted luciferase assays on these variants and found a pattern of allele-specific expression for three regulatory variants. Microarray analysis of cells knocked down for *LRRC34* disclosed perturbations in genes involved in cellular movement and cell cycle. The effects of *LRRC34* knockdown reduced the migration of thyroid cancer cells.

To further explore the role of the five functional variants (two missense and three regulatory variants) at 3q26 with PTC risk we performed haplotype analysis in an Ohio cohort (2718 cases/3210 controls). We observed three significant haplotypes: all risk alleles (odds ratio [OR] = 1.23), all protective alleles (OR = 0.79), and risk missense alleles with protective regulatory alleles (OR = 0.72). These data indicate that regulatory variants might negate or mitigate the risk of bearing risk alleles for *LRRC34* missense variants.

In conclusion we have demonstrated two separate mechanisms dictating PTC risk at 3q26 using both biochemical and genetic techniques. Further studies to refine and characterize additional risk loci will enable a comprehensive understanding of the germline genetics of thyroid cancer.

## PgmNr 920: Alternative splicing of *APOBEC3B* as a preventive mechanism for APOBEC-mediated mutagenesis.

### Authors:

A.R. Banday<sup>1</sup>; O.O. Onabajo<sup>1</sup>; SHY. Lin<sup>1</sup>; J.M. Vargas<sup>1</sup>; A. Obajemu<sup>1</sup>; K.A. Delviks-Frankenberry<sup>2</sup>; A. Bayanjargal<sup>1</sup>; P. Lamy<sup>3</sup>; C. Zettelmeyer<sup>1</sup>; O. Florez-Vargas<sup>1</sup>; L. Dyrskjøl<sup>3</sup>; V.K. Pathak<sup>2</sup>; L. Prokunina-Olsson<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, NCI, NIH, Bethesda, Maryland.; 2) Viral Mutation Section, HIV Dynamics and Replication Program, Center for Cancer Research, National Cancer Institute, Frederick, MD, USA; 3) Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark

---

**Background:** APOBEC3B (A3B), a member of the cytidine deaminase family of enzymes, is considered as one of the major endogenous mutagens in tumor genomes. A3B can cause C-to-T or C-to-G mutations within the TCA and TCT motifs commonly referred to as APOBEC mutational signature. The loads of APOBEC-signature mutations have been correlated with A3B mRNA expression levels in tumors, but the factors that regulate the expression of this enzyme remain largely unknown. Here, we explored the role of alternative splicing on the expression of A3B and its consequences on mutagenesis.

**Results:** The human genome assembly GRCh37 (hg19) annotates three mRNA isoforms of A3B. These include a canonical isoform with eight exons, here referred as A3B1, and two alternatively spliced isoforms – one occurs due to cryptic acceptor site in exon 6 (A3B2), and the other – due to skipping of exon 5 (A3B3). Using a combination of *in silico* protein structural analysis and *in vitro* deamination assays, we determined that A3B2 and A3B3 protein isoforms are non-mutagenic, but their production reduces the levels of the canonical mutagenic isoform, A3B1. Specifically, exon 5 skipping, a major splicing event in A3B, not only reduced expression of the mutagenic A3B1 isoform but also targeted the resulting transcript for nonsense-mediated decay. Increased exon 5 skipping was significantly associated with lower APOBEC mutational signature load in bladder and lung tumors in TCGA, and longer progression-free survival in patients with non-muscle-invasive bladder cancer. Using exon-trap assays, we determined that skipping of exon 5 was dependent on weak branch point sites in intron 4. Treatment with pladienolide B, which inhibits SF3B1 splicing factor and prevents spliceosome engagement at weak branch point sites, reduced the expression of mutagenic A3B enzyme and its *in vitro* deamination activity, validating the potential role of exon 5 skipping in the regulation of A3B associated mutagenesis.

**Conclusions:** Our results suggest that alternative splicing of A3B is an intrinsic preventive mechanism for APOBEC-mediated mutagenesis, and can be harnessed to control this mutational process in clinically relevant conditions.

## PgmNr 921: A functional network of gastric-cancer-associated splicing events controlled by dysregulated splicing factors.

### Authors:

S. Cheng <sup>1,2,3</sup>; D. Ray <sup>3</sup>; R. Lee <sup>3</sup>; K.B. Naripogu <sup>3</sup>; P. Yusoff <sup>3</sup>; P. Goh <sup>3</sup>; Y. Liu <sup>2,3,4</sup>; Y. Suzuki <sup>2,3</sup>; K. Das <sup>3</sup>; P. Tan <sup>3</sup>; D.M. Epstein <sup>3</sup>; S.G. Rozen <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology and Biostatistics, Hwazhong University of Science and Technology, Wuhan, Hubei, China; 2) Center for Computational Biology, Duke-NUS Medical School, Singapore; 3) Cancer & Stem Cell Biology Programme, Duke-NUS Medical School, Singapore; 4) Singapore MIT Alliance, Singapore

---

Although several genes have known oncogenic splice forms in gastric cancer, the full scope of the role of aberrant splicing in gastric oncogenesis remains unclear. In this study, we developed a network of aberrant splicing in gastric cancer by elucidating tumor-associated (TA) alternative splicing events (ASEs) and their upstream regulators. RNA-sequencing of 19 tumor-normal pairs revealed 118 TA ASEs and 8 splicing factors (SFs) with dysregulated protein and RNA expression in gastric cancer. We constructed a network model of gastric-cancer-associated splicing based on correlations between ASE splicing ratios and SF transcript levels, and independently validated the model by analysis of TCGA data (323 gastric tumors, 31 normal samples). This model contained three clusters centered on the SFs *PTBP1*, *ESRP2*, and *MBNL1*. Individual knockdown of these SFs in gastric cell lines drove TA ASEs observed in gastric tumors, confirming causal links between the dysregulated SFs and downstream ASEs. Knockdown of SFs altered transwell migration rates in directions consistent with expected roles in oncogenesis and with the observation that TA ASEs disproportionately affected cytoskeletal-organization genes. This computational analysis followed by in-vitro experimentation indicated that concomitant SF dysregulation and generation of TA ASEs contribute to gastric oncogenesis.

## PgmNr 922: Non-coding somatic mutations converge on the PAX8 pathway in epithelial ovarian cancer.

### Authors:

R.I. Corona <sup>1,2</sup>; J.-H. Seo <sup>3,4</sup>; X. Lin <sup>1</sup>; D.J. Hazelett <sup>2</sup>; J. Reddy <sup>1</sup>; M.A.S. Fonseca <sup>1</sup>; F. Abassi <sup>1</sup>; Y.G. Lin <sup>5</sup>; P.Y. Mhawech-Fauceglia <sup>6</sup>; S.P. Shah <sup>7,8,9</sup>; D.G. Huntsman <sup>8,9,10</sup>; A. Gusev <sup>3,11</sup>; B.Y. Karlan <sup>1</sup>; B.P. Berman <sup>2</sup>; M.L. Freedman <sup>3,4,12</sup>; S.A. Gayther <sup>1,2</sup>; K. Lawrenson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Women's Cancer Program, Cedars-Sinai Medical Center, Los Angeles, CA.; 2) Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 3) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; 4) Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA; 5) Department of Obstetrics and Gynecology, Keck School of Medicine, University of Southern California, Los Angeles, CA; 6) Department of Pathology, University of Southern California, Los Angeles, California, USA; 7) Department of Computer Science, University of British Columbia, Vancouver, BC, Canada; 8) Department of Molecular Oncology, BC Cancer Agency, Vancouver, BC, Canada; 9) Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada; 10) Department of Gynecology and Obstetrics, University of British Columbia, Vancouver, BC, Canada; 11) McGraw/Patterson Center for Population Sciences, Dana-Farber Cancer Institute, Boston, MA, USA; 12) The Eli and Edythe L. Broad Institute, Cambridge, MA, USA

Epithelial ovarian cancer (OC) is the fifth most lethal cancer among women, with more deaths than any other gynecological cancer. Each ovarian cancer subtype (clear cell, endometrioid, high-grade serous and mucinous) is characterized by distinct molecular drivers, risk factors and clinical behaviour. Transcriptional regulation is highly histotype-specific. While each OC harbors around 10,000 non-coding somatic variants, the contribution of non-coding somatic mutations to OC development is currently not known.

We performed H3K27ac chromatin immunoprecipitation and transcriptome sequencing in primary tumors for the four different subtypes of invasive OC. Histotype-specific regulatory elements (REs) were enriched in enhancers ( $P < 0.001$ ). In silico prediction of putative target genes for histotype-specific REs identified genes (*WFDC2*,  $P = 5.5 \times 10^{-5}$ ) and pathways (PI3K-Akt signaling,  $P < 0.002$ ) known to be involved in OC development. Some genes (e.g. *PAX8* and *CA125*) are associated with super-enhancers (SEs) in all OCs, while others are histotype-specific, including *PPP1R3B* which is associated with SEs in clear cell OCs only. Knockdown of *PPP1R3B* in clear cell OC cells significantly reduced intracellular glycogen content, a signature feature of this histotype.

To identify frequently mutated regulatory elements (RE) we integrated the ovarian cancer epigenetic profiles with whole genome sequencing data from 232 OCs. We identified a cluster of nine single nucleotide variants in the *KLF6* promoter ( $P = 8.2 \times 10^{-8}$ ) at chromosome 10p15, and seven mutations in an enhancer near the *HIST1* gene cluster at 6p22.1. Stable knockout of a 635 bp region in the 6p22.1 enhancer induced downregulation of two of our predicted target genes, *ZSCAN16* and *ZSCAN12* ( $P = 6.6 \times 10^{-4}$  and  $P = 0.02$ ), and *ZKSCAN3* and *HIST1H2AI*, which are previously predicted targets of this

enhancer. Globally, there was an enrichment of mutations in active binding sites for TEAD4 ( $P=6 \times 10^{-11}$ ) and its binding partner PAX8 ( $P=2 \times 10^{-10}$ ), a known lineage-specific transcription factor in OC. In addition, REs associated with *PAX8* expression comprised the most frequently mutated set of enhancers in OC ( $P=0.003$ ).

In summary, we have characterized histotype-specific epigenomic and transcriptomic landscapes in OC and defined likely functional REs based on somatic mutation analysis of ovarian tumors. Taken together these data indicate that somatic mutations disrupt the PAX8 transcriptional network during tumor development.

## PgmNr 923: Hemimethylation analysis.

**Authors:**

S. Sun

[View Session](#) [Add to Schedule](#)**Affiliation:** Mathematics, Texas State University, San Marcos, TX.

---

DNA methylation is an epigenetic event that involves the addition of a methyl-group to a cytosine (C) site that pairs with a guanine (G) site (i.e., CG site). This event plays an important role in both cancerous and normal cell development. Previous studies often assume symmetric methylation on both DNA strands. However, asymmetric methylation, or hemimethylation (methylation that occurs only on one DNA strand), does exist and has been reported in several cancer studies. Due to the limitation of previous DNA methylation sequencing technologies, researchers could only study hemimethylation on specific genes, but the overall genome-wide hemimethylation landscape remains relatively unexplored. With the development of advanced next generation sequencing techniques, we can now measure methylation levels on both the forward and reverse strands at single CG sites in an entire genome. Analyzing hemimethylation patterns may potentially reveal genes or regions related to tumor growth. For our research, we identify hemimethylated CG sites in cancer by analyzing publicly available methylation sequencing data. Meanwhile, we also identify hemimethylation patterns by grouping consecutive hemimethylated sites based on their methylation states, methylation “M” or un-methylation “U”. These patterns include regular hemimethylation clusters (e.g., MMM on one strand and UUU on another strand) and polarity (or reverse) clusters (e.g., MU on one stand and UM on another strand). We then map these hemimethylation clusters and sites to corresponding genes and study the functions of these genes. Our results reveal that hemimethylation does occur across the entire genome and highly hemimethylated genes may influence tumor growth or suppression.

## PgmNr 924: Dissecting the chromatin landscape of relafusion driven ependymoma.

### Authors:

S.C. Mack<sup>1,2</sup>; A. Arabzade<sup>1,2</sup>; Y. Zhou<sup>1,2</sup>; J. Yi<sup>2,3</sup>; B. Deneen<sup>2,4</sup>; C.Y. Lin<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Baylor College of Medicine, Houston, Texas.; 2) Dan L Duncan Cancer Center, Houston TX; 3) Cell and Gene Therapy, Baylor College of Medicine, Houston, TX; 4) Human and Molecular Genetics, Baylor College of Medicine, Houston, TX

---

Ependymoma is the third most common pediatric brain tumor, with treatment to this date remaining surgical resection and radiation. Over 70% of supratentorial ependymoma are characterized by an oncogenic fusion between *C11ORF95* and *RELA*. ***C11ORF95-RELA* fusion is frequently the sole genetic driver detected in ST ependymoma, thus ranking this genomic event as a lead target for therapeutic investigation.** *RELA* is a transcription factor central to mediating NF- $\kappa$ B pathway activation in processes such as inflammation, cellular metabolism, and chemotaxis. Whether *C11ORF95-RELA* hyper-activates the NF- $\kappa$ B pathway during ependymoma development and is required for tumor maintenance is unclear. How *C11ORF95-RELA* protein expression directly modulates the epigenome and transcriptome is unknown. **We hypothesize that: (1)** *C11ORF95-RELA* establishes specific transcriptional programs through remodeling of distinct chromatin patterns, and **(2)** *C11ORF95-RELA* stability is required for tumor maintenance and represents a direct target for chemical and genetic screening. To address these hypotheses we utilized both chromatin profiling of *C11ORF95-RELA* paired with proteolysis targeting chimera (PROTAC) based perturbation experiments. We found that *C11ORF95-RELA* fusion: 1) Drives nuclear localization, 2) Associates with active chromatin regions, and 3) Essential for tumor growth. We conclude that ependymomas arise during development through a single oncogenic fusion event that remodels chromatin landscapes to promote tumorigenesis.



## PgmNr 925: Multi-omic approach highlights novel role for mismatch repair protein in Lynch syndrome.

**Authors:**

M. Devall; C. Dampier; S. Eaton; J. Bryant; G. Casey

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of Virginia, Charlottesville, VA.

---

**Introduction:** Lynch Syndrome (LS), is characterized by germline mutations in one of four key mismatch repair genes. This leads to a 50-70% lifetime risk for colorectal cancer (CRC) as well as a significantly increased risk for other cancers such as endometrial cancer and stomach adenocarcinoma. Furthermore, LS cancers are phenotypically distinct from cancer evolved from other molecular events. However, our understanding of LS is somewhat limited. Indeed, as of 2019 very few studies have investigated epigenetic alterations in LS patients, an analysis of which may provide additional insight into this complex syndrome and the pathways in which it is differentially regulated. To bridge this gap, here we utilize a novel 3D organoid model derived from human colon biopsies taken during routine colonoscopy to explore differences in non-cancerous cells between LS and healthy controls (CTL) at the methylomic and transcriptomic level.

**Methods:** Normal 3D colon organoids were generated from 7 LS and 9 control subjects and subjected to Illumina EPIC array, a platform that targets over 850k methylation sites, and RNA-Sequencing. Following standard analysis guidelines for both platforms, normalized gene expression levels and M-values were correlated to explore potential overlay between methylome and the transcriptome. Further, we employed WGCNA at both omic layers to identify common networks of dysregulation. TCGA data from CRC were used as a validation dataset.

**Results:** Here, we identify over 33,000 differentially methylated positions and over 20 spatially correlated differentially methylated regions (DMR). Importantly, our most significant DMR correlates to a mismatch repair gene not currently associated with LS. Further, we find a strong correlation between aberrant methylation and gene expression changes in matched samples. Finally, by using publicly available CRC data downloaded from TCGA we are able to validate a number of differences both at the methylation and expression level.

**Conclusions:** By utilising 3D organoid models, we not only identify significant differential methylation across 20 differentially methylated regions, but also highlight a role for a novel DNA mismatch repair protein in LS. Our study provides a new insight into the role of DNA methylation in LS and should warrant further investigation.

## PgmNr 926: Epigenetic features of repetitive elements in breast cancer development.

### Authors:

C. He <sup>1</sup>; N. Lin <sup>1</sup>; J. Castle <sup>1</sup>; J. Liu <sup>2</sup>; A. Shendre <sup>1</sup>; C. Wang <sup>2</sup>; Y. Liu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Kentucky, Lexington, Kentucky.; 2) Indiana University, Indianapolis, Indiana

---

Repetitive elements (REs) are rich regulatory factors in the human genome. Recent studies have reported that REs are involved in the epigenetic reactivation of oncogenes in tumorigenesis, and the global hypo-methylation of REs has been associated with cancer development in various cancer types. However, the role of the site-specific RE epigenetic features in cancer development is not well understood. In this study, we investigated the site-specific DNA methylation (DM) at REs and their influence on breast cancer development. For each ethnic group of African American (AA) and European American (EA) women, we generated genome-wide DM profiles for 17 tumor and 17 age-matched normal breast tissue samples, using the TruSeq Methyl Capture EPIC platform that covers 31,159 Alu and 25,374 LINE-1 elements. RE-specific DM levels were calculated as average methylation levels across CpGs in each RE sequence. We identified 18 REs (10 Alu and 8 LINE-1) and 64 REs (23 Alu and 41 LINE-1) were differentially methylated (*Bonferroni* corrected p-value < 0.05) between tumor and normal in EA and AA women, respectively. Race-specific differential methylation of REs was observed. Differentially-methylated Alu elements are enriched in signaling pathways for EA women and cell replication, enzyme regulation, and transcription regulation for AA women. Further, differentially-methylated LINE-1 elements are enriched in immune system activities for EA women and cell apoptosis and cell differentiation for AA women. To further understand the regulatory potential of RE methylation on gene expression, we conducted association analyses between site-specific RE methylation and gene expression. We observed both *cis*- and *trans*- regulation of the RE methylation on gene expression, and the identified genes were enriched in cancer-related pathways and networks. Our study suggests that the site-specific RE methylation may play a role in breast cancer development through regulating cancer-related genes and pathways. Future research is warranted to replicate these findings and understand the complex regulatory mechanisms related to REs.

## PgmNr 927: A pan-cancer approach to predict core transcriptional regulatory circuitries identifies important nodes in high-grade serous ovarian cancer.

### Authors:

K. Lawrenson<sup>1</sup>; J. Reddy<sup>1</sup>; M. Fonseca<sup>1</sup>; B. Abraham<sup>2</sup>; I. Klein<sup>3</sup>; L. Afeyan<sup>3</sup>; F. Segato<sup>1</sup>; R. Corona<sup>1</sup>; J. Seo<sup>4</sup>; D. Chaves-Moreira<sup>5</sup>; B. Karlan<sup>1</sup>; S. Gayther<sup>1</sup>; H. Noushmehr<sup>6</sup>; R. Drapkin<sup>5</sup>; M. Freedman<sup>4</sup>; R. Young<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cedars-Sinai Medical Center, 290 W 3rd Street, Los Angeles, CA.; 2) Department of Computational Biology, St. Jude Children's Research Hospital, Memphis, TN; 3) Whitehead Institute for Biomedical Research, Cambridge, MA, USA; 4) Dana-Farber Cancer Institute, Boston, MA, USA; 5) University of Pennsylvania, Philadelphia, PA, USA; 6) Henry Ford Hospital, Detroit, MI, USA

---

Lineage-specific transcriptional 'master' regulators can be co-opted to drive oncogenesis and thus represent potential therapeutic targets as they often are tumor-specific genetic dependencies. We therefore developed an approach to predict these transcription factors across 34 distinct tumor types by identifying those with high levels of cell-type-specific expression. As a proof-of-concept, we mapped the core regulatory circuitry (CRC) of master regulators in high-grade serous ovarian cancer (HGSOC) as it is the most lethal gynecologic malignancy. This analysis revealed *PAX8*, *SOX17*, and *MECOM* as putative CRC members, of which *PAX8* is a known urogenital regulator and ovarian cancer biomarker. Genes encoding these factors were found proximal to especially large super-enhancers in primary HGSOC tissues. Chromatin immunoprecipitation sequencing experiments in ovarian cancer cell lines confirmed that *PAX8*, *SOX17*, and *MECOM* co-occupy the genome at genes critical for ovarian cancer biology as well as at those encoding each putative master regulator; RNA-seq after knockdown of each factor identified largely overlapping target gene sets for *PAX8* and *SOX17*, with a convergence on cell cycle and angiogenesis pathways. *MECOM* fulfils some expectations of a master regulator – it is a lineage-specific essential gene, positively regulates *PAX8*, but *negatively* regulates *SOX17*; with *SOX17* and *MECOM* exhibiting opposing effects on critical epithelial and mesenchymal genes. In *in vitro* assays, knockdown of *PAX8* and *SOX17* reduces cell number and perturbs cell cycle progression, *PAX8* and *MECOM* knockdown impairs cancer cell clonogenicity, and *SOX17* knockdown increases apoptosis. Finally, ovarian cancer cells are sensitive to transcriptional inhibitors THZ1, JQ1 and THZ531, and *PAX8*, *SOX17* and *MECOM* are all downregulated following treatment with one or more of these small molecules, indicating a dependency on continuous high-level expression of these factors. In summary, our data suggest that *PAX8* and *SOX17* are important CRC factors in HGSOC, with largely overlapping but also factor-specific molecular and functional roles. Our data also implicate *MECOM* as a non-canonical master regulator in HGSOC. Contribution to the better understanding of the CRC governing HGSOC, as well as other tumor types, aids in predicting key molecular regulators and future therapeutic avenues in this devastatingly underserved disease.

## PgmNr 928: Association of DNA methylation marks with breast cancer risk in a Singapore population.

### Authors:

A. Lee <sup>1,2,3</sup>; P. Munusamy <sup>1</sup>; C. Chan <sup>1</sup>; G.L. Koh <sup>1</sup>; P. Ang <sup>4,5</sup>; M.H. Tan <sup>4,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cellular and Molecular Research, National Cancer Centre Singapore, Singapore; 2) Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore; 3) Office of Clinical and Academic Faculty Affairs, Duke-NUS Medical School, Singapore; 4) Department of Medical Oncology, National Cancer Centre Singapore, Singapore; 5) Oncocare Cancer Centre, Gleneagles Medical Centre, Singapore; 6) Institute of Bioengineering and Nanotechnology, Singapore; 7) Lucence Diagnostics Pte Ltd, Singapore

---

**Background:** Heritable methylation marks associated with breast cancer risk have recently been identified through epigenome-wide DNA methylation analysis of germline DNA samples from Australian multiple-case breast cancer families. Of the 24 methylation marks identified, four were replicated in an independent nested case-control study. **Aim:** The objective of this current study was to determine if these four methylation marks are associated with breast cancer risk in our Asian population. **Methods:** Germline DNA from 185 *BRCA*-negative breast cancer cases and 204 matched healthy controls were subjected to bisulfite conversion. Pyrosequencing was performed for the CpG sites cg18584561, cg01741999, cg03916490 and cg27639199 using the PyroMark (Qiagen) platform. Odds ratios (OR) and 95% confidence intervals (CI) for breast cancer risk were calculated using logistic regression based association analysis, with a *P*-value of  $\leq 0.05$  considered statistically significant. **Results:** Only one of the four methylation marks, cg03916490, was found to be associated with breast cancer risk in our Singapore cohort (OR = 1.0486, 95% CI = 1.0283 - 1.0702, *P* = 3.06E-06). The remaining three methylation marks cg18584561 (OR = 1.0041, 95% CI = 0.9844 - 1.0242, *P* = 0.686), cg01741999 (OR = 1.0043, 95% CI = 0.9836 - 1.0267, *P* = 0.687) and cg27639199 (OR = 1.0047, 95% CI = 0.9975 - 1.0121, *P* = 0.199) did not have statistically significant associations with breast cancer risk. **Conclusion:** Our findings suggest that replication studies in additional populations are necessary for newly discovered DNA methylation marks.

## PgmNr 929: Mechanisms by which mutant p53 drives metastatic triple negative breast cancer.

### Authors:

J. McDaniel<sup>1</sup>; Y. Zhang<sup>1,2</sup>; S. Xiong<sup>1</sup>; R. Morrissey<sup>1</sup>; D. Chachad<sup>1</sup>; X. Yu<sup>1</sup>; X. Su<sup>3</sup>; Q. Yu<sup>3</sup>; J. Wang<sup>3</sup>; X. Rao<sup>3</sup>; G. Lozano<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, UT MD Anderson Cancer Center, Houston, TX; 2) Department of Pharmaceutical Sciences Texas Southern University, Houston, TX; 3) Department of Bioinformatics and Computational Biology, UT MD Anderson Cancer Center, Houston, TX

---

Breast cancer is a heterogeneous disease comprised of four molecular subtypes defined by whether the tumor-originating cells are luminal or basal epithelial cells. Breast cancers arising from the luminal mammary duct often express estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth receptor 2 (HER2). Tumors expressing ER and/or PR are treated with anti-hormonal therapies, while tumors overexpressing HER2 are targeted with monoclonal antibodies. Breast tumors that do not express these proteins are known as “triple negative breast cancer” (TNBC) and are the most aggressive subtype with the highest mortality rates and no targeted therapy. There is a pressing need to identify regulatory factors in TNBC. 88% of human TNBCs harbor a *TP53* alteration yet the regulatory roles of mutant p53 are poorly understood. We will investigate the *in vivo* gain-of-function (GOF) mechanisms of mutant p53 involved in development of metastatic TNBC.

We have generated a somatic mouse model for p53R245W (murine counterpart to human p53R248W). The conditional allele expresses wild type p53, which converts to mutant p53 upon Cre recombinase activation. Ductal injection of adenovirus-Cre into the mouse mammary gland allows mutant p53 expression in a few cells surrounded by a normal stroma and immune system. Studies in our somatic breast cancer model indicate p53<sup>R245W</sup> mice give rise to metastatic triple negative tumors. We hypothesize that transcriptomic changes associated with p53R245W GOF mechanisms drive metastatic TNBC.

Using an integrated genomics approach, we will investigate how p53R245W contributes to breast cancer metastasis by ChIP-seq and RNA-seq to discover transcriptional co-factor(s) and target genes mediating mutant p53 GOF. We have queried the publicly available METABRIC human breast cancer dataset to identify genes differentially expressed between breast tumors harboring p53R248W/Q mutations compared to p53 null breast tumors. RNA-seq of p53R245W tumors from our model revealed upregulation of genes common to human breast tumors. Gene set enrichment analysis revealed enrichment of cell cycle pathways. This common gene signature will be used to predict co-factors mediating p53R245W GOF and perform ChIP-seq to define the transcriptome driven by mutant p53 in metastatic TNBC. Data from these studies will provide a better understanding of the GOF mechanism of the p53<sup>R245W</sup> hotspot mutation and its contribution to metastatic TNBC development.

## PgmNr 930: Enhancer-promoter chromatin interactions in endometrial cells reveals biologically relevant candidate target genes at endometrial cancer risk regions.

### Authors:

T. O'Mara; A. Spurdle; D. Glubb; Endometrial Cancer Association Consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

---

Genome-wide association studies (GWAS) have identified 16 genetic regions associated with endometrial cancer risk. As is typical of GWAS, the vast majority of endometrial cancer risk-associated variants at these regions are non-coding. To identify potential target genes of these endometrial cancer risk variants, we have generated chromatin looping data associated with the H3K27Ac histone mark (characteristic of promoters and enhancers) using HiChIP in an E6E7/hTERT immortalised glandular endometrial cell line and three endometrial cancer cell lines (Ishikawa, JHUEM-14 and ARK-1).

Stratified LD score regression of promoter-associated loops demonstrated significant enrichment of endometrial cancer heritability in the immortalised normal and two of the three cancer cell lines after conditioning on the baseline LD-model (3.83-8.84 fold,  $P_{\text{Bonferroni}} < 0.03$ ), highlighting the potential role of these loops in mediating the effects of GWAS risk variants.

Intersection of promoter-associated loops with endometrial cancer risk-associated variants identified 103 candidate target genes at 13 of the 16 endometrial cancer risk regions. Among these candidates there was an over-representation of targets for miR196a, a microRNA that was itself one of the 103 candidate target genes identified by our HiChIP data. Mining expression quantitative trait loci (eQTL) data from whole blood ( $N = 31,684$ ; Võsa *et al* BioRxiv 2018) or endometrial tumours ( $N = 336$ ; Lim *et al* PNAS 2018), we found expression of 22 of the 103 candidate genes to be associated with risk variants.

Using protein-protein interactions, coupled with functional annotation and “guilt-by-association” analysis, we prioritised 387 proteins that interacted with proteins encoded by the candidate target genes. This set was highly enriched for proteins encoded by known endometrial cancer driver genes (10-fold;  $P = 8 \times 10^{-9}$ ), suggesting that candidate target genes may affect endometrial cancer development through these interactions. Pathway analysis including the 387 prioritised interacting proteins and 103 candidate genes revealed enrichment of many cancer-related pathways, including “Pathways in Cancer” ( $P = 2 \times 10^{-56}$ ), as well as immune- and hormone-related pathways. Our study illustrates the power of H3K27Ac HiChIP chromatin looping analyses in identifying biologically relevant targets of risk associated variants, and provides avenues for translation of GWAS results.

## PgmNr 931: Aberrant C<sub>p</sub>G islands promoter hypermethylation in *Runx3* has potential as a diagnostic predictor in oral squamous cell carcinoma.

### Authors:

K.C. Onyegbula <sup>1</sup>; B.O. Emikpe <sup>2</sup>; C.I. Anumudu <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical Laboratory Science, Faculty of Basic Medical Sciences, College of Med, University of Ibadan, Nigeria, Ibadan, Oyo, Nigeria; 2) Department of Veterinary Pathology, Faculty of Veterinary Medicine, University of Ibadan, Nigeria; 3) Department of Zoology, Faculty of Science, University of Ibadan, Nigeria

---

**Introduction:** Oral squamous cell carcinoma (OSCC) is traditionally associated with extrinsic factors such as tobacco use and alcohol consumption. No such associations have been established in studies conducted in Nigeria so far. Therefore, OSCC development may involve molecular events, which information is largely lacking in Nigerian cohorts, and extrinsic factors. **Objective:** As such, this study to our knowledge is the first of its kind in Nigeria designed to identify molecular predictive diagnostic biomarkers for OSCC. **Methods:** A 12 year hospital-based retrospective analysis of patients' demographic data was conducted. A profile of aberrant C<sub>p</sub>G islands methylation in *Runx3* was done by methylation-specific polymerase chain reaction (MSP) with DNA extracted from 67 formalin-fixed paraffin-embedded (FFPE) blocks of OSCC tissue. Numerical data of C<sub>p</sub>G islands methylation in *Runx3* for gender, age, tumor location and histologic class were computed and presented as frequencies and percentages. Pearson's Chi-square test was used to assess association between gender, age, tumor location, histologic class and *Runx3* promoter hypermethylation. For all analyses,  $P \leq 0.05$  was considered significant. **Results:** *Runx3* promoter hypermethylation were observed in 45.0% (30/67) of the OSCC tumor samples, while 55.0% (37/67) were not methylated. No association was established between *Runx3* promoter hypermethylation with gender ( $P = 0.157$ ), age ( $P = 0.223$ ) and histologic class ( $P = 0.199$ ). A strong association was however observed between promoter hypermethylation of *Runx3* and tumor location ( $P = 0.010$ ). **Conclusion:** This study showed that aberrant C<sub>p</sub>G islands promoter hypermethylation of *Runx3* seem to be prevalent in OSCC, and that in combination with tumor location, may serve as a diagnostic predictor for OSCC.



## **PgmNr 932: The Ovarian Cancer Regulatory Atlas (OCRA): An integrative epigenomic-transcriptomic approach to characterizing ovarian cancer cells and their cells of origin.**

### **Authors:**

J. Plummer <sup>1</sup>; R. Corona <sup>2</sup>; B. Davis <sup>1</sup>; F. Dezem <sup>2</sup>; S. Chen <sup>1</sup>; K. Lawrenson <sup>2</sup>; S. Gayther <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Bioinformatics & Functional Genomics, Cedars Sinai Medical Center, Los Angeles, California.; 2) Women's Cancer Program, Cedars Sinai Medical Center, Los Angeles, California.

Ovarian cancer is a heterogeneous disease that can be subdivided pathologically into five histotypes: high-grade serous (HGSOC), low-grade serous (LGSOC), mucinous (MOC), endometrioid (ENOC) and clear cell (CCOC) carcinoma. HGSOC is thought to originate from secretory epithelial cells of the fallopian tube, while CCOC and ENOC appear to develop from endometriosis. The cell of origin for MOC is unknown, however MOC resembles gastric and intestinal adenocarcinoma. These histotypes differ not only in their cell of origin but also their genetics, risk factors, treatment and ultimately prognosis. Given gene expression is highly regulated and cell type specific, studies have attempted to distinguish these histotypes based on their transcriptional profiles. Despite consortia efforts (ENCODE) having shown that epigenomic landscapes of disease are also highly cell type specific and ultimately control gene regulation, a large scale atlas of the ovarian cancer epigenomic landscape has been missing from the field. Ovarian cancer risk variants identified through GWAS require evaluation in their relevant histotype-specific cell type to identify candidate causal genes. Hence we set forth to generate the Ovarian Cancer Regulatory Atlas (OCRA) to characterize the epigenomic and transcriptomic profiles of normal and ovarian cancer cells in a histotype specific manner to provide a landscape for functional annotation of risk SNPs, eQTL SNPs, TWAS SNPs and somatic mutations.

Here we describe the integrative analysis of >25 ovarian normal and cancer cell epigenomic profiles for histone modification patterns, DNA methylation and RNA expression. From integrative RNAseq analysis with ChromHMM/StatepaintR state calls, we have established global maps of regulatory elements, defined histotype specific regulatory modules and identified key regulators of these modules based on transcription factor motif enrichment and regulator expression. This atlas reveals that ovarian cancer genetic risk variants are enriched in cell type specific epigenomic marks and highlights OCRA as a valuable resource for interpreting the molecular basis for ovarian cancer risk and progression.



## PgmNr 933: Single cell transcriptomic and epigenetic characterization of inter and intra tumor heterogeneity in Glioblastoma.

### Authors:

A.M. Raman <sup>1,6</sup>; R. Raviram <sup>1,2</sup>; S. Preissl <sup>4</sup>; K. VanDeyenze <sup>3</sup>; X. Wang <sup>4</sup>; J.Y. Han <sup>4</sup>; R. Fang <sup>1,6</sup>; C. Chen <sup>5</sup>; B. Ren <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ludwig Institute for Cancer Research, University of California San Diego, La Jolla, CA, USA; 2) Department of Chemistry, University of California San Diego, La Jolla, CA, USA; 3) Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA, USA; 4) Center for Epigenomics, University of California San Diego, La Jolla, CA, USA.; 5) Department of Neurosurgery, University of Minnesota, Minneapolis, MN, USA; 6) Bioinformatics and Systems Biology Program, University of California, San Diego, La Jolla, CA 92093-0419, USA

---

Glioblastoma is the most common form of primary brain tumor in adults and remains one of the deadliest of human cancers. Intra-tumoral heterogeneity in glioblastomas remains a major challenge in therapeutic development. While single cell based RNA profiling has yielded insights into this heterogeneity, the resolution of these techniques is limited and the pertinence of RNA-based profiling to epigenetic landscape remains unclear. To better address these issues, we have developed and applied single nuclei ATAC-Seq (snATAC-Seq) methods to clinical glioblastoma specimens secured from distinct regions of a tumor and compared these profiles to those derived using single nuclei RNA-Seq (snRNA-Seq). We found that snATAC-seq could identify CNVs at the single cell level with high resolution (20KB), and these CNVs define distinct tumor sub-populations in different geographic locations of the tumor. Using aberrant CNVs as a marker for cells derived from the glioblastoma, we found significant representation of cells with aberrant CNVs that harbor transcriptomic profiles suggestive of normal oligodendrocyte progenitor cell and astrocytes, suggesting potential for these cells to be tumor cells of origin. Our results indicate a combination of snRNA-seq and snATAC-seq as a valuable molecular platform for characterization of glioblastoma heterogeneity.

## PgmNr 934: Open chromatin informed analyses of transcription factor cooperation in high-grade serous ovarian cancer.

### Authors:

F. Segato Dezem; R. I. Corona; R. Nameki; S. A. Gayther; K. Lawrenson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Cedars Sinai Medical Center, Los Angeles, CA.

---

Ovarian cancer is the most lethal gynecological cancer, with approximately 14,000 deaths per year in the US alone. High-grade serous ovarian cancer (HGSOC) is the most common histological subtype of ovarian cancer, characterized by specific genetic risk factors, biomarkers and cells of origin. Transcription factors (TFs) are critical players in gene regulation, are commonly deregulated in several cancer types, and are a major class of cancer cell dependencies. TFs are often studied in isolation, although they typically operate within complexes, with interactions with other DNA-binding proteins (including other TFs) often occurring in a context-dependent manner. We have identified putative master TFs in ovarian cancer histotypes and precursor cells, and in this study, our goal was to identify evidence of TF cooperation specific to HGSOC. Our approach is to use a motif-based approach to identify TF binding sites in open chromatin regions assayed using ATAC-seq on HGSOCs and normal fallopian tube epithelial cells, the major precursor for this tumor type. The tumor samples profiled were also included as part of The Cancer Genome Atlas cohort, for which gene expression, methylation and miRNA profiling is also available. We identified ~60-185,000 open regions per sample, ~11% of open regions coincided with promoters, the remaining peaks were categorized as putative enhancers (flanked by H3K27ac signal, a mark of active chromatin) or putative repressors (not flanked by H3K27ac). Within putative enhancers we observed an enrichment of motifs for factors already known to be associated with HGSOC, such as WT1, KLF5, FOXM1 and PAX8, plus novel factors not previously implicated in ovarian cancer (TGIF2, NR2F6). We focussed on PAX8, a known lineage-specific essential gene and attractive therapeutic target for ovarian cancer. We searched for evidence of TF cooperation using SpaMo, which detects significantly enriched spacings in a set of sequences between pairs of motifs. We identified hybrid binding sites for PAX8 together with SOX-family and TEAD-family TFs. We are now characterizing the expression of gene targets of the hybrid motifs during tumorigenesis. In parallel we are exploring small molecule and genetic disruption of the PAX:SOX interface targeting as a novel therapeutic strategy for HGSOC, a tumor type for which novel treatment approaches are urgently needed.

## PgmNr 935: Robust messenger RNA (mRNA) markers in blood for lung cancer diagnosis.

### Authors:

A.H. Skogholt <sup>1,2</sup>; E.D. Jensen <sup>2</sup>; E. Ryeng <sup>2</sup>; B.H. Grønberg <sup>2</sup>; F. Skorpen <sup>2</sup>; S.A. Schønberg <sup>2</sup>; P. Sætrum <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) K.G. Jebsen Center for genetic epidemiology, NTNU, Trondheim, Norway; 2) Department of Clinical and Molecular Medicine, NTNU, Trondheim, Norway; 3) Department of Computer and Information Science, NTNU, Trondheim, Norway

---

Lung cancer (LC) is the leading cause of cancer deaths worldwide with more than 1.6M deaths each year. The major risk factor for developing LC is cigarette smoking, but air pollution is also rapidly rising as a risk factor. Despite progress in therapies, the overall survival rate is low, with only 18.6% 5 year survival. These devastating numbers are mainly due to LC being diagnosed at late stages, for which few can be cured. Early detection of LC is crucial to achieve efficient treatment and to increase survival. However, there are limitations to current methods for early detection of LC as they are too invasive and represent potential health hazards, necessitating improved tools to detect LC at an early stage. Blood-based gene expression profiling is an alternative, or additional tool as blood samples are easily available, essentially non-invasive, and can be collected at a low cost.

We sampled blood from 123 LC patients at the St. Olavs university hospital, and 180 controls from two different biobanks sampled on two different blood tubes (PAXgene and Tempus) and used Illumina (HT-12 v4) microarrays to measure whole blood gene expression. We did three analyses: (i) finding biologically relevant gene expression differences between cases and controls, (ii) finding a robust set of genes to be used as a panel to identify presence of LC in blood, and (iii) identifying differences between patients with different pathological traits such as stage and histology. We collected phenotype data from questionnaires and hospital medical records, and evaluated the potential effects of CRP, tumor size, gender, and smoking habits.

By comparing cases and controls, we identified 382 significant genes (Bonferroni adjusted  $p < 0.05$ ) with biological relevance. When evaluating these on our test set, we found 51 genes that were robust as they were unaffected by either technical issues in sampling systems, biobank differences, gender, or pathological traits such as LC subtype or stadium. These findings were confirmed in three validation sets. By analysing the cases separately we found 6 genes distinguishing squamous cell carcinoma from other LC subtypes, and 6 genes distinguishing early from late stage LC patients. Pathway analyses and a literature survey indicated that the identified genes show biological relevance for cancer development and lung related diseases.

## PgmNr 936: *Prdm14* transcriptional output is a determinant of testicular germ cell risk.

### Authors:

N.J. Webster<sup>1,2</sup>; R.L. Maywald<sup>1,2</sup>; E.P. Dawson<sup>3</sup>; D.G. Lanza<sup>1,2</sup>; J.D. Heaney<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Dan L Duncan Comprehensive Cancer Center, Houston, TX; 3) NYU School of Medicine, Cell Biology, New York, NY

Genome wide association studies (GWAS) in humans have identified 25 Type II Testicular Germ Cell Tumor (TGCT) susceptibility loci that contain candidate risk genes with putative roles in germ cell biology or tumorigenesis. Functionalization of these GWAS discoveries can potentially lead to advances in screening of at-risk individuals and novel targeted therapies. Spontaneous TGCTs occur at a measurable frequency only in the 129 family of inbred mouse strains. Several single gene mutations modify TGCT susceptibility by enhancing or suppressing baseline incidence in 129/SvImJ mice. We propose that 129/SvImJ inbred mice can serve as a model to reveal risk genes within human TGCT susceptibility loci and to determine the mechanisms through which these genes influence tumor risk.

We used CRISPR/Cas9 genome editing directly in 129/SvImJ mice to generate a knockout allele of PR Domain 14 (*Prdm14*), a candidate GWAS risk gene for human TGCTs. *Prdm14* is a critical determinant of primordial germ cell specification during embryogenesis and an important epigenetic regulator of the naïve pluripotent state. We report that both partial and complete loss of *Prdm14* expression significantly increase tumor incidence in 129/SvImJ mice. Tumors in both heterozygous and homozygous *Prdm14* knockout animals originate during embryogenesis as pluripotent embryonal carcinoma cells (ECCs) that differentiate to form teratomas in adult testes. In homozygous *Prdm14* knockout embryos, the few germ cells that populate the testis express core components of pluripotency; however, most germ cells fail to initiate expression of the naïve pluripotency marker and germ cell licensing factor DAZL. By embryonic day (E)13.5, *Prdm14* deficient germ cells robustly express the primed pluripotency factor OTX2, likely in response to somatic cell-derived TGF-beta superfamily signaling. Moreover, *Prdm14* deficient germ cells begin to transform into foci of OTX2-positive ECCs by E13.5, two days prior to wild-type 129 embryonic testes.

We conclude that reduced transcriptional output of *Prdm14* promotes tumor initiation. We hypothesize that loss of *Prdm14* expression disrupts epigenome remodeling events necessary for regulating genes critical to establishing and maintaining germ cell lineage identity. In the absence of germ cell lineage restriction, *Prdm14* deficient “germ cells” are proposed to transform into primed pluripotent ECCs in response to TGF-beta superfamily signaling in the embryonic testis.

## PgmNr 937: An integrative ENCODE resource for cancer genomics.

### Authors:

J. Zhang <sup>1,2</sup>; D. Lee <sup>1</sup>; V. Dhiman <sup>4,5</sup>; P. Jiang <sup>6</sup>; J. Xu <sup>7,8</sup>; P. McGillivray <sup>1</sup>; H. Yang <sup>7,8</sup>; J. Liu <sup>2</sup>; W. Meyerson <sup>1,2</sup>; C. Clarke <sup>1,2</sup>; M. G <sup>1</sup>; S. Li <sup>1</sup>; S. Lou <sup>2</sup>; J. Xu <sup>1,2</sup>; L. Lochovsky <sup>1,2</sup>; R. Klein <sup>12</sup>; M. Snyder <sup>13</sup>; D. Gilbert <sup>14</sup>; K. Yip <sup>11</sup>; C. Cheng <sup>9,10</sup>; F. Yue <sup>7,8</sup>; X. Liu <sup>6</sup>; K. White <sup>4,5</sup>; M. Gerstein <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Computational Biology & Bioinformatics, Yale University, New Haven, CT 06520, USA; 2) Department of Molecular Biophysics & Biochemistry, Yale University, New Haven, CT 06520, USA; 3) Department of Computer Science, Yale University, New Haven, CT 06520, USA; 4) Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL 60637, USA; 5) Departments of Human Genetics and Medicine, University of Chicago, Chicago, Illinois 60637, USA; 6) Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute and Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA; 7) Department of Biochemistry and Molecular Biology, Penn State School of Medicine, Hershey, PA 17033; 8) Bioinformatics and Genomics Program, The Pennsylvania State University, University Park, Pennsylvania 16802; 9) Department of Biomedical Data Science, Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Lebanon, NH 03765; 10) Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Lebanon, NH 03765; 11) Department of Computer Science and Engineering, The Chinese University of Hong Kong; 12) Icahn Institute for Genomics and Multiscale Biology and Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 13) Department of Genetics, Stanford School of Medicine, Stanford, CA, USA; 14) Department of Biological Science, Florida State University, Tallahassee, FL, 32306-4295, USA

---

ENCODE comprises thousands of functional genomics datasets, and the encyclopedia covers hundreds of cell types, providing a universal annotation for genome interpretation. However, for particular applications, it may be advantageous to use a customized annotation. Here, we develop such a custom annotation by leveraging advanced assays, such as eCLIP, Hi-C, and whole-genome STARR-seq on a number of data-rich ENCODE cell types. A key aspect of this annotation is comprehensive and experimentally derived networks of both transcription factors and RNA-binding proteins (TFs and RBPs). Cancer, a disease of system-wide dysregulation, is an ideal application for such a network-based annotation. Specifically, for cancer-associated cell types, we put regulators into hierarchies and measure their network change (rewiring) during oncogenesis. We also extensively survey TF-RBP crosstalk, highlighting how SUB1, a previously uncharacterized RBP, drives aberrant tumor expression and amplifies the effect of the well-known oncogenic TF MYC. Furthermore, we show how our annotation allows us to place oncogenic transformations in the context of a broad cell space; here, many normal-to-tumor transitions move towards a stem-like state, while oncogene knockdowns show an opposing trend. Finally, we organize the resource into a coherent workflow to prioritize key elements and variants, in addition to regulators. We showcase the application of this prioritization to somatic burdening, cancer differential expression and GWAS. Targeted validations of the prioritized regulators, elements and variants using siRNA knockdowns, CRISPR, and luciferase assays demonstrate the value of the ENCODE resource.

## PgmNr 938: Extrachromosomal DNA function as mobile enhancers and global transcriptional amplifiers in cancer.

### Authors:

Y. Zhu <sup>1</sup>; A. Gujar <sup>1</sup>; H. Tjong <sup>1</sup>; C.H. Wong <sup>1</sup>; C.Y. Ngan <sup>1</sup>; L. Gong <sup>1</sup>; J. Liu <sup>1</sup>; H. Kim <sup>1</sup>; A.C. deCarvalho <sup>2</sup>; R. Verhaak <sup>1</sup>; C.L. Wei <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; 2) Department of Neurosurgery, Henry Ford Hospital, Detroit, MI, USA

---

Extrachromosomal, circular DNA (ecDNA) is emerging as a prevalent oncogenic alteration in cancer genomes, which frequently carries amplified oncogenes and plays important roles in tumorigenesis and progression by offering cancer cells selective growth advantages. While the presence of ecDNA and their structure information were characterized by standard genomic approaches, the mechanism(s) by which ecDNA are deployed to modulate cancer progression and to contribute to cancer drug resistance is not yet well understood. Here we leveraged the ChIA-PET chromatin interaction assay, which interrogates both general spatial chromatin organization and protein factor mediated long-range chromatin interactions, to advance the discovery of ecDNAs and to characterize genome-wide ecDNA-mediated chromatin contacts that functionally impact transcriptional programs in cancers. Data from RNA polymerase II-associated chromatin interactions in glioblastoma patient-derived neurosphere cultures suggested that ecDNA mediated extensive intra-extrachromosomal and widespread trans-chromosomal interactions. EcDNA-chromatin contact foci were characterized by broad and high-level H3K27ac signals converging predominantly on chromosomal promoters, and corresponding increased gene expression levels, suggesting a major regulatory role in genome-wide activation of chromosomal gene transcription in *trans*. Deciphering the chromosomal targets of ecDNAs revealed an association with actively expressed oncogenes spatially clustered within ecDNA-chromatin connectivity networks. Our results suggest that ecDNAs, beyond the manifestations of oncogene amplification, function as transcription-amplifying elements to activate oncogene expression in human cancers.

## PgmNr 939: Chromatin folding domain disruptions by somatic genomic rearrangements in human cancers.

**Authors:**

K. Akdemir

[View Session](#) [Add to Schedule](#)**Affiliation:** Genomic Medicine Department, MD Anderson Cancer Center, Houston, TX.

---

Genomic material within the nucleus is folded into successive layers in order to package and organize the long string of linear DNA. This hierarchical level of folding is closely associated with transcriptional regulation and DNA replication. Genes within the same folding domain demonstrate similar expression and histone-modification profiles. Therefore, boundaries separating different domains have important roles in reinforcing the stability of these domain-wide features. Indeed, domain boundary disruptions in human genetic disorders or human cancers lead to misregulation of certain genes, due to *de novo* enhancer exposure to promoters. However, the frequency of boundary disruptions in human cancers, and whether there are recurrently affected boundaries in specific cancer types, remains unclear. In this study, in order to understand effects and distributions of somatic structural variations (SVs) across chromatin folding domains, we utilized 288,457 high-confidence somatic structural variations from 2658 tumor-normal paired high-coverage whole genome sequencing datasets across 43 different tumor types. We comprehensively profiled SVs effects on the domain boundaries and the regulation of genes in human cancers. Our findings demonstrate that the impact of chromatin folding domain disruptions is substantially varied across tumor types. Notably, different types of structural variants affect distinct genomic regions such as deletions affect inactive chromatin regions whereas duplications predominantly affect active boundary sites. Complex rearrangements such as chromothripsis (involving concomitant SVs) are the most prevalent class of structural variations in cancer genomes which can drastically change chromatin folding maps in the cancer genomes. To test the effects of SVs on genome architecture in primary human tumors, we generated high-resolution Hi-C data from fresh frozen osteosarcoma samples. Indeed, we were able to observe the catastrophic effects of the complex rearrangements on chromatin folding patterns in tumor samples. Overall, our work highlights the role of structural variations on restructuring the chromatin organization and rewiring transcriptional programs in cancer cells.



## PgmNr 940: Abnormal X chromosome dosage causes replication delays and polyploidization.

### Authors:

A. Koren<sup>1</sup>; M. Edwards<sup>2</sup>; I. Sagi<sup>3</sup>; D. Egli<sup>2</sup>; N. Benvenisty<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York.; 2) Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University, Jerusalem, Israel.; 3) Department of Pediatrics, Columbia University, New York, New York.

---

We previously described the derivation of haploid human embryonic stem cells by artificial activation of unfertilized eggs. Haploid stem cells are a powerful tool for genetic screens and for investigating why normal human development requires a diploid genome. However, haploid cells spontaneously diploidize by endoreplication, due to unknown reasons thought to be related to aborted mitoses. We sequenced the genomes of proliferating haploid and isogenic diploid cells and inferred the dynamics of genomic DNA replication by copy number analysis. Strikingly, ~2% of the genome failed to normally replicate in haploids, being drastically delayed such that the duration of S phase was extended by ~50% and replication continued into mitosis. The genomic regions with replication delays were composed of “naked” chromatin and corresponded almost exactly to under-replicated regions in polyploid cells of the mouse placenta. Haploid cells have one X chromosome that is expressed at similar levels as autosomes, in contrast to female diploid cells in which one X chromosome is developmentally inactivated leading to reduced levels of gene expression. We show that diploid cells that have undergone X chromosome reactivation, and recently-diploidized cells that have yet to inactivate the X chromosome, have replication delays similar to haploids. X chromosome dosage is also altered in polyploid placental cells and in breast cancers, which are often polyploid. We conclude that abnormal X chromosome gene expression exerts a trans-effect on autosomal replication, causing replication delays that lead to endoreplication and polyploidization in normal, pathological, and experimental systems.



## **PgmNr 941: Harnessing the immune response in chronic hepatitis B to regulate the epigenome in hepatocellular carcinoma (HCC).**

### **Authors:**

N. Nguyen <sup>1</sup>; V. Upasani <sup>1</sup>; D. Thompson <sup>2</sup>; C. Bui <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Immunology, Pastuer Institute, Cambodia, Phnom Penh, Cambodia; 2) Wellcome Trust Sanger Institute — Cambridge Immunology Network, UK; 3) Functional Genomics, University of Medicine, Vietnam

---

Current treatments for HCC are limited and 5-year survival for all stages combined has been reported less than 10% and with a median survival of 10 months following diagnosis. In addition, the mortality rates of patients with HCC remained poor in comparison to patients with other leading causes of cancer. The main reason for the poor survival seen in patients with HCC is partly due to the late presentation in major patients. In addition, for HCC patients the availability of certain treatments, chronic hepatitis B (CHB) infection is one of the major risk factors for cirrhosis of the liver and hepatocellular carcinoma in developing country like Cambodia. The chromatin remodeling access in response of CHB to liver tissue has been linked to various immune cell populations and understand their role in the pathologies of the diseases. In this study, we detected large numbers of tumor specific effector T-cells in the blood and biopsy of HCC patients but fail to mount effective immune responses. Moreover, the presence of suppressive regulatory T-cells (Treg) inhibit anti-tumor immune function. A set of experiments was designed an attempt to increase our understanding on how CD8+ Treg may disrupt anti-tumor response and by what mechanisms they are induced. CD8+Treg infiltrating HCC demonstrated a suppressive phenotype. The co-culture of nai?ve CD8+ T-cells with tumour-conditioned mouse dendritic model induces a population of CD8+Treg through an IDO dependent mechanism. This population of induced T-cells could suppress via the CD39-adenosine pathway. The findings of the mechanisms involved in the induction of CD8+Treg by DC and the involvement of CD39 in the suppressive capacity of these novel T-cells, may guide the development of future immunotherapeutic in HCC.

## PgmNr 942: Competing endogenous RNA network signature in glioblastoma.

### Authors:

E. Toraih <sup>1,2</sup>; M. Fawzy <sup>3,4</sup>; H. Abdallah <sup>1,2</sup>; A. El-Wazir <sup>1,2</sup>; S. Hosny <sup>2,5</sup>; A. Matter <sup>6</sup>; M. Hussein <sup>7</sup>; M. Tantawy <sup>8</sup>; E. Rashed <sup>9,10</sup>; M. Khashana <sup>11</sup>; D. Badran <sup>2,3</sup>; N. Aly <sup>3</sup>; A. Ellawindy <sup>1</sup>; A. Shaalan <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics unit, Department of Histology and Cell Genetics unit, Histology and Cell Biology Department, Faculty of Medicine, Suez Canal University, Egypt; 2) Center of Excellence in Molecular and Cellular Medicine, FOM/SCU, Ismailia, Egypt; 3) Department of Medical Biochemistry and Molecular Biology, Faculty of Medicine, Suez Canal University, Ismailia, Egypt; 4) Department of Biochemistry, Faculty of Medicine, Northern Border University, P.O. Box: 1321, Arar, Saudi Arabia; 5) Histology and Cell Biology Department, Faculty of Medicine, Suez Canal University, Egypt; 6) Department of Neurological surgery, FOM/SCU, Ismailia, Egypt; 7) Ministry of Health and Population, Cairo, Egypt; 8) Hormones Department, Medical Research Division, National Research Center, Cairo, Egypt.; 9) Department of Computer Science, Faculty of Informatics and Computer Science, The British University in Egypt, Cairo 11837, Egypt.; 10) Department of Mathematics, Faculty of Science, Suez Canal University, Ismailia 41522, Egypt.; 11) Faculty of Medicine, Suez Canal University, Ismailia, Egypt

### Introduction

Glioblastoma multiforme (GB) is one notable example of miRNA-modulated neoplasms. Given its unique tissue-specific expression signature, proper miRNA profiling can help to discriminate from other types of brain tumors.

The study aimed (1) to evaluate the expression signature of a panel of miRNAs and targets as diagnostic and prognostic biomarkers in primary GB, and (2) to develop a validated GB-specific custom-designed miRNA assays.

### Methods

#### *Pilot study phase:*

Expression pattern of a bioinformatically-selected panel of seven miRNAs (miR-34a, -16, -17, -21, -221, -326, -375), ten genes (*E2F3*, *PI3KCA*, *TOM34*, *WNT5A*, *PDCD4*, *DFFA*, *EGFR*, *SOX2*, *OCT3/4*, *NANOG*), and four lncRNAs (*ROR*, *GAS5*, *MALAT1*, *H19*) was performed using qRT-PCR.

Immunohistochemistry staining for five proteins (VEGFA, BAX, caspase 3, P53, BCL2) was carried out. *MGMT* promoter methylation was assessed for molecular subtyping.

#### *Biomarker panel development phase:*

A comprehensive data analysis of bioinformatics databases, previous literature and commercially available pre-designed miRNA PCR arrays within the market were carried out to develop a putative GB-specific transcriptomic (96-well plate) panel.

### Results

A significant up-regulation of five miRNAs (miR-16, -17, -21, -221, and -375), four genes (*E2F3*, *PI3KCA*, *WNT5A*, *SOX2*), two lncRNAs (*ROR* and *H19*), and two proteins (VEGFA and BCL2), and down-regulation of miR-34a, five other genes (*DFFA*, *PDCD4*, *EGFR*, *OCT3/4*, *NANOG*) and two lncRNA (*MALAT1* and *GAS5*) in brain cancer tissues. ROC analysis revealed that miR-34a, miR-17, and *MALAT1* had the highest diagnostic performance. Over-expression of *ROR*, *H19* and VEGFA and down-

regulation of TOM34, MALAT1, BAX, and p53 had poor survival. Hierarchical clustering classified patients into four groups based on gene panel signature. Competitively endogenous RNAs network of the lncRNA-miRNA-mRNA was constructed.

Next, we have developed a highly enriched panel of 84 deregulated miRNAs that could have a potential role in GB pathogenesis. Functional analysis classified miRNAs into four distinct clades. Further GB specimens, primary cultured cells and sera were collected from different geographic locations and transcriptomic analysis is still in progress.

### **Conclusion**

The explored microRNA-target dysregulation could pave the road toward developing potential therapeutic strategies for GB. Application of the newly designed array, after its validation, can provide more accurate data about diagnosis and/or prognosis.

## PgmNr 943: Tissue relevant super-enhancer activity and epithelial ovarian cancer risk loci.

### Authors:

K.C. Vavra<sup>1,2</sup>; D.J. Hazelett<sup>2</sup>; J. Tyrer<sup>3</sup>; R.I. Corona<sup>1,2</sup>; N. Rodriguez-Malave<sup>2</sup>; J. Dennis<sup>3</sup>; J.H. Seo<sup>4</sup>; L. Li<sup>4,5</sup>; R. Drapkin<sup>6</sup>; B.Y. Karlan<sup>1</sup>; H. Long<sup>4,5</sup>; M.L. Freedman<sup>4,5,7</sup>; T.A. Sellers<sup>8</sup>; P. Pharoah<sup>3</sup>; S.A. Gayther<sup>2</sup>; K. Lawrenson<sup>1,2</sup>; Ovarian Cancer Association Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Women's Cancer Program, Cedars-Sinai Medical Center, Los Angeles, CA; 2) Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA; 3) Strangeways Research Laboratory, University of Cambridge, UK; 4) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA; 5) Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA; 6) Ovarian Cancer Research Center, Department of Obstetrics and Gynecology, University of Pennsylvania, Philadelphia, PA; 7) The Eli and Edythe L. Broad Institute, Cambridge, Massachusetts; 8) Department of Cancer Epidemiology, Moffitt Cancer Center, Tampa, FL

---

Super-enhancers (SEs) are large domains of active chromatin that typically regulate genes important for cell state. SE landscapes are reprogrammed during tumorigenesis to dysregulate the expression of key oncogenes and tumor suppressor genes. Many studies have now shown that variants discovered through genome-wide association studies are enriched in noncoding transcriptional enhancers; but currently, the role of SEs in disease susceptibility is less well understood. We evaluated the role of SEs, defined by H3K27ac chromatin immunoprecipitation sequencing analysis of epithelial ovarian cancers (EOC) and precursor cells, in mediating the genetic risk of EOC. We identified 311-1338 SEs per sample; these were enriched at genes associated with actin cytoskeleton organization and extracellular structural pathways (in normal cells) and cell signaling and cellular organization pathways in tumors. At 25/76 of candidate causal loci, multiple candidate causal alleles overlap with ovarian super-enhancers, suggesting SEs may be involved in mediating risk. Risk SNPs associated with an overall risk for EOC tend to be distributed across EOC, prostate, and breast cancer-relevant SEs. In contrast, histotype-specific risk-associated SNPs tend to be distributed across OC-relevant SEs. A risk locus at 3q25 marked by index SNP rs7651446, (odds ratio = 1.59,  $p = 1.5 \times 10^{-34}$ ), is tagged by 79 candidate causal risk SNPs, more than half of which coincide with SEs detected in EOC precursor cells (ovarian and fallopian tube epithelia) but not in ovarian cancer cells. We hypothesized that SNPs at the 3q25 locus modulate activity of SEs that regulate *in cis* the expression of genes involved in the identity of normal tissue precursors, and that these genes are downregulated during tumorigenesis. Preliminary experiments comparing EOC samples and precursor cell lines highlight two candidate genes at the locus, *SSR3*, and *TIPARP*, as candidate susceptibility genes regulated by the normal tissue SEs at this locus. We therefore established stable shRNA-mediated knockdown models for *SSR3* and *TIPARP* in normal EOC precursor cell lines to study the roles of these genes in ovarian cancer development. RNA-seq profiling identified an enrichment of cell cycle-associated genes following *SSR3* knockdown, and flow cytometry confirmed a G1 arrest. In summary, we have identified a potential role for super-enhancers at the 3q25 serous EOC risk locus and identified *SSR3* as a novel candidate risk gene.

## PgmNr 944: Epigenome-wide association study of whole blood from kidney cancer patients indicates hypomethylation of a *PCBD2* intronic region.

### Authors:

H. Ohmomo<sup>1</sup>; S. Komaki<sup>1</sup>; Y. Shiwa<sup>1,3</sup>; R. Otomo<sup>1</sup>; Y. Sutoh<sup>1</sup>; T. Hachiya<sup>1</sup>; K. Ono<sup>1</sup>; M. Satoh<sup>1</sup>; M. Sasaki<sup>1</sup>; E. Arai<sup>2</sup>; H. Fujimoto<sup>4</sup>; T. Yoshida<sup>5</sup>; Y. Kanai<sup>2</sup>; A. Shimizu<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Iwate Tohoku Medical Megabank Organization, Disaster Reconstruction Center, Iwate Medical University, 2-1-1 Nishitokuta, Yahaba-cho, Shiwa-gun, Iwate 028-3694, Japan.; 2) Department of Pathology, Keio University School of Medicine, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan.; 3) Department of Molecular Microbiology, Tokyo University of Agriculture, 1-1-1 Sakuragaoka, Setagaya-ku, Tokyo 156-8502, Japan.; 4) Department of Urology, National Cancer Center Hospital, 5-1-1 Tsukiji, Chuo-ku, Tokyo, 104-0045, Japan.; 5) Department of Clinical Genomics, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo, 104-0045, Japan.

Kidney cancer (KC) is the fourteenth most common cancer worldwide, the incidence of which accounts for approximately 4% of all cancers. Since it has no pathognomonic signs or valuable biomarkers, KC is usually found by chance during abdominal examination via palpation or medical examination for hypertension. Recent studies have shown that DNA methylation status can serve as biomarkers in the diagnosis of several diseases. However, epigenetic biomarkers for KC still remain unidentified; therefore, we aimed to identify blood-based DNA methylation biomarkers for KC.

For this, we performed targeted-bisulfite sequencing of whole blood samples obtained from 50 KC patients and 50 control individuals. Epigenome-wide association analysis for KC was then conducted. We found 6 significantly hypomethylated CpGs in KC blood cells (Bonferroni-corrected  $p < 1.59 \times 10^{-8}$ ). They were located in the second intron of the *PCBD2* gene in chromosome 5 at dense DNase I hypersensitivity sites. Moreover, they did not have transcription factor binding sites. Interestingly, the region exhibited high homology with the light strand of mitochondrial DNA in humans, and likely translocated from mitochondrial DNA as a cluster of 5 kilobases during the evolutionary process from Chimpanzees and Neanderthals. Cis-expression-quantitative trait methylation (cis-eQTM) analysis results using our multi-omics database, iMETHYL (<http://imethyl.iwate-megabank.org/>), also suggested that the DNA methylation profiles of 3 out of 6 CpGs significantly affected *TCF7* and *VDAC1* expression, the abnormal expression of which is related to the onset of cancer. Furthermore, protein-protein interaction analysis using the GPS-Prot database showed that the *PCBD2* protein interacted with 34 proteins, including products of the cancer-related gene, *HNF1A*, and proto-oncogene, *MET*. In summary, we identified 6 significantly hypomethylated CpGs in a *PCBD2* intronic region in KC whole blood samples. Therefore, blood-based examination of *PCBD2* methylation profiles has the potential to clinically support KC diagnosis.

## **PgmNr 945: A two-stage epigenome wide association study identifies novel pancreatic cancer susceptibility loci by leveraging public controls.**

### **Authors:**

Z. Wang <sup>1</sup>; Y. Lu <sup>2</sup>; M. Fornage <sup>3</sup>; L. Jiao <sup>4</sup>; D. Li <sup>5</sup>; P. Wei <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Department of Epigenetics & Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Smithville, TX; 3) Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX; 4) Department of Gastroenterology, Baylor College of Medicine, Houston, TX; 5) Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

---

To investigate the role of DNA methylation in pancreatic cancer, we conducted an epigenome-wide association study (EWAS) using DNA samples from peripheral blood and the Illumina Infinium EPIC BeadChip at over 850,000 CpG sites across the epigenome. In this two-stage EWAS study, 44 pancreatic cancer cases and 20 controls in the discovery stage and 23 cases and 22 controls in the validation stage from a case control study conducted at MD Anderson Cancer Center were matched by age, gender, BMI, diabetes and smoking status. We increased the number of controls in the discovery stage from 20 to 556 by integrating public data from the Framingham Heart Study. We successfully removed the batch effects between the two datasets from different sources, as shown by the visualization of unsupervised learning. We discovered and replicated six significantly differentially methylated CpG probes (DMPs) and three regions (DMRs) between cases and controls in the discovery and validation stages. Furthermore, one of the three DMRs, which is a non-coding RNA region, was replicated in an external prospective cohort of 13 incident cases and 26 matched controls from the Women's Health Initiative Study. By performing causal inference using bidirectional Mendelian randomization analysis, we found evidence of directional relationships of the associations between DMPs and pancreatic cancer. RNA-sequencing analysis also illustrated the functional consequences of DMPs/DMRs on the cancer risk.

## PgmNr 946: Transcriptome interactions between Epstein-Barr viruses and human host cells at single-cell resolution.

**Authors:**

D. Osorio; J. Cai

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Veterinary Integrative Biosciences. Texas A&M University , College Station, TEXAS.

---

Epstein-Barr virus (EBV), is a common human virus that affects people all over the world differently. EBV infection can be asymptomatic, or cause infectious mononucleosis as well as lymphoid or epithelial malignancies. Research oriented to identify factors underlying the outcome of the infection have focused on the viral life cycle as well as in the interactions between the host and the EBV by using RNA-seq data. However, the mechanisms underlying the infection outcomes remain unknown.

Using high-throughput droplet-based scRNA-seq technology, we measured the single-cell transcriptome profile of GM12878 and GM18502—two lymphoblastoid cell lines established by in-vitro infection with EBV of B cells of female donors with European and African ancestry, respectively. We compared the transcriptome profiles of 7,045 cells from GM12878 and 5,189 from GM18502 against the profile of 1,228 uninfected B cells from a peripheral blood sample, and evaluated the regulatory interactions between viral and host genes using a single-cell gene regulatory network (scGRN) method for cells in different cell cycle phases.

We found 255 and 299 differentially expressed genes when compared the expression profile of the uninfected B cells against cells of GM12878 and GM18502, respectively. Enrichment analysis showed that, for both cell lines, these genes are related with EBV infection and the B cell receptor signaling pathway; however, an association with the Kaposi-sarcoma related with a herpesvirus infection pathway was only recovered on the GM18502 cell line (q-value = 0.016). This result match with previous reports suggesting an increased susceptibility to develop cancer in African populations due to oncogenic viruses as EBV. Comparison of the scGRN between EBV transcriptomes showed that genes are expressed variably between cycle phases and across samples. When the host-EBV interaction was compared, we found evidence that the interaction is more active in GM18502 than in GM12878 where the association with EBV infection pathway decreases after G1. In GM18502, we found an association with the viral carcinogenesis pathway during G1, which may be associated with cancer development and the activation of the EBV infection pathway during all the cell cycle phases.

Together, our results provide the transcriptome profile of EBV and host interaction with an unprecedented resolution, allowing to identify several possible mechanisms underlying the different outcome of the EBV infection.

## PgmNr 947: The role of transposable element methylation in aging and breast cancer development.

### Authors:

J.R. Castle <sup>1</sup>; N. Lin <sup>1</sup>; J. Liu <sup>1</sup>; A. Shendre <sup>2</sup>; C. Wang <sup>1</sup>; Y. Liu <sup>2</sup>; C. He <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Markey Cancer Center, University of Kentucky, Lexington, Kentucky.; 2) Indiana University School of Medicine, Indianapolis, IN

---

Nearly half of the human genome is comprised of repetitive sequences, most of which are derived from transposable elements (TE). TE can move and replicate in the genome and contribute to genome instability through insertional mutagenesis, gene expression regulation, and recombination. It is suggested that the mobility of TE is regulated by DNA methylation. Previous studies found that global hypo-methylation of TE increases with age, likely leading to an increase in mobility of TE and thus an increased susceptibility to cancer development. Limited studies have been reported on site-specific TE methylation and their relation to aging and cancer development. To understand the role of site-specific TE methylation in aging and breast cancer risk, we generated genome-wide DNA methylation profiles for 571 breast tissue (462 healthy normal and 109 tumor) samples using the Illumina TruSeq Methyl Capture EPIC Sequencing technology. We examined the association between the methylation of each of 176,987 Alu/LINE-1 CpGs and chronological age in normal breast tissue, and identified 438 Alu/LINE-1 CpGs associated with age. Genes annotated to these 438 TE-CpG markers are enriched in functions regulating transcription frequency, tumor necrosis factor production, cellular proliferation, and apoptosis. We further investigated the behaviors of the identified age-related TE-CpG markers in tumor and normal breast tissue. In age-matched analyses, we found that on average the identified 438 Alu/LINE-1 CpGs were hypomethylated in tumor compared to normal breast tissue (mean  $\Delta\beta = -0.03$ ;  $p=2E-11$ ). Further analyses with RNA-seq data from these tissue samples suggested that these age-related TE are overexpressed in tumor tissue compared to normal tissue, supporting the increased mobility of TE in the tumorigenesis. In line with our previous observations from epigenetic aging research that tumor breast tissue ages at a faster rate compared to healthy normal tissue, the current study identified specific age-related TE and showed that hypo-methylation of these TE was correlated with their overexpression, and the latter could lead to increased mobility of these TEs for cancer development. Our study provided support of the role of methylation at the identified TE in the processes of aging and tumorigenesis. Our current study examined approximately 3% of all TE CpGs. Future research will focus on the imputation of missing TE CpGs and the understanding the underlying molecular mechanisms.



## PgmNr 948: Epigenetic regulation of the *miR-137* Host Gene CpG island that spans *miR-137* and adjacent downstream sequence is associated with head and neck cancer.

### Authors:

H. Celesnik <sup>1,2</sup>; B. Cizmarevic <sup>3</sup>; U. Potocnik <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculty of Medicine, Center for human molecular genetics and pharmacogenomics, University of Maribor, Maribor, Slovenia; 2) Faculty of Chemistry & Chemical Engineering, University of Maribor, Maribor, Slovenia; 3) Department of Otorhinolaryngology, Cervical and Maxillofacial Surgery, University Medical Centre Maribor, Maribor, Slovenia

---

**BACKGROUND:** Head and neck cancer, including oral squamous cell carcinoma (OSCC), is one of the most common malignancies worldwide. Despite numerous advances in diagnosis and treatment, OSCC prognosis is still poor due to its aggressive nature, which includes strong local invasiveness and lymph node metastasis. The 5-year survival rate is among the lowest in cancer patients at <50%. Elucidation of molecular mechanisms underlying this type of cancer is crucial for early detection and treatment and for increasing patient survival. MicroRNAs are endogenous small RNAs that control gene expression in various cellular and metabolic pathways. MicroRNA *miR-137* has been shown to suppress cell proliferation, migration and invasion, and its dysregulation has been observed in cancerous tissues. In OSCC, methylation of *miR-137* promoter has been linked to cancer development. However, methylation of the CpG island (CGI) that spans *miR-137* and adjacent downstream sequence has not yet been investigated in OSCC. We have set out to elucidate whether epigenetic regulation of this CpG island plays a role in head and neck cancer.

**PATIENTS AND METHODS:** Paired tumor and normal adjacent tissue samples were collected from 118 Head and neck cancer patients that were treated at the University Medical Centre in Maribor from 2016 to 2019. Specimens included oropharynx (36%), hypopharynx (21%), cavitas oris (19 %) and larynx (11%) primary tumors. Medical history and clinical data was collected for each patient, including tumor TNM classification, perivascular/perineural invasion, cancer recurrence and patient survival. The methylation levels of portions of the *miR137* Host Gene were determined by Combined bisulfite restriction analysis (COBRA) of bisulfite-modified tissue DNA and by bisulfite sequencing. Expression analyses were performed by qRT-PCR.

**RESULTS:** Methylation at various CpGs of the CpG island that contains *miR-137* and adjacent downstream sequence was significantly higher in cancer tissues compared to paired normal adjacent tissues ( $P < 0.05$ ). UCSC Genome Browser tracks reveal that this CGI contains various transcription factor binding sites and DNaseI hypersensitivity clusters. In agreement with our methylation data, *miR-137* expression was decreased in tumor tissues. The aberrant methylation levels in OSCC were examined in relation to patient survival, tumor grade, invasion and cancer recurrence.

## PgmNr 949: Three-day workflow and analysis pipeline for the low-input NGS Ion AmpliSeq™ Methylation Panel for cancer research.

### Authors:

A. Hatch <sup>1</sup>; L. Pickle <sup>1</sup>; H. Shin <sup>2</sup>; V. Mozhayskiy <sup>1</sup>; G. Lowman <sup>1</sup>; M. Andersen <sup>1</sup>; D. Berman <sup>3</sup>; R. Bennett <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Thermo Fisher Scientific, 5781 Van Allen Way, Carlsbad, CA 92008; 2) Thermo Fisher Scientific, 180 Oyster Point Blvd, South San Francisco, CA 94080; 3) Cancer Research Institute, Queen's University, 10 Stuart St, QCRI A302, Kingston, ON K7L 3N6, Canada

---

**Introduction:** Strong associations between DNA methylation patterns and clinical phenotypes can be used as biomarkers for diagnosing diseases and guiding treatment. Currently there are multiple ways to determine methylation status, including whole genome sequencing, microarrays and amplicon-based approaches. The recently released Ion AmpliSeq™ Methylation Panel for Cancer Research is the first of its kind and demonstrates the capability to create a targeted panel leveraging the AmpliSeq™ technology on the NGS Ion S5™ sequencing platform.

The panel combines an easy three-day protocol with manual or automated library options, low input (10 ng to 20 ng starting), and a flexible multiplexed approach with quantitative information at single base pair resolution. The bioinformatics analysis has been streamlined into a downloadable plugin providing DNA methylation calls on both Watson and Crick strands and methylated:unmethylated ratios for each CpG. Samples tested to date include standard fresh frozen and FFPE.

**Methods:** A complete three-day end-to-end workflow has been developed starting from bisulfite conversion and progressing through library construction, template preparation, sequencing and data analysis. A subset of clinically relevant markers from the BLUEPRINT consortium was used to create the panel, which contains a total of 40 amplicons, with 31 of those being cancer markers associated with colon cancer, prostate cancer, leukemia and lymphoma. Amplicons were designed for both original strands, but only one of those two was chosen for the final panel.

Two control gDNA samples were used in the evaluation of the panel. The expected average methylation state across all CpGs was >98% for the first sample and <5% for the second. Testing was also performed with an equal mixture of these two samples.

**Results:** The Ion AmpliSeq™ Methylation Panel for Cancer Research showed good performance on both the >98% and <5% samples as well as the equal mixture.

**Conclusions:** The AmpliSeq™ Methylation Panel for Cancer Research offers high resolution, targeted and quantitative methylation analysis of clinically relevant targets with a three-day end-to-end workflow from low DNA input. The potential exists to design custom methylation panels.

For Research Use Only. Not for use in diagnostic procedures. Early access materials were provided by

Thermo Fisher Scientific to Dr. David Berman in support of this abstract, but no remuneration.

# PgmNr 950: Epigenome-wide association study of the herbicide glyphosate and DNA methylation.

## Authors:

R.M. Lucia <sup>1</sup>; W. Huang <sup>1</sup>; K.V. Pathak <sup>2</sup>; M. McGilvrey <sup>2</sup>; A. Alvarez <sup>1</sup>; A. Ziogas <sup>1</sup>; D. Goodman <sup>1</sup>; A.O. Odegaard <sup>1</sup>; P. Pirrotte <sup>2</sup>; T.M. Norden-Krichmar <sup>1</sup>; H.L. Park <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Epidemiology, School of Medicine, University of California, Irvine, CA; 2) Translational Genomics Research Institute, Phoenix, AZ

## Background

Glyphosate, the most commonly used herbicide worldwide, is associated with endocrine disruption and potentially carcinogenesis, although the association with cancer is controversial. Exposure to many environmental toxicants has been linked to altered DNA methylation patterns in white blood cells, but the impact of glyphosate on the epigenome has not yet been evaluated.

## Methods

We conducted the first epigenome-wide association study of glyphosate exposure in a cohort of 182 postmenopausal women aged 45-65. Peripheral blood samples and two first-morning urine samples were collected. Glyphosate concentration was assessed in urine samples using liquid chromatography coupled to a triple quadrupole tandem mass spectrometer. DNA was extracted from white blood cells for quantification of DNA methylation using the Illumina Infinium MethylationEPIC BeadChip. For each CpG site, a linear model adjusted for age and race/ethnicity was fitted with creatinine-corrected glyphosate tertile as the predictor and methylation (M) value as the outcome. Sites with a false discovery rate  $q < 0.05$  were considered statistically significant.

## Results

178/182 (98%) of study participants had detectable glyphosate in their urine (LOD = 0.014 ng/mL). Glyphosate levels varied widely, with a mean of  $0.41 \pm 0.47$   $\mu\text{g}$  glyphosate/g creatinine, but did not significantly differ according to key confounding variables (including race/ethnicity, body mass index, smoking status, alcohol intake, and estimated white blood cell subtype distribution). DNA methylation at 156 CpG sites was significantly associated with glyphosate tertile. For most sites (125/156, 80%), methylation was inversely correlated with urinary glyphosate.

Glyphosate-associated sites were located in or near 115 genes, including *ESR1* and *IGF2*, which both have important roles in breast and other cancers. Gene ontology terms significantly enriched (adjusted  $p < 0.05$ ) for these 115 genes included regulation of transcription, Wnt signaling, and stem cell proliferation. Pathway analysis showed enrichment of several KEGG pathways, including proteoglycans in cancer and viral carcinogenesis.

## Conclusions

In a cohort of postmenopausal women, urinary glyphosate levels were associated with differential methylation at 156 CpG sites. Analysis of the genes found in or near these CpG sites indicated enrichment of cancer-related pathways. Further investigations are warranted to evaluate the effect of

glyphosate exposure on cancer risk.

# PgmNr 951: Genetically predicted methylation biomarkers and pancreatic cancer risk: A comprehensive study using genetic instruments.

## Authors:

L. Wu <sup>1</sup>; Y. Yang <sup>2</sup>; J. Zhu <sup>2</sup>; X. Guo <sup>2</sup>; X.O. Shu <sup>2</sup>; Q. Cai <sup>2</sup>; X. Shu <sup>2</sup>; W. Zheng <sup>2</sup>; J. Long <sup>2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Cancer Epidemiology Division, Population Sciences in the Pacific Program, University of Hawaii Cancer Center, University of Hawaii at Manoa, Honolulu, HI, USA; 2) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville, TN, USA

## Background

Existing epidemiologic studies on pancreatic cancer risk associated DNA methylation biomarkers are limited by relatively small sample sizes and potential biases caused by selection bias, confounders, and reverse causation. To overcome these limitations, we performed a comprehensive study to systematically investigate associations of genetically predicted DNA methylation with pancreatic cancer risk.

## Methods

We used genotyping and DNA methylation data obtained in white blood cells from subjects of European descent included in the Framingham Heart Study (FHS, N=1,595) and established models using the elastic net method to predict DNA methylation based on genetic variants. We further conducted model external validation using data from the Women's Health Initiative (WHI, N=883), and selected 62,994 models with a prediction performance ( $R^2$ ) of  $>0.01$  in both internal and external validation stages. For association analyses with pancreatic cancer risk, we used data from 8,280 cases and 6,728 controls of European ancestry included in PanScan and PanC4 consortia downloaded from dbGaP. For CpG sites showing a significant association with pancreatic cancer risk, we further evaluated correlations of their methylation levels with expression levels of genes flanking these loci in blood, by using data from the Offspring Cohort of the FHS (N=1,367). Furthermore, for genes whose expression levels were correlated with DNA methylation levels, we assessed whether their genetically predicted mRNA expression levels in blood were also associated with pancreatic cancer risk.

## Results

We identified 47 CpGs showing an association with pancreatic cancer risk at a false discovery rate (FDR)  $< 0.05$ , including 15 CpGs located  $>500$  kb away from any risk variant reported in previous GWAS of pancreatic cancer. Among these 47 CpGs, methylation levels of 14 were correlated with expression levels of eight adjacent genes at  $p < 0.05$ . Among these eight genes, two had genetic prediction models built for their mRNA expression levels. Gene *ABO* showed a significant association with pancreatic cancer risk ( $p = 2.85 \times 10^{-12}$ ), confirming previous findings.

## Conclusion

In this large association study, we integrated genetic variants, DNA methylation and gene expression

data and identified multiple novel DNA methylation biomarkers for pancreatic cancer.

## PgmNr 952: *NOTCH1* expression correlates to *miR-27a*, *miR-34a*, *miR-145*, *miR-150*, and *miR-335* in human squamous cell carcinoma of the oral cavity.

### Authors:

G. Grubelnik; E. Bostjancic; N. Zidar

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of Pathology, Faculty of Medicine, University of Ljubljana, Korytkova 2, 1000 Ljubljana

*NOTCH1* is an important regulator of stemness, involved in epithelial-mesenchymal transition (EMT), which can be an initial step in the development of cancer metastasis. *NOTCH1* expression has been documented in a variety of cancers, including head and neck squamous cell carcinomas (HNSCC), where it was associated with poor prognosis. In HNSCC, it is believed that it can act as a tumor suppressor gene. However, based on expression in different cell lines, it has been considered as tumor suppressor and as oncogene, suggesting that NOTCH1 signaling is dependent on many regulatory factors, including microRNAs. Different microRNAs have also been shown to be related to *NOTCH1* expression in various cancers, but not in oral squamous cell carcinomas (OSCC). The aim of our study was therefore to analyze the expression of *NOTCH1* and its correlations to the expression of *miR-27a*, *miR-34a*, *miR-150*, *miR-145* and *miR-335* in OSCC tumor samples.

Our study included tumor samples of 90 patients with HPV-negative OSCC that have been divided in 4 groups: 29 patients with pT1/2 N0 M0, 35 patients with pT1/2 N+ M0, 5 patients with pT3/4 N0 M0, 21 patients with pT3/4 N+ M0. Expression of *NOTCH1* and miRNAs (*miR-27a*, *miR-34a*, *miR-150*, *miR-145* and *miR-335*) has been analyzed using quantitative real-time polymerase chain reaction.

Analysis has shown that *NOTCH1* was in moderate positive correlation with expression of all investigated miRNAs; with *miR-27a* ( $r_s=0.250$ ,  $p=0.02$ ), *miR-34a* ( $r_s=0.290$ ,  $p=0.01$ ), *miR-145* ( $r_s=0.233$ ,  $p=0.04$ ), *miR-150* ( $r_s=0.428$ ,  $p=0.001$ ) and with *miR-335* ( $r_s=0.254$ ,  $p=0.02$ ). Expression of all the tested miRNAs was also in correlation to each other.

Our results show a correlation between the expression of the analyzed miRNAs and *NOTCH1* in tumor samples of OSCC, suggesting regulation of these miRNAs by NOTCH1 or vice versa. To the best of our knowledge, this is the first report on positive correlation of NOTCH1 with analyzed miRNAs in OSCC tumor samples.



## PgmNr 953: The landscape of long non-coding RNAs in pediatric cancers.

### Authors:

A. Modi <sup>1,2</sup>; G. Lopez <sup>2</sup>; S. Ramanan <sup>1,2</sup>; K. Conkrite <sup>2</sup>; D. Cheung <sup>1,2</sup>; M.E. Johnson <sup>2</sup>; E. Manduchi <sup>2</sup>; C. Su <sup>2</sup>; A.D. Wells <sup>1,2</sup>; S.F.A. Grant <sup>1,2</sup>; S.J. Diskin <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Pennsylvania, Philadelphia, PA.; 2) Children's Hospital of Pennsylvania, Philadelphia, PA, USA

---

**Background:** Long-non coding RNAs (lncRNAs) are emerging as key components of gene regulation in cancer. However, very few lncRNAs have been associated with pediatric cancers.

**Methods:** To characterize the lncRNA expression landscape of pediatric cancers, we analyzed RNA-sequencing data for 1,070 tumor samples from the Therapeutically Applicable Research Project (TARGET). This data set included three blood cancers: 282 acute myeloid leukemia (AML), 194 acute B-cell leukemia (B-ALL), and 244 acute T-cell leukemia (T-ALL), two kidney cancers: 128 Wilm's tumor (WT) and 61 rhabdoid tumors (RT), and 161 neuroblastoma (NBL). We performed a global analysis of lncRNA function using the lncMod method, in which expression data is modelled to identify lncRNA modulators that perturb transcription factor (TF) regulation of target genes. To further elucidate cancer-associated lncRNAs, we integrated ChIP-sequencing and chromatin interaction data, to identify lncRNAs regulated by TFs involved in the core transcriptional regulatory circuitry (CRC) of NBL and T-ALL.

**Results:** Our curated transcriptome catalog contains 14,636 known lncRNAs, of which 2087 were robustly expressed. We observed that 25-50% of all lncRNA expression is explained by approximately 5-10 highly expressed lncRNAs in each cancer. These included *MALAT1*, *GAS5*, *MEG3*, and *H19*, lncRNAs that are commonly dysregulated in various cancers. Across all the pediatric cancers, we observed 185,154 significantly dysregulated lncRNA-TF-target gene (lncMod) triplets. We assigned lncRNA modulators to hallmark cancer pathways via gene set enrichment analysis of their associated target genes. We observed significant enrichment for the MSigDB cell proliferation hallmark pathways (FDR < 0.01), across all the cancers, implicating 641 lncRNAs with this biological function. Analysis of lncRNA regulation revealed 449 and 80 lncRNAs regulated by cancer driver genes of the NBL and T-ALL CRC respectively; and included the *PVT1* and *CASC15* lncRNAs, both with known roles in driving these cancers. Differential expression analysis further prioritized the CRC-associated lncRNAs, *NR2F1-AS1* and *TBX2-AS1* in NBL, for functional validation. Silencing of *TBX2-AS1* with siRNAs (>80% knockdown) in an NBL cell line resulted in 50% decreased cell growth (p-value < 0.01) and elongated cell morphology.

**Conclusion:** Our results provide a comprehensive catalog of lncRNA expression and putative functional pathways for lncRNAs in pediatric cancers.

## PgmNr 954: Variable trametinib-driven epigenomic adaptation in basal-like TNBC cell lines.

### Authors:

K.A. Torkenczy<sup>1</sup>; E. Langer<sup>1</sup>; A. Fields<sup>1</sup>; T. Risom<sup>1,2</sup>; M. Turnidge<sup>1</sup>; J.W. Gray<sup>3,4,5</sup>; R. Sears<sup>1</sup>; A.C. Adey<sup>1,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oregon Health & Science University, OREGON, USA; 2) Stanford University, Department of Pathology, CA, USA; 3) Center for Spatial Systems Biomedicine, Oregon Health & Science University, Portland, OR, USA; 4) Knight Cancer Institute, Oregon Health & Science University, Portland, OR, USA; 5) Knight Center for Early Detection Advanced Research, Oregon Health & Science University, Portland, OR, USA; 6) Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR, USA

---

Triple negative breast cancers (TNBC: ER-,PR-,HER2-) constitute one sixth of invasive female breast cancer cases and are the most likely to develop resistance to treatment. This is largely due to the absence of successful targeted therapies, which are further hindered by the high intra-tumor heterogeneity associated with TNBC tumors. Cells in these tumors exist in distinct genetic and epigenetic populations that gain treatment resistance via clonal Darwinian evolution and through the epigenetic adaptation into drug tolerant persister (DTP) states. While breast cancer clonal evolution has been more thoroughly studied on the single-cell level, the underlying epigenetic heterogeneity and treatment induced cell-state transitions in TNBC tumors remain enigmatic. Single-cell Combinatorial Indexed Assay-for-Transposase-Accessible-Chromatin (sci-ATAC-seq) allows for the accurate description of the single-cell regulatory landscape through the mapping of promoter and enhancer accessibility and the identification of putative cell-state specific DNA binding transcription factor proteins. We carried out sci-ATAC-seq on four TNBC cell lines that were MEK inhibitor (trametinib) and DMSO treated (control), which revealed separate cell line specific MEK inhibitor driven DTP epigenetic states along with some regulatory features shared between the DTP cell states. This indicated that the epigenetic shifts were primarily cell line specific with some shared underlying epigenetic mechanisms for resistance. Using a recently-described framework for linking regulatory elements to target genes (Cicero), we identified commonly deregulated genes in cancer, such as EGR1, which were further confirmed via differential accessibility analysis between treated and control groups. Finally, we looked at the pseudotemporal ordering of cells to further understand the changing chromatin dynamics of cell lines along a treatment response curve. Next, we aim to understand the influenced pathways of cell line specific epigenetic shifts during DTP state acquisition.

## PgmNr 955: Unearthing a new wild-type p53 transcriptional signature *in vivo*.

### Authors:

S.M. Moyer <sup>1,2</sup>; Y. Qi <sup>3</sup>; A.R. Wasylishen <sup>1</sup>; N. Fowlkes <sup>4</sup>; X. Su <sup>3</sup>; G. Lozano <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, MD Anderson Cancer Center, Houston, Texas.; 2) Genetics and Epigenetics Program, The University of Texas MD Anderson Cancer Center UT Health Graduate School of Biomedical Sciences, Houston, Texas.; 3) Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center, Houston, Texas.; 4) Department of Veterinary Medicine and Surgery, MD Anderson Cancer Center, Houston, Texas.

---

Cell stress activates the tumor suppressor and transcription factor p53. This subsequently triggers downstream p53-dependent pathway activation through a myriad of p53 target genes. However, it has become apparent that focusing on a select handful of p53 target genes, out of the vast array of genes mediated by p53 transcriptional activation, may not only be too exclusive but could also be an inaccurate representation of transcriptional activity. By deleting *mdm2* and thus activating a p53 transcriptional program *in vivo*, we overlayed tissue-specific RNA-sequencing data with previously published p53 ChIP-sequencing and ATAC-sequencing data. We identified a new p53 transcriptional signature of 7 common differentially expressed genes: *Eda2r*, *Gtse1*, *Polk*, *Mdm2*, *Psrc1*, *Polk*, *Ccng1*, and *Zfp365*. We propose that these genes be added to the classically studied p53 target genes to serve as an *in vivo* signature for p53 transcriptional activation. Not only does this study provide novel insight into acute p53 transcriptional activity, but it can also serve as a resource for the field by providing invaluable p53-dependent tissue-specific transcriptomes.

## PgmNr 956: A novel method to detect genetic interactions associated with age of onset in a genome-wide association study.

### Authors:

J. Luyapan<sup>1</sup>; X. Ji<sup>2</sup>; X. Xiao<sup>3</sup>; D. Zhu<sup>2,3</sup>; E.J. Duell<sup>4</sup>; D.C. Christiani<sup>5,6</sup>; M.B. Schabath<sup>7</sup>; S.M. Arnold<sup>8</sup>; S. Zienolddiny<sup>9</sup>; H. Brunnstrom<sup>10</sup>; O. Melander<sup>11</sup>; M.D. Thornquist<sup>12</sup>; T.A. MacKenzie<sup>1,2</sup>; C.I. Amos<sup>1,2,3</sup>; J. Gui<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Quantitative Biomedical Science Program, Geisel School of Medicine, Dartmouth College, Hanover, 03755, NH, USA.; 2) Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, 03756, NH, USA.; 3) Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, 77030, TX, USA.; 4) Unit of Nutrition and Cancer, Catalan Institute of Oncology (ICO-IDIBELL), Barcelona 08908, Spain.; 5) Department of Environmental Health, Harvard School of Public Health, Boston, 02115 MA, USA.; 6) Department of Medicine, Massachusetts General Hospital, Boston, 02115 MA, USA.; 7) Department of Cancer Epidemiology, H. Lee Moffitt Cancer Center and Research Institute, Tampa, 33612 FL, USA.; 8) Markey Cancer Center, University of Kentucky, First Floor, 800 Rose Street, Lexington, 40508 KY, USA.; 9) National Institute of Occupational Health, 0033 Gydas vei 8, 0033, Oslo, Norway.; 10) Lund University, Laboratory Medicine Region Skåne, Department of Clinical Sciences Lund, Pathology, Lund, Sweden.; 11) Lund University, Department of Clinical Sciences Malmö, Sweden.; 12) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, 98109, WA USA.

---

Genome-wide association studies (GWAS) have proven successful in predicting genetic risk of disease using single-locus models; however, identifying single nucleotide polymorphism (SNP) interactions at the genome-wide scale is limited due to computational and statistical challenges. We address the computational burden encountered when detecting SNP interactions for survival analysis, such as age of disease-onset. To confront this problem, we developed a novel algorithm, called the Efficient Survival Multifactor Dimensionality Reduction (ES-MDR) method, which uses Martingale Residuals as the outcome parameter to estimate survival outcomes, and implemented the Quantitative Multifactor Dimensionality Reduction method to identify significant interactions associated with age of disease-onset. To demonstrate efficacy, we evaluated this method on two simulation sets to estimate the type I error and power. Simulations show that ES-MDR identifies interactions using less computational workload with adjustment for covariates. We applied ES-MDR on OncoArray Consortium data with 14,935 cases and 12,787 controls (SNPs = 108,254) to search over all two-way interactions to identify SNP interactions associated with lung cancer age-of-onset. We tested the best model in an independent data set from the OncoArray. Our experiment on the OncoArray identified many one-way and two-way models with a single-base deletion in the noncoding region of *BRCA1* (HR = 1.24,  $P = 3.15 \times 10^{-15}$ ), as the top marker to predict age of lung cancer onset. Through the results of our extensive simulations and analysis of a large GWAS study, we demonstrate that our method is an efficient algorithm that identifies genetic interaction models to predict survival outcomes.

## PgmNr 957: LDscore regression identifies novel associations between glioma and auto-immune conditions.

### Authors:

Q.T. Ostrom <sup>1</sup>; J. Edelson <sup>2</sup>; J. Byun <sup>1,2</sup>; Y. Han <sup>1,2</sup>; K.M. Walsh <sup>3,4</sup>; C.I. Amos <sup>1,2</sup>; M.L. Bondy <sup>1</sup>; GLIOGENE Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Section of Epidemiology and Population Sciences, Baylor College of Medicine, Houston, Texas.; 2) Institute for Clinical and Translational Research, Dan L. Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, Texas, United States; 3) Duke Cancer Institute, Duke University Medical Center, Durham, North Carolina.; 4) Department of Neurosurgery, Duke University School of Medicine, Durham, North Carolina.

### BACKGROUND

Prior epidemiological studies in glioma have identified 25 germline risk variants, as well as risk associations with exposure to ionizing radiation (which increases risk) and history of allergies and aspirin use (which decrease risk). In this analysis we LDscore regression, which leverages single SNP associations and known patterns of linkage disequilibrium (LD) to estimate the genetic correlation between phenotypes, to confirm prior associations as well as attempt to identify novel phenotype associations for traits not previously assessed that may improve genetic prediction for glioma.

### METHODS

Summary statistics for all glioma, GBM, and non-GBM were obtained from a prior meta-analysis conducted by Melin, et al. Summary statistics for 13 immune- and atopy-related traits were obtained from the prior case-control studies and the UK biobank. Data were filtered to include only SNPs with imputation INFO value  $>0.7$ , and minor allele frequency  $>0.01$ , excluding SNPs within the HLA region. Pairwise genetic correlation between these traits was generated using LDSC. Associations were considered significant at  $p < 0.05$

### RESULTS

Significant negative correlations were identified between glioma and ulcerative colitis ( $rg = -0.4039$ ,  $p = 4.91 \times 10^{-10}$ ), celiac disease ( $rg = -0.2028$ ,  $p = 1.18 \times 10^{-4}$ ), lupus ( $rg = -0.0956$ ,  $p = 0.0083$ ), and multiple sclerosis ( $rg = -0.5755$ ,  $p = 4.46 \times 10^{-9}$ ). These associations were generally consistent in both GBM and non-GBM. There was a significant correlation between both self-reported ( $rg = -0.102$ ,  $p = 0.0233$ ) and doctor diagnosed ( $rg = -0.116$ ,  $p = 0.0305$ ) hayfever/allergic rhinitis and GBM only.

### CONCLUSIONS

This analysis demonstrates a genetic basis for previously identified protective effect of allergic rhinitis on GBM, and identifies novel associations between multiple auto-immune traits and glioma. Further studies are necessary in order to confirm these associations and identify the mechanism through which increased immune activity may lower risk of glioma.

## PgmNr 958: European genetic ancestry associated with risk of childhood ependymoma.

### Authors:

C. Zhang <sup>1</sup>; H. Hansen <sup>2</sup>; J. Gonzalez-Maya <sup>2</sup>; I. Smirnov <sup>2</sup>; J. Wiemels <sup>3</sup>; K. Walsh <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology and Biostatistics, Univ California San Francisco, San Francisco, CA; 2) Department of Neurological Surgery, University of California San Francisco, San Francisco, CA; 3) Center for Genetic Epidemiology, University of Southern California, Los Angeles, CA; 4) Department of Neurosurgery, Duke University, Durham, NC

**Background:** Ependymoma is a histologically-defined central nervous system tumor most commonly occurring in children. Population-based differences in incidence by race/ethnicity exist within the US and worldwide, with individuals of European ancestry at highest risk. No large-scale genomic analyses of ependymoma predisposition have been conducted to date. We aimed to determine whether extent of European genetic ancestry is associated with ependymoma risk in US populations.

**Methods:** In a multi-ethnic study of Californian children (327 cases, 1970 controls), we estimated the proportions of European, African, and Native American ancestry among admixed Hispanic and African-American subjects and estimated European substructure among non-Hispanic White subjects using genome-wide data. We tested whether genome-wide ancestry differences were associated with ependymoma risk and performed admixture mapping to identify associations with local European ancestry.

**Results:** A 20% increase in European ancestry was associated with 1.31-fold higher odds of ependymoma among Hispanic and African-Americans (95% CI: 1.08-1.59,  $P_{meta}=6.7 \times 10^{-3}$ ), with a stronger association observed among Hispanic individuals (OR: 1.34, 95% CI: 1.09-1.67,  $P=6.2 \times 10^{-3}$ ). Among non-Hispanic Whites, European ancestral substructure was also significantly associated with ependymoma risk. Local admixture mapping revealed a peak at 20p13 associated with increased local European ancestry, and genotype association analysis in the region identified an association at rs6039499 that survived Bonferroni correction (OR=1.99; 95% CI=1.45-2.73;  $P=2.2 \times 10^{-5}$ ) but was not validated in an independent set of posterior fossa type A (PF-EPN-A) patients.

**Conclusions:** Inter-ethnic differences in ependymoma risk are recapitulated in the genomic ancestry of ependymoma patients, implicating regions to target in future association studies.

## PgmNr 959: Arsenic (+3) methyltransferase gene polymorphisms and arsenic metabolism in schistosomiasis associated bladder pathology in Eggua, Nigeria.

### Authors:

C.I. Anumudu <sup>1</sup>; A.T. Adeboye <sup>2</sup>; A.S. Adebayo <sup>2</sup>; H.O. Awobode <sup>3</sup>; R.D. Isokpehi <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cellular Parasitology Programme, Department of Zoology, University of Ibadan, Ibadan, Nigeria; 2) Cell Biology and Genetics Unit, Zoology, University of Ibadan, Nigeria, Ibadan, Oyo State, Nigeria; 3) Parasitology Unit, Department of Zoology, University of Ibadan, Ibadan, Nigeria; 4) College of Science, Engineering and Mathematics, Bethune Cookman University, Daytona Beach, Florida, United States

---

Exposure to toxic inorganic Arsenic (iAs) in urogenital schistosomiasis (UGS) endemic areas may confer increased risk for bladder cancer. The severity of the adverse effects of iAs however depends on its metabolism which is highly variable. Genetic polymorphism in AS3MT enzyme responsible for arsenic metabolism, accounts significantly for these variations. The aim of this study was to investigate the association between *AS3MT* gene polymorphisms and Arsenic metabolism in relation to schistosomiasis associated bladder pathology. In total, 119 individuals from Eggua community in Southwest Nigeria were screened for schistosomiasis and bladder pathology. A digital arsenic test kit was used to assess total urinary arsenic concentrations of participants and determine level of arsenic exposure. The SNP *AS3MT/MET287THR* T>C (rs1191439) was genotyped using AS-PCR method. Prevalence of schistosomiasis and bladder pathology in the sample population was 5% and 30.25% respectively. Of the participants diagnosed with bladder pathology, only 5.6% tested positive for UGS. Total urinary arsenic concentration observed ranged from 0 to >0.8mg/L (mean of 0.16mg/L) with 75% above the World Health Organization (WHO) limit of 0.05mg/L. Observed allelic frequencies were 0.96 and 0.04 for wild-type T and mutant C alleles respectively. There was no significant relationship between *AS3MT* SNP, arsenic metabolism and schistosomiasis associated bladder pathology. However, low allelic frequency of the C-allele may suggest increased enzyme reactivity as urinary arsenic concentrations were relatively high indicating reduced arsenic body burden. Logistic regression analysis also showed that high total urinary arsenic concentration was significantly associated with reduction in likelihood of inducing bladder pathology.



## PgmNr 960: Genetic correlations between lung cancer and biomarkers using LD score regression.

### Authors:

J. Byun<sup>1,2</sup>; J. Edelson<sup>1,2</sup>; Y. Han<sup>1,2</sup>; Q. Ostrom<sup>2</sup>; K. Walsh<sup>3</sup>; M. Bondy<sup>2</sup>; C. Amos<sup>1,2</sup>; INTEGRAL-ILCCO group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX; 2) Department of Medicine, Epidemiology and Population Sciences, Baylor College of Medicine, Houston, TX; 3) Department of Neurosurgery, Duke University, Durham, NC

A common approach in genetic association studies is to perform meta-analysis from several cohort-level genome-wide association studies (GWAS) of a single trait. Common diseases often show comorbidity with other traits and diseases, and understanding the genetic architecture that results in disease co-development may be more informative than studying individual phenotypes. Here, we examined the degree of overlap in the genetic contribution between lung cancer and biomarkers based on the genetic correlation.

Linkage disequilibrium (LD) score regression is an approach that aims to quantify the genetic correlation between phenotypes based on summary statistics from GWAS. This study aimed to identify possible biomarkers that are associated with lung cancer overall and histological subsets including adenocarcinoma, squamous cell carcinoma, and small cell carcinoma, within smoker versus nonsmoker subsets. GWAS summary statistics from OncoArray lung consortium data and the UK Biobank were used for multi-trait analyses and chromosomal regions related to cigarettes per day, smoking cessation, and initiation were removed to correct the potential confounding effects of smoking.

We identified numerous biomarkers showing high SNP heritability ( $h^2$ ), which is the proportion of phenotypic variance explained by all SNPs, and genetic correlation ( $r^2$ ) with lung cancer, but many effects varied according to whether or not SNPs associated with smoking were included or excluded from analysis.

Without removing smoking-related SNPs, both high concentrations of lactate and lower concentrations of acetate in the blood were genetically correlated with lung cancer. After removal of smoking-related SNPs, the presence of CD3 on double negative T-cells was the strongest association ( $r^2=0.54$ ,  $P=0.01$ ,  $h^2=0.38$ ) in never smokers, while the percentage of immunoglobulin-E in the B memory cells ( $r^2=-0.50$ ,  $P=0.04$ ,  $h^2=0.17$ ) and fatty acid lengths ( $r^2=-0.57$ ,  $P=0.04$ ,  $h^2=0.04$ ) were the strongest protective associations. Elevations of selected immunological biomarkers may identify a subset of nonsmokers at higher disease risk.

LD score regression provides an improved understanding of the genetic architecture between lung cancer and biomarkers. Mendelian randomization analyses will help elucidate the potentially causal relationship between immune-related biomarkers and risk of lung cancer.



## **PgmNr 961: LFSPRO: A risk prediction R package for probabilities of age-of-onset of multiple primary cancers and specific cancer types in families with Li-Fraumeni syndrome.**

### **Authors:**

J. Chen <sup>1</sup>; S. Shin <sup>2</sup>; E. Dodd <sup>1</sup>; G. Peng <sup>3</sup>; W. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Bioinformatics and Computational Biology, MD Anderson Cancer Center, Houston, TEXAS.; 2) Department of Statistics, Korea University, Seoul, South Korea; 3) Department of Biostatistics, Yale University, New Haven CT

---

Li-Fraumeni syndrome (LFS) is a rare autosomal dominant disorder associated with a germline *TP53* mutation that increases an individual's risk of a spectrum of cancers and multiple primary diagnoses. Penetrance estimations of multiple primary cancer (MPC) risk and cancer-specific (CS) risk associated with *TP53* germline mutations remains a challenge because of limited data available for statistically meaningful estimation, the difficulty of accounting for variables like previous primary cancer effect and ascertainment bias inevitable in rare diseases. We developed a non-homogenous Poisson process model to estimate the MPC penetrance from a large pediatric sarcoma cohort from MD Anderson Cancer Center (MDACC). We also developed a model under a competing risk framework that considers the pedigree structure into the penetrance estimation and corrects for ascertainment bias to address the *TP53* penetrance for three cancer subtypes, breast (BR), sarcoma (SA), and others (OT), on the same pediatric cohort. We have validated the risk prediction performance of the two sets of penetrance estimates via independent cohort datasets.

Using the validated penetrance estimates, we have updated our R package LFSPRO to calculate probabilities based on family cancer history for 1) whether the counselee is a *TP53* germline mutation carrier, 2) the counselee develops breast cancer, sarcoma or any other cancers as a first primary in future years, or 3) if the counselee already had a cancer, he/she develops a second primary cancer in future years. Incorporating LFSPRO into genetic counseling sessions could potentially facilitate clinical decision-making to both patients and their physicians. The accurate cancer risk assessment will benefit cancer survivors who desire better risk management of any future cancer development, as well as healthy individuals who are concerned about their future risk of cancers based on their family history. Genetic counselors and clinicians will be able to quantitatively assess counselee's risks in the future, provide evidence-based counseling for clinical actions and save evitable diagnostic resources.

## PgmNr 962: Association between obesity and the colorectal cancer prognosis using pathway activity inference approach from multi-omics data.

### Authors:

E.K. Choe <sup>1,2</sup>; S.Y. Kim <sup>1,3</sup>; M. Shivakumar <sup>1</sup>; D. Kim <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) 1Department of Biostatistics, Epidemiology & Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Department of Surgery, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, 06236, South Korea; 3) Department of Software and Computer Engineering, Ajou University, Suwon 16499, South Korea; 4) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA

---

Obesity is increasing worldwide, potentially influencing prognostic outcomes in colorectal cancer (CRC) patients. Obesity is reported to have an association with CRC progression and prognosis in many previous epidemiologic studies. There have been several studies to explain its biological interaction using respective genomic profiles, but the results are inconsistent, and the mechanisms are not fully elucidated. Recently, methods for integrating the multi-omics profiles are applied in disease association studies. Here, we performed the association analysis between obesity and CRC prognosis using the integrative effect of multi-omics data. Clinical information and genomic data, including gene expression, methylation and copy number variation, from colorectal cancer patients were downloaded from the Cancer Genome Atlas (TCGA) database. We ran linear regression with each genomic data for body mass index (BMI) to select multi-omic features associated with obesity. To integrate these multi-omic features associated with obesity, a directed random walk-based pathway activity inference method was used to generate an integrated pathway activity matrix. KEGG pathway was used as a reference panel. Association analyses between the integrated pathway activities and CRC survival, locoregional lymph node metastasis and distant metastasis were done. Among 254 CRC patients, 120 (47.2%) had positive lymph nodes, and 29 (11.4%) had distant metastasis. There were 46 mortalities during follow-up (median, 20.5 months). Based on FDR < 0.1, there were 13 pathways significantly associated with overall survival, 2 with lymph node metastasis and 3 with distant metastasis. Alcoholism pathway was found to be significant both for overall survival and lymph node metastasis, which is typically observed in obesity population. In addition, Glycan synthesis, which has a crucial role in distant metastasis, was a significant pathway both for overall survival and distant metastasis. Integrated pathway activities for BMI showed a significant relationship with CRC overall survival, lymph node and distant metastasis. The pathophysiological effects of the pathways might be the potential mechanism linking obesity and CRC progression. This study warrants further validation in a larger CRC patient population to demonstrate the validity of biomarkers for CRC survival prediction, further potential targets for intervention to improve CRC prognosis.

## PgmNr 963: Sequencing DNA repair genes in 5,545 men with aggressive and non-aggressive prostate cancer.

### Authors:

B.F. Darst<sup>1</sup>; T. Dadaev<sup>2</sup>; E. Saunders<sup>2</sup>; X. Sheng<sup>1</sup>; P. Wan<sup>1</sup>; L. Pooler<sup>1</sup>; L.Y. Xia<sup>1</sup>; S. Chanock<sup>3</sup>; S.I. Berndt<sup>3</sup>; S.M. Gapstur<sup>4</sup>; V. Stevens<sup>4</sup>; D. Albanes<sup>3</sup>; S. Weinstein<sup>3</sup>; V. Gnanapragasam<sup>5</sup>; G.G. Giles<sup>6</sup>; R.J. MacInnis<sup>6</sup>; R.L. Milne<sup>6</sup>; M.C. Southey<sup>7</sup>; M. Pomerantz<sup>8</sup>; T. Key<sup>9</sup>; L. Mucci<sup>10</sup>; W. Catalona<sup>11</sup>; R.A. Eeles<sup>2</sup>; F. Wiklund<sup>12</sup>; Z. Kote-Jarai<sup>2</sup>; D.V. Conti<sup>1</sup>; C.A. Haiman<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Epidemiology, University of Southern California, Los Angeles, California.; 2) The Institute of Cancer Research, London, U.K.; 3) National Cancer Institute, National Institutes of Health, Bethesda, Maryland, U.S.; 4) American Cancer Society, Atlanta, Georgia, U.S.; 5) Academic Urology Group, Department of Surgery and Oncology, University of Cambridge, Cambridge, U.K.; 6) Cancer Council Victoria, Melbourne, A.U.; 7) Monash University, Melbourne, A.U.; 8) Dana Farber Cancer Institute, Boston, Massachusetts, U.S.; 9) University of Oxford, Oxford, U.K.; 10) Harvard University, Cambridge, Massachusetts, U.S.; 11) Northwestern University, Evanston, Illinois, U.S.; 12) Karolinska Institute, Solna, S.E.

---

**Introduction:** Few common genetic variants have been associated with the aggressive form of prostate cancer (PCa). Aggressive PCa may be attributed to multiple rare high-risk coding variants. We investigated whether such pathogenic variants in 20 DNA repair genes (Pritchard et al., NEJM 2016) associate with risk of aggressive versus non-aggressive PCa. **Methods:** Participants were from stage 1 of a multi-stage case-only investigation of men of European ancestry, including 2,770 with aggressive disease (died from PCa, stage T4/M1, or were both T3 and Gleason score $\geq$ 8 at diagnosis) and 2,775 with non-aggressive PCa (T1/T2 and Gleason score $\leq$ 6), who had germline whole-exome sequencing. A subset of highly aggressive cases died from PCa or had metastatic disease at diagnosis (n=2,134). Variants analyzed were rare (MAF<0.01) and either pathogenic based on ClinVar or likely to disrupt the protein based on IMPACT rating. Logistic regression was used to test associations between the burden of rare variants and aggressive PCa, adjusting for age at diagnosis, country, and three principal components. Linear regression was used to investigate the burden of rare variants on age at PCa onset. **Results:** Of non-aggressive cases, 9.3% carried at least one pathogenic variant in the 20 genes, compared to 13.0% of aggressive cases (OR=1.40, P=1.4x10<sup>-5</sup>) and 13.4% of highly aggressive cases (OR=1.47, P=3.8x10<sup>-6</sup>). *BRCA2* had the strongest evidence of association, with 2.5% of non-aggressive cases, 4.4% of aggressive cases (OR=1.73, P=1.2x10<sup>-4</sup>), and 4.7% of highly aggressive cases (OR=1.87, P=2.6x10<sup>-5</sup>) carrying rare pathogenic variants. Rare pathogenic *PALB2* variants were found in 0.11% of non-aggressive cases, 0.65% of aggressive cases (OR=6.19, P=5.1x10<sup>-4</sup>), and 0.70% of highly aggressive cases (OR=6.44, P=6.4x10<sup>-4</sup>). Carrying one pathogenic variant among the 20 genes was associated with a 1.0-year younger age at PCa onset (P=2.1x10<sup>-4</sup>). Carrying one pathogenic variant in *NBN* was associated with a 7.2-year younger onset (P=0.003), *BRCA2* with a 1.4-year younger onset (P=0.004), and *ATM* with a 2.3-year younger onset (P=0.01). **Conclusion:** In this large sequencing study, rare pathogenic variants in DNA repair genes were more common in aggressive than non-aggressive cases and associated with age at PCa onset. Nearly 2,000 genes identified in stage 1 are being sequenced in stage 2 in 14,000 additional aggressive and non-aggressive cases, which is expected to identify novel aggressive PCa genes.

## PgmNr 964: Fine-mapping of two differentiated thyroid carcinoma susceptibility loci at 2q35 and 8p12 in Europeans and Oceanians.

### Authors:

J. Guibon <sup>1,2</sup>; P.-E. Sugier <sup>1</sup>; M. Karimi <sup>1</sup>; O. Kulkarni <sup>2</sup>; A. Boland-Auge <sup>3</sup>; J.-F. Deleuze <sup>3</sup>; E. Ostroumova <sup>4</sup>; A. Kesminiene <sup>4</sup>; M.-C. Boutron-Ruault <sup>1</sup>; P. Guenel <sup>1</sup>; F. De-Vathaire <sup>1</sup>; F. Lesueur <sup>2</sup>; T. Truong <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Paris-Saclay University, Paris-Sud University, CESP (Center for Research in Epidemiology and Population Health), Inserm, Villejuif, France; 2) Inserm, U900, Curie Institute, PSL University, Mines ParisTech, Paris, France; 3) National Center of Human Genomics Research (CNRGH), François Jacob Institute of biology, CEA, Paris-Saclay University, F-91057, Evry, France; 4) Radiation group, International Agency for Research on Cancer, Lyon, France

---

Differentiated thyroid carcinomas (DTC) are the most frequent type of thyroid cancers. DTC incidence is characterized by wide ethnic and geographic variation, with high incidence rates observed in Oceanian populations. Genome-wide association studies (GWAS) identified several DTC susceptibility loci including 2q35, 8p12, 9q22 and 14q13. While the 9q22 and 14q13 loci have been extensively investigated, signals at 2q35 and 8p12 have not been explored so far.

We investigated the role of the two later regions in the EPITHYR consortium (1,554 cases and 1,973 controls) and performed fine-mapping in Europeans, Melanesians and Polynesians to identified likely causal variants of DTC risk. Odds ratios (OR) testing the association between SNP and DTC were derived from a logistic mixed model. To identify likely causal variants, we employed several strategies, including conditional analyses to test for residual association after adjusting on the lead SNP and two Bayesian methods. The first one, Paintor, integrates association strength with genomic functional annotation data to improve accuracy in selecting plausible causal variants; the second one consists of a colocalisation analysis that integrates the tests of SNP-disease association and eQTLs data from thyroid tumour cells available in the PancanQTL database.

The SNPs previously reported to be associated with DTC in European GWAS at 2q35 and 8p12 were replicated while no significant association was found in Polynesians and Melanesians.

In Europeans, a variant located in intron 5 of *DIRC3* at2q35 was highlighted by the Bayesian methods with a posterior probability of causality (PP) of 0.84 with Paintor and a PP of 0.41 with the colocalisation analysis. The associated OR was 1.4 ( $p=1.6 \times 10^{-10}$ ). Although association with DTC in Melanesians and Polynesians was not significant, this SNP was also identified as potentially causal in these populations. This SNP is strongly associated with *DIRC3* and *IGFBP5* expression in thyroid tumour cells. At locus 8p12, the colocalisation analysis identified a SNP located in the promoter region of *NRG1* as potentially causal in the 3 ethnic groups ( $PP_{\text{Europeans}} = 0.67$ ,  $PP_{\text{Melanesians}} = 0.55$  and  $PP_{\text{Polynesians}} = 0.38$ ), with a significant association in Europeans ( $OR=1.2$ ,  $p=1.4 \times 10^{-3}$ ). This SNP is associated with *NRG1* expression.

Hence, we confirmed the role of 2q35 and 8p12 loci in DTC susceptibility.  
Furthermore, our fine-mapping data will help prioritizing variants for further functional studies.

## PgmNr 965: Polygenic risk score analysis in familial lung cancer.

### Authors:

Y. Han <sup>1,2</sup>; J. Byun <sup>1,2</sup>; J. Edelson <sup>1,2</sup>; J. Bailey-Wilson <sup>3</sup>; S. Pinney <sup>4</sup>; A. Schwartz <sup>5</sup>; M. Andrade <sup>6</sup>; C. Gaba <sup>7</sup>; D. Mandal <sup>8</sup>; R. Hung <sup>9</sup>; C. Amos <sup>1,2</sup>; The Genetic Epidemiology of Lung Cancer Consortium, Integrative Analysis of Lung Cancer Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, TX; 2) Medicine, Epidemiology and Population Sciences, Baylor College of Medicine, Houston, TX; 3) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 4) University of Cincinnati College of Medicine, Cincinnati, OH; 5) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 6) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 7) University of Toledo Dana Cancer Center, Toledo, OH; 8) Louisiana State University Health Sciences Center, New Orleans, LA; 9) Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, Ontario, Canada

Many studies have indicated that individuals with a family history of lung cancer are two- to three-fold more susceptible to lung cancer development than those without such a history. Familial cancer is defined as having two or more first- and second-degree relatives diagnosed with the same type of cancer. Genome-wide association studies (GWAS) have successfully identified thousands of associations for many complex traits and common disorders. One application of GWAS results is to predict an individual's risk of disease development using genetic propensities.

We used the summary statistics from previously published familial lung cancer GWAS and data from the Oncoarray lung cancer dataset for lung cancer cases with a family history (1,719 cases and 11,135 controls) to develop polygenic risk scores. We randomly selected 70% and 30% as discovery and validation data sets, respectively. The summary data were filtered by  $MAF > 0.01$  and  $INFO > 0.8$ . We applied different combinations of tuning parameters including genome-wide thresholds, LD clumping values, and weights in genetic risk score models. GWAS threshold P-values were set to  $5 \times 10^{-6}$ ,  $5 \times 10^{-5}$ ,  $5 \times 10^{-4}$ ,  $5 \times 10^{-3}$ , and,  $5 \times 10^{-2}$  and LD clumping  $r^2 = 0.2$ , 0.4, 0.6, and, 0.8, respectively. Polygenic risk scores were calculated by computing the sum of risk alleles, weighted by 1,  $\log(ORs)$ , Z, and  $Z^2$  scores from the summary statistics, respectively. We evaluated models by the area under the ROC curve (AUC) and odds ratios (OR) per 2 standard deviations.

We chose the combination of tuning parameters (threshold  $5 \times 10^{-6}$ , LD clumping  $r^2 = 0.2$ , weight  $Z^2$ ) that identified 6 SNPs since this combination had the best overall AUC in the polygenic risk score model. From the discovery stage, the AUC was 0.7591(95% CI=0.7460-0.7722) with an OR of 1.560(95% CI=1.370-1.775). The AUC was 0.7590(95% CI=0.7393-0.7787) with an OR of 1.501(95% CI=1.234-1.825) in the validation stage.

Identification and assessment of subgroups at high risk of lung cancer on the basis of genetic characteristics are critical. Polygenic risk score analysis can stratify individuals into risk categories based on their genetic variation and improve personalized interventions for precision genomics based medicine.

## PgmNr 966: Ethno-geographic germline protein-altering mutations in ATM predispose to lung adenocarcinoma.

### Authors:

Ji <sup>1</sup>; J. McKay <sup>2</sup>; S. Mukherjee <sup>3</sup>; D. Zhu <sup>4</sup>; J. Byun <sup>4</sup>; C. Amos <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dartmouth College; 2) World Health Organization, Lyon, France.; 3) Memorial Sloan Kettering Cancer Center; 4) Baylor College of Medicine

---

Germline mutations cause 3–10% of cancers diagnosed yearly. However, only a few germline mutations responsible for lung cancer pathogenesis have been identified. Here, by performing associations of germline mutations in two independent datasets with a total of 39,146 individuals of European ancestry and investigating isoform expression values in 433 samples, we identified large-effect associations with a *ATM* L2307F (rs56009889) germline mutation in lung adenocarcinoma. We found that L2307F carriers had a significant 5.23-fold increased risk for lung adenocarcinoma (LAD) in the discovery dataset ( $P=6.47 \times 10^{-9}$ ) and a 2.48-fold increased risk in the replication dataset ( $P=0.01$ ), and exhibited no association with the risk of lung squamous cell carcinoma or of small cell lung cancer in either dataset. Additionally, all *ATM*-L2307F homozygotes had lung adenocarcinoma ( $P=0.004$ ), suggesting a significant role for L2307F in LAD pathogenesis. Despite being very rare in the general European population, L2307F was more common in the Israeli population (MAF=0.023) with conferring lung adenocarcinoma risk differently in geographic populations. Further investigating the difference in the main ethnic groups in Israel, including Jews and Arabs, we observed L2307F was found in 8.68% (MAF=0.042) Jews from Israel and 8.53% Jews (MAF=0.045) who lived in other countries, though Arabs have very low frequency. Among L2307F carriers, the ORs for LAD risk were 7.86 in Israeli Jews ( $P=2.12 \times 10^{-6}$ ) and 3.40 for the Jews in other countries ( $P=0.005$ ). Since *ATM* has alternative splicing in tissues, we identified rs56009889 can transcript in two isoforms that highly expressed in normal lung tissue. The two isoforms are the only ones that can produce longest *ATM* protein and contain all of 4 *ATM* functional domains. While L2307F was predicted to enhance *ATM* protein aggregation and have a pathogenic effect on the *ATM* protein, it has not been previously associated with either lung cancer or ataxia-telangiectasia. Our results demonstrate L2307F increased lung adenocarcinoma risk with large effects among heterozygotes and with very high effects among homozygotes, indicating the importance of *ATM* for lung cancer pathogenesis. The findings may be beneficial for lung cancer panel screening and familial inheritance research, suggest targets for lung cancer treatment for individuals with mutations of *ATM*, and greatly advance our understanding of lung cancer etiology.



## PgmNr 967: Development of a longitudinal tissue repository from a prospective veteran colonoscopy screening cohort.

### Authors:

A.N. Madison <sup>1</sup>; M.C. O'Leary <sup>1</sup>; R.L. Whitley <sup>1</sup>; R. Thomas <sup>1</sup>; A.J. Bullard <sup>1</sup>; K.J. Sims <sup>1</sup>; B. Chesnut <sup>1,2</sup>; T.S. Redding, IV <sup>1</sup>; E.R. Hauser <sup>1,2</sup>; D. Provenzale <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cooperative Studies Program Epidemiology Center-Durham, Durham VA Health Care System, Durham, NC.; 2) Duke University, Durham, NC

---

Colorectal cancer (CRC) prevention and risk stratification efforts are greatly informed by further understanding the longitudinal variation in the DNA of individuals who progress from healthy to advanced neoplasia and cancer. The creation of a longitudinal tissue repository provides an opportunity to initiate these CRC progression analyses. We developed a longitudinal data and tissue repository from a prospective colonoscopy screening cohort of veterans. The Cooperative Studies Program (CSP) study #380 enrolled 3,121 asymptomatic veterans who underwent screening colonoscopy from 1994-1997 from 13 VA medical centers. Following participation in CSP #380, participants may have received colon-related procedures for routine clinical care. We sought to enhance the original study database with data and specimens from these clinical events. From a master list of CSP #380 participants, we utilized the VA electronic health records (EHR) to identify their exam dates from 1994 to 2009 and corresponding specimen identification numbers. Then each original site's pathology lab was contacted to determine whether these specimens were available for retrieval for research purposes. Following retrieval, specimens were recoded and stored at a central VA biorepository. We also utilized the VA EHR to capture medical history and physical exam information. Of the 13 original sites, 11 had specimens available for these participants. Specimen retrieval and storage has been completed for 8 sites with 3,404 tissues obtained from 680 participants. Due to local site specimen storage or destruction policies, 4 sites were only able to provide specimens from 2003-2009. An additional 4 sites were able to provide specimens across the full period of 1994-2009 (n=3,320; 641 participants). Among these 641 participants, 47.1% (n=302) have longitudinal specimens with both baseline specimens and specimens from any subsequent exam. There were 1,222 specimens preserved with Buoin's fixative, 1,580 with formalin-fixed paraffin (FFPE), and 602 in an unknown fixative. A DNA extraction pilot yielded mean DNA quantities of 8.32 ng/μL (SD= 8.94) in Buoin's and 66.42 ng/μL (SD= 42.65) in FFPE tissues. Continued development of this repository, which captures over 15 years of longitudinal data and specimens, will be a powerful tool for examining the CRC progression pathway. Details about the CSP #380 repository may be found here: <https://www.research.va.gov/programs/csp/cspeg/initiatives.cfm>.



## PgmNr 968: Breast cancer subtype GWAS in 1,312 Peruvian breast cancer patients.

### Authors:

K. Marker<sup>1,2</sup>; V. Zavala<sup>1</sup>; T. Vidaurre<sup>3</sup>; J. Navarro<sup>3</sup>; J.N. Vásquez<sup>3</sup>; R. Meza Florez<sup>3</sup>; S. Casavilca<sup>3</sup>; M. Calderon<sup>3</sup>; J.E. Abugattas<sup>3</sup>; H.L. Gómez<sup>3</sup>; H.A. Fuentes<sup>3</sup>; C.L. Monge Pimentel<sup>3</sup>; J.M. Cotrina<sup>3</sup>; S.P. Neciosup<sup>3</sup>; C.A. Castaneda<sup>3</sup>; L. Tamayo<sup>4</sup>; S. Song<sup>1</sup>; D. Cherry<sup>5</sup>; S. Huntsman<sup>1</sup>; D. Hu<sup>1</sup>; E. Ziv<sup>1</sup>; J. Zabaleta<sup>6</sup>; L. Fejerman<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of General Internal Medicine, Department of Medicine, Institute of Human Genetics, University of California San Francisco, San Francisco, California; 2) Division of Epidemiology, School of Public Health, University of California Berkeley, Berkeley, California; 3) Instituto Nacional de Enfermedades Neoplásicas, Lima, Peru; 4) University of Chicago, Chicago, Illinois; 5) University of California San Diego, San Diego, California; 6) Louisiana State University Health Sciences Center, New Orleans, Louisiana

**Background:** Breast cancer genome wide association studies (GWAS) have been mostly conducted in European populations. A previous GWAS study in Latinas found a protective variant on chromosome 6q25 (rs140068132), that is relatively common in populations with Indigenous American ancestry and almost absent in all other groups. In order to discover other population-specific variants, we conducted a GWAS in a sample of highly Indigenous American breast cancer patients from Peru. The median Indigenous American ancestry in the Peruvian breast cancer patients is high at 79%.

**Methods:** Blood samples and clinical data were collected from 1,619 women with breast cancer at the Instituto Nacional de Enfermedades Neoplásicas in Lima, Peru. Genotypic profiles for 1,307 women were generated using the Affymetrix Precision Medicine Research Array. Four major breast cancer subtypes were identified using immunohistochemical markers [luminal A, luminal B, triple negative (TN), and HER2 amplified]. Luminal A and Luminal B are hormone receptor positive subtypes and HER2 amplified and TN are negative. Genetic ancestry was determined using ADMIXTURE. Genome wide association analysis was conducted for luminal subtypes (luminal A and B) compared to non-luminal subtypes (TNBC and HER2 amplified) in PLINK with age at diagnosis and genetic ancestry as covariates. There were 865 women with luminal tumors and 357 women with non-luminal tumors. To confirm the relevance of the top GWAS hit, we analyzed tumor gene expression based on RNA-seq data by genotype in 47 samples.

**Results:** The top GWAS hit, rs34533620 and another SNP in LD with it, rs6459569, are located in the intron and coding region of the kinesin family member 13A gene (*KIF13A*), respectively, at 6p22 (OR = 2.75 and 2.34, respectively, 95% CI 1.85-4.08 and 1.62-3.37, P = 5.71\*10<sup>-07</sup> and 5.12\*10<sup>-06</sup>). RNA-seq data for *KIF13A* in a subset of 47 women showed that the expression level of this gene is higher in women with luminal tumors (N=28) compared to non-luminal tumors (N=16) (P=0.05). RNA-seq data for rs34533620 and rs6459569 showed the expression level of this gene is higher in women without the minor allele, G and C, respectively (P=0.13 and 0.03).

**Conclusions:** Our data suggests that a germline variant within the *KIF13A* gene may be associated with the risk of developing luminal breast cancer tumors compared to non-luminal tumors. This supports the importance of conducting genetic association studies in diverse samples.



## PgmNr 969: Validation of previously identified breast cancer SNPs in Black women.

### Authors:

J. Moore <sup>1</sup>; S. Reid <sup>1</sup>; B. Mautz <sup>1</sup>; F. Ye <sup>1</sup>; R. Fan <sup>1</sup>; X. Wang <sup>2</sup>; A. Tezak <sup>1</sup>; A. Weidner <sup>1</sup>; T. Pal <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt University Medical Center, Department of Epidemiology, 2525 West End Avenue, Suite 600, Nashville, Tennessee, 37203; 2) H. Lee Moffitt Cancer Center and Research Institute, Biostatistics and Bioinformatics Department, 12902 USF Magnolia Drive, Tampa, FL 33612

Young Black women bear a disproportionate burden of breast cancer (BC) compared to White women yet are underrepresented in clinical studies. Prior research suggests that 18% of the familial BC risk can be explained by common single nucleotide polymorphisms (SNPs); however, most research has been conducted on White women. The evaluation of BC SNPs can be used to assess BC risk for patients and physicians to make informed decisions about risk reduction options. Few studies have assessed the association of high-risk BC SNPs among Black women; therefore, the purpose of this study was to assess the association of previously identified high risk BC SNPs among a population-based sample of young Black women with BC.

Methods: Black women diagnosed with invasive BC at or before age 50 from 2009 to 2012 were recruited to the study through the Florida Cancer Registry. Clinical information was collected through data obtained from the state cancer registry supplemented by data abstracted from medical records. Saliva specimens were collected for SNP analysis on the Oncoarray platform. The SNPs included in the analysis were identified from previously published literature. Breast cancer subtype specific logistic regression models (BRCA+ vs. BRCA-, TNBC vs. Non-TNBC, BRCA+/TNBC vs. other patients, BRCA-/TNBC vs. other patients) were developed, odds ratios (OR), 95% Confidence Intervals (CIs) and uncorrected multiple comparison p-values ( $p < .05$ ) were compared to validate the statistical significance of the previously identified SNPs.

Results: Among 358 study participants, a total of 42 previously identified BC susceptibility SNPs were included in the analysis. Of the SNPs included in the analysis, SNP rs11242675 (OR 2.505; 95% CI: 1.267-4.953;  $p=0.008$ ), and SNP rs67397200 (OR 2.409; CI: 1.204-4.82;  $p=0.013$ ) increased the odds of BRCA-/Non-TNBC compared to other genetic and phenotypic subgroups. SNP rs67397200 (OR 0.34; CI: 0.163-0.71;  $p=0.004$ ) is suggestive of a lower odds of TNBC compared to Non-TNBC.

Conclusion: Our results suggest that a proportion of high-risk BC SNPs previously identified among White women also contributes to subgroup specific BC among Black women; however, it is likely that additional SNPs are yet to be identified among this understudied population. This information is needed to develop robust polygenic risk scores among Black women to estimate their individual risk for developing BC.

## PgmNr 970: Patterns of loss-of-heterozygosity (LOH) in tumor identifies deleterious germline variants in DNA damage repair genes in patients with lung cancer.

### Authors:

S. Mukherjee<sup>1</sup>; P. Srinivasan<sup>2</sup>; C. Bandlamudi<sup>4</sup>; M.D. Hellmann<sup>1</sup>; H. Rizvi<sup>1</sup>; M.E. Selvan<sup>5</sup>; D. Mandelker<sup>2</sup>; M. Zauderer<sup>1</sup>; M. Walsh<sup>1</sup>; Y. Kemel<sup>1</sup>; A. Zehir<sup>2</sup>; M.I. Carlo<sup>1</sup>; K. Cadoo<sup>1</sup>; K. Tkachuk<sup>1</sup>; Z. Steinsnyder<sup>1</sup>; S.M. Lipkin<sup>6</sup>; R.J. Klein<sup>5</sup>; M. Ladanyi<sup>2</sup>; M. Robson<sup>1</sup>; L. Zhang<sup>2</sup>; D. Jones<sup>1</sup>; D. Solit<sup>1,4,6</sup>; J. Vijai<sup>1</sup>; C. Rudin<sup>1,6</sup>; Z.H. Gumus<sup>5</sup>; Z. Stadler<sup>1</sup>; B. Taylor<sup>3,4</sup>; M. Berger<sup>2,4</sup>; C. Amos<sup>7</sup>; K. Offit<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York.; 2) Department of Pathology, Memorial Sloan Kettering Cancer Center, New York; 3) Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York; 4) Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York; 5) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York; 6) Department of Medicine, Weill Cornell Medicine, New York, New York; 7) Dan L Duncan Comprehensive Cancer Center Baylor College of Medicine Houston, TX

The bi-allelic inactivation model for pathogenic germline mutations is a widely accepted mechanism for tumor suppressor genes (TSG). We hypothesized that deleterious single nucleotide polymorphism (SNPs) in TSG could also be targeted by LOH in tumor leading to a cancer-promoting role. We employed an analytical framework by assessing tumor LOH for deleterious SNPs in TSG. 2686 lung cancer patients (pts) who underwent germline and somatic sequencing analysis for 468 genes by MSK IMPACT between January 2014 and May 2016 were analyzed. For each truncating or missense SNP predicted as deleterious/damaging (minor allele frequency <2% in genome aggregation database gnomAD), we assessed whether pts were heterozygous in their normal DNA, and what proportion of these with LOH were homozygous for the mutant allele in the tumor DNA. The background LOH rate was estimated using synonymous variants in the same gene. Statistical significance was computed using fisher exact test. Frequencies of the candidate SNPs were compared to population frequencies in gnomAD. We identified *ATM* *p.Leu2307Phe* and *RAD52* *p.Ser346Ter* as significantly enriched in lung cancer pts compared to gnomAD controls corrected for population stratification. 63 (2.3%) pts carried the *ATM* SNP, of whom 58 had lung adenocarcinoma. 29% pts showed LOH of the WT allele compared to 12% background LOH ( $p<0.0001$ ) whereas LOH of the mutant allele was in 12.7% pts. The frequency of the *ATM* SNP was higher in females (4.1% vs male 2.1%;  $p\text{-value}=0.03$ ), light-smokers (<5 pack years of smoking history (6.6%) vs smokers (2.8%)  $p=0.003$ ) and pts with *EGFR* oncogenic mutations [8.9% vs 3%,  $p=0.02$ ]. The pattern of LOH in tumor for 89 (3.3%) pts carrying the *RAD52* SNP varied based on somatic mutation profile. Pts with oncogenic *EGFR* mutations who carried the *RAD52* SNP showed 35% LOH of WT allele compared to 12% background LOH rate ( $p<0.01$ ). Interestingly, LOH of the mutant allele in *EGFR* driven tumors was observed in 14% compared to 45% in pts with oncogenic *KRAS* mutations. In lung adenocarcinoma patients, *RAD52* SNP carriers had a higher frequency of somatic *MYC* amplification (12% vs 5% in non-carriers,  $p=0.003$ ) and a poorer prognosis compared to patients carrying the wild type allele (adjusted HR=1.6, 95%CI=1.06-2.6;  $p\text{-value}=0.02$ ). In conclusion, an analytical framework for germline discovery by

assessing patterns of LOH in tumors identified novel deleterious lung cancer susceptibility alleles in DNA damage repair pathways.

## **PgmNr 971: Genetic evidence for a causal relationship between testosterone levels and cancer risk in females, but no effect in males.**

### **Authors:**

J. Ong <sup>1</sup>; J. An <sup>1</sup>; X. Han <sup>1</sup>; G. Chenevix-Trench <sup>1</sup>; P.M. Webb <sup>2</sup>; S. MacGregor <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, 300 Herston Road, Queensland, Australia.; 2) Department of Population Health, QIMR Berghofer Medical Research Institute, 300 Herston Road, Queensland, Australia.

---

Testosterone is the primary circulating androgen. Testosterone is produced in the testes of males although it is also produced at much lower levels in female ovaries. Previous observational findings have suggested that serum testosterone level might be linked with increased risk of developing cancer among both males (particularly prostate cancer) and females, however it is difficult to establish causality. Using large-scale phenotypic and genetic data from UK Biobank (UKB), we performed Mendelian randomization (MR) analyses to explore the relationship between serum testosterone levels and overall cancer risk. We extracted data for 46,155 adults with malignant cancer (excluding keratinocyte tumours) and 264,638 healthy controls of white British ancestry. Given the differing production of testosterone in males and females, we used recently released UKB testosterone data to derive separate instruments for MR using autosomal SNPs in cancer-free males (213 SNPs, total  $r^2=0.09$ ) and females (147 SNPs, total  $r^2=0.05$ ). Reflecting the sex-specific metabolism and production, the genetic correlation in testosterone levels across the sexes was very low (LD-score,  $rg=0.01$ ,  $p=0.7$ ).

Genetically-predicted elevated serum testosterone level was associated with overall cancer risk in females (OR 1.09; 95% C.I.[1.03, 1.15] per 1 SD increase in serum testosterone level), but not in males. Stratified analyses for individual female cancers showed most common cancers were not associated with testosterone, apart from breast (OR 1.18 [1.06, 1.31]) and endometrial cancers (OR 1.80 [1.49, 2.19]). Although testosterone is required for prostate cancer growth, there was no association with prostate cancer (OR 0.99 [0.90, 1.09]). We used a multivariable framework to adjust for BMI, measured levels of serum lipids and sex-hormone binding globulin; effect sizes for genetically predicted testosterone and breast cancer and endometrial cancer were essentially unchanged.

We replicated our UKB findings using two-sample MR using summary statistics from publicly available consortia data from BCAC (breast OR 1.15 [1.08, 1.22]) and PRACTICAL (prostate OR 1.01 [0.93, 1.09]). While Our UKB analysis found a small non-significant increase in risk of ovarian cancer with increased testosterone, this result was not borne out in OCAC (OR 0.89 [0.80, 0.97]). These powerful genetic analyses suggest a causal effect of testosterone on some hormone-driven cancers in females.

## PgmNr 972: Cervical somatic mutations detected in HPV18 and HPV45 precancerous lesions are related to disease progression.

### Authors:

M. Pinheiro <sup>1</sup>; M. Dean <sup>1,2</sup>; M. Schiffman <sup>1</sup>; B. Zhu <sup>1</sup>; M. Yeager <sup>1,2</sup>; Y. Xiao <sup>1</sup>; M. Cullen <sup>1,2</sup>; J.F. Bolland <sup>1,2</sup>; M. Steinberg <sup>1,2</sup>; T. Raine-Bennett <sup>3</sup>; T. Lorey <sup>4</sup>; P.E. Castle <sup>5</sup>; R.D. Burk <sup>5,6</sup>; N. Wentzensen <sup>1</sup>; L. Mirabello <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc., Frederick, MD; 3) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 4) Regional Laboratory, Kaiser Permanente Northern California, Oakland CA; 5) Department of Epidemiology and Population Health, at Albert Einstein College of Medicine, Bronx, NY; 6) Departments of Pediatrics, Microbiology and Immunology, and Obstetrics & Gynecology and Women's Health, at Albert Einstein College of Medicine, Bronx, NY

Invasive cervical cancer (ICC) is a heterogeneous disease caused by persistent human papillomavirus (HPV) infection. Worldwide, HPV16, HPV18 and HPV45 are responsible for 76% of ICC cases. For unknown reasons, ~90% of HPV infections clear and only a minority of women progress to precancer and subsequently ICC. Distinct somatic mutational profiles have been well-characterized for cervical cancer but had never been evaluated in precancers.

To investigate whether somatic mutations in the human DNA are related to disease progression, we evaluated exfoliated cervical cells from 2,991 women (plus 228 serial specimens) HPV16-, 18-, or 45-positive with benign transient infections ( $\leq$ CIN1), precancers (CIN2, CIN3, adenocarcinoma *in situ* [AIS]), or ICC in the NCI-Kaiser PaP prospective cohort.

We evaluated 1303 controls, 1532 precancers and 156 ICC cases using a custom Ion AmpliSeq gene-panel of 40 established gynecological somatic driver genes at ~600x coverage. We first compared the frequency of known recurrent somatic mutations (i.e. hotspots), reported in multiple TCGA ICC, in our controls, precancers and cancers, by histology and calculated the risk (odds ratio [OR]) of progression from a benign infection to precancer and cancer. We also looked for somatic mutations in the serial samples collected prior to the final diagnosis of precancer/cancer. We plan to integrate the human somatic data with viral genomic data to investigate the viral-host genetic interaction.

Somatic mutations were detected in earlier lesions at a lower frequency, and they were distinct for adenocarcinoma vs. squamous cell carcinoma. For HPV18+ and HPV45+ specimens, the prevalence of mutations was significantly higher in AIS (11%;  $p=2.3 \times 10^{-3}$ ; OR=4.4, CI=1.7-11.6) and adenocarcinoma (28%;  $p=6.4 \times 10^{-8}$ ; OR=13.4, CI=5.2-34.2), compared to the controls (3%). The mutations observed in the controls were distinct and the average allele fraction, a proxy for cellular clonal expansion, was significantly lower in controls (0.03) compared to precancers (0.08) and cancers (0.15) ( $p=6.7 \times 10^{-5}$ ). In all case samples with a mutation and a serial sample, we detected the same mutation in the samples collected from 1 to 4 years prior to diagnosis of precancer/cancer. This is the first study to investigate somatic driver mutations in cervical precancers and serial samples, and our findings suggest that these somatic mutations help drive the transition of

precancers to cancer.



## PgmNr 973: Genetic colorectal cancer risk variants are associated with increasing adenoma counts.

### Authors:

B. Sullivan<sup>1,2</sup>; X. Qin<sup>1</sup>; T.S. Redding IV<sup>1</sup>; Z. Gellad<sup>1,2</sup>; A. Stone<sup>3</sup>; D. Weiss<sup>4</sup>; A. Madison<sup>1</sup>; K. Sims<sup>1</sup>; C. Williams<sup>1,2</sup>; D. Lieberman<sup>5,6</sup>; E. Hauser<sup>1,2</sup>; D. Provenzale<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cooperative Studies Program Epidemiology Center-Durham, Durham VA Health Care System, Durham, NC; 2) Duke University, Durham, NC; 3) Cooperative Studies Program Pharmacogenomics Analysis Lab, Central Arkansas Veterans Health System, Little Rock, AR; 4) Perry Point VA Medical Center, Perry Point, MD; 5) VA Portland Health Care System, Portland, OR; 6) Oregon Health & Science University, Portland, OR

---

**Background:** Colorectal cancer (CRC) screening guidelines suggest genetic evaluation in those with more than 10 lifetime pre-cancerous colonic polyps, termed “adenomas.” However, the genetic basis of most individuals with high cumulative adenoma counts will remain unexplained. Further studies are needed to characterize low-penetrance germline factors in those with increased cumulative adenomas.

**Aim:** To investigate if known CRC or adenoma-risk single nucleotide polymorphisms (SNPs) are associated with increasing cumulative adenoma counts in a prospective **screening** cohort of Veterans.

**Methods:** The CSP #380 screening colonoscopy cohort includes a biorepository of selected individuals with baseline advanced neoplasia and matched individuals without neoplasia (n=612). Blood samples were genotyped using the Illumina Infinium Omni2.5-8 GWAS chip and associated cumulative adenoma counts were summed over 10 years. A corrected Poisson regression (adjusted for age at last colonoscopy, gender, and race) was used to evaluate associations between higher cumulative adenoma counts and 43 pre-specified CRC-risk SNPs or a subset of these SNPs shown also to be associated with adenomas in published literature. SNPs were evaluated singly or combined in a Genetic Risk Score (GRS). The GRS was constructed from only the eight literature adenoma-risk SNPs and calculated based on the total number of present *risk* alleles (0-2) summed across all SNPs per individual (both weighted for published effect size and unweighted).

**Results:** Four CRC-risk SNPs were associated with increasing mean adenoma counts ( $p < 0.05$ ): rs12241008 (gene: VTG1A), rs2423279 (BMP2/HAO1), rs3184504 (SH2B3), and rs961253 (FERMT1/BMP2), with *risk* allele risk ratios (RR) of 1.31, 1.29, 1.24, and 1.23, respectively. Only one of eight known adenoma-risk SNPs was significant in our dataset (rs961253; OR 1.23 per *risk* allele;  $p = 0.01$ ). An increasing weighted GRS was associated with increased cumulative adenoma counts (weighted RR 1.58,  $p = 0.03$ ; unweighted RR 1.03,  $p = 0.39$ ).

**Conclusion:** In this CRC screening cohort, four known CRC-risk SNPs were associated with increasing cumulative adenoma counts. Additionally, an increasing burden of adenoma-risk SNPs, as measured by a weighted GRS, was associated with higher cumulative adenoma counts. Future work will evaluate predictive tools based on a precancerous, adenoma GRS to better risk stratify patients during CRC screening, and compare to current CRC genetic risk scores.

## PgmNr 974: Genome-wide association study of differentiated thyroid cancer in European and Oceanian populations in EPITHYR consortium.

### Authors:

T. Truong <sup>1</sup>; F. Lesueur <sup>2</sup>; P.E. Sugier <sup>1</sup>; J. Guibon <sup>1,2</sup>; O. Kulkarni <sup>2</sup>; M. Karimi <sup>1</sup>; J.F. Deleuze <sup>3</sup>; A. Boland-Auge <sup>3</sup>; A.F. Herzig <sup>4</sup>; A.L. Leutenegger <sup>5</sup>; C. Rubino <sup>1</sup>; E. Ostroumova <sup>6</sup>; A. Kesminiene <sup>6</sup>; M.C. Boutron-Ruault <sup>1</sup>; P. Guénel <sup>1</sup>; F. de Vathaire <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) CESP Inserm, U1018, Paris-Saclay University, Paris-Sud University, Villejuif, France; 2) Inserm, U900, Curie Institute, PSL University, Mines ParisTech, Paris, France; 3) National Centre of Human Genomics Research (CNRGH), François Jacob Institute of biology, CEA, Paris-Saclay University, Evry, France; 4) Inserm, U1078, Université de Bretagne Occidentale, EFS, Brest, France; 5) Inserm, U1141, Paris University, Paris, France; 6) Radiation group, International Agency for Research on Cancer, Lyon, France

---

Differentiated thyroid cancer (DTC) is the most frequent endocrine malignancy. Its incidence varies considerably around the world with high incidence reported in some Pacific islands such as Hawaii, New Caledonia (NC), and French Polynesia. Ethnic differences in incidence have also been noted with for instance higher rates in Melanesians than in other ethnic groups in NC. The causes underlying these geographic and ethnic variations are still unknown and a role of environmental and genetic factors, as well as changes in screening practices are suspected. Few susceptibility loci have been identified so far and most of the recent genome-wide association studies (GWAS) have had limited sample sizes.

In order to identify new DTC susceptibility loci, we conducted a GWAS using data from the EPITHYR consortium that included subjects from 7 population-based case-control studies from Metropolitan France, Polynesia, New Caledonia, Cuba, and Belarus. A total of 3,527 individuals of European ancestry (1,554 cases/1,973 controls) and 649 individuals of Oceanian ancestry (301 cases/348 controls) were genotyped using the OncoArray (Illumina). After imputation, the analyses were based on about 9 million SNPs in Europeans and 7 million SNPs in Oceanians. Odds ratios (OR) were calculated using logistic mixed models with the genetic relationship matrix as random effect, adjusted for age, sex, and study. Analyses were stratified by sex, ethnic group, age group (<50, ≥50 years-old), histology (papillary/follicular), and size of carcinoma (<10mm/≥10 mm).

Our results confirmed the susceptibility loci reported previously at *2q35*, *8p12*, *9q22* and *14q13*. While no significant loci were reported in Oceanian populations, we reported a new significant locus at *1p31* in Europeans (OR= 0.52,  $p=3.10^{-8}$ ), which was more striking in subjects younger than 50 years. In subgroup analyses, we observed another significant signal at *1p32* in the oldest group and in cases with a microcarcinoma. Interestingly, the *1p31* lead variant was also associated to the expression of *NFIA* in thyroid tumour cells, a gene that is involved in the regulation of cell proliferation. The *1p32* variant is correlated with a SNP in *FAF1*, a gene involved in apoptosis that had previously been associated with DTC risk in a candidate gene study. Replication analyses are ongoing and will be presented.

In conclusion, this study provides further insights into the genetic susceptibility of DTC.

## PgmNr 975: Exploring the causal role of the human gut microbiome on colorectal cancer: Application of Mendelian randomization.

### Authors:

K.H. Wade <sup>1,2</sup>; L.H. Nguyen <sup>3</sup>; A.D. Joshi <sup>3</sup>; R. Carreras-Torres <sup>4</sup>; A.T. Chan <sup>3</sup>; M. Gunter <sup>5</sup>; Y. Lin <sup>6</sup>; G. Casey <sup>7</sup>; J.C. Figueiredo <sup>8</sup>; S.B. Gruber <sup>9</sup>; J. Hampe <sup>10</sup>; H. Hampel <sup>11</sup>; M.A. Jenkins <sup>12</sup>; T.O. Keku <sup>13</sup>; U. Peters <sup>14</sup>; C.M. Tangen <sup>15</sup>; A.H. Wu <sup>16</sup>; D.A. Hughes <sup>1,2</sup>; M.C. Rühlemann <sup>17</sup>; J. Raes <sup>18,19</sup>; N.J. Timpson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MRC Integrative Epidemiology Unit at University of Bristol, Bristol, BS8 2BN, UK; 2) Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, BS8 2BN, UK; 3) Massachusetts General Hospital, Boston, MA 02114, USA; 4) IDIBELL Bellvitge Biomedical Research Institute, Gran Via de L'Hospitalet, 08908, Barcelona, Spain; 5) International Agency for Research on Cancer, 150 Cours Albert Thomas, 69372 Lyon CEDEX 08; 6) Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109, USA; 7) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; 8) Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 9) Department of Preventive Medicine, USC Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA; 10) Department of Medicine I, University Hospital Dresden, Technische Universität Dresden (TU Dresden), Dresden, Germany; 11) Division of Human Genetics, Department of Internal Medicine, The Ohio State University Comprehensive Cancer Center, Columbus, Ohio, USA; 12) Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia; 13) Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, North Carolina, USA; 14) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 15) SWOG Statistical Center, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 16) University of Southern California, Preventative Medicine, Los Angeles, California, USA; 17) Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Kiel, Germany; 18) Department of Microbiology and Immunology, Rega Instituut, KU Leuven, University of Leuven, Leuven, Belgium; 19) Center for Microbiology, VIB, Leuven, Belgium

---

Background: Recent small-scale epidemiological and experimental studies have provided evidence supporting an association between variation in the human gut microbiome and colorectal cancer (CRC). However, despite compelling results from *in vivo* and *in vitro* models, few findings have translated between model organisms and human studies have been unconvincing in their ability to offer causal evidence. Reasons for these discrepancies are mainly due to the challenges in multi-omic technologies (e.g., sequencing and metagenomics) and sensitive experimental models. Additionally, epidemiological studies within this context have been of cross-sectional or case-control design; therefore, confounding of lifestyle and behavioural factors, reverse causation and bias are particularly important limitations to consider. The evidence that has been lost in translation impedes any opportunity for harnessing the gut microbiome – a uniquely malleable trait – for improving population health. Mendelian randomization (MR) is a method that uses human genetic variation (usually single nucleotide polymorphisms, SNPs) as instruments to proxy for a clinically relevant trait to improve

causal inference.

**Methods:** Here, we used MR to interrogate the causal impact of the gut microbiome on CRC combining summary-level data from genome-wide association studies (GWASs) of host genotype and gut microbiome variation from the Flemish Gut Flora Project (FGFP) and two German cohorts (n=3890) with the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO, n=120328). All analyses were conducted in MR-Base and pleiotropy was assessed using PhenoScanner and MR-TRYX.

**Results:** Of the 157 microbial traits (MTs), including presence, abundance and diversity metrics, assessed in the GWAS of the gut microbiome, there was evidence for a host genetic contribution to 13 MTs. Of these, we found evidence that presence of a genus within a certain order of bacteria increased the risk of CRC by 8% (95% CI: 2-15%; P=0.02). There was no strong evidence that the SNP used as an instrument was associated with other traits, reducing the likelihood of pleiotropy.

**Conclusions:** Our study confirmed previous observational evidence suggesting that genera within this certain bacterial order are more present in CRC cases than controls and provides further evidence that this observation may be due to a causal impact of these bacteria on CRC.

## **PgmNr 976: Mapping the genomic and epigenomic interaction landscape in triple negative and non-triple negative breast cancer.**

### **Authors:**

J. Wu; T. Mamidi; C. Hicks

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, Louisiana.

---

Breast cancer (BC) is the most diagnosed cancer among women in the US and the second leading cause of cancer-related death in women. Majority of the BC respond to endocrine and targeted therapy and generally have good prognosis. However, a significant proportion are triple negative breast cancer (TNBC), the most aggressive and lethal form of BC. Patients with TNBC have poor prognosis, a significantly increased risk of relapse and shorter survival rates than patients affected by non-TNBC. Currently, there are no targeted therapies for TNBC, cytotoxic chemotherapy remains the only effective therapeutic modality. The molecular mechanisms underlying the differences between TNBC and non-TNBC are poorly understood. Genes may play a strong role but epidemiological studies suggest that BC risk is largely determined by a complex interplay between genetic (both germline and somatic) and epigenetic alterations, and a broad range of environmental factors. Therefore, discovery of molecular markers and targets for the development of novel therapeutics must consider joint analysis of genomics and epigenomic alterations. Recent advances in high-throughput genotyping and sequencing technologies have provided opportunities for comprehensive analysis of germline, somatic and epigenetic alterations. Traditionally, analysis of genomic and epigenomic alteration has largely been carried out as separate research endeavors. Here we advocate an integrated approach that leverages germline mutation information from GWAS and somatic mutation information from The Cancer Genome Atlas (TCGA) and integrates it with information on DNA methylation using gene expression as the intermediate phenotype from TCGA to map the genomic and epigenomic interaction landscape in TNBC and non-TNBC. Preliminary results from this integrative analysis reveal the genes containing germline and somatic mutations also epigenetically altered. We discovered genomically and epigenomic altered molecular signatures associated with each type and distinguishing the two types of BC. Additionally, we discovered multiple gene regulatory networks and multiple signaling pathways enriched for germline, somatic and epigenetic alterations in each type of BC. Our investigation demonstrates that integrative analysis is a powerful approach to mapping oncogenic interactions and cooperation among germline, somatic and epigenomic changes shaping the TNBC and non-TNBC phenotypes.

## PgmNr 977: A novel next-generation sequencing platform for hereditary cancer risk evaluation.

### Authors:

J. Yang <sup>1</sup>; L. Wang <sup>1</sup>; L. Li <sup>1</sup>; J. Xiao <sup>1</sup>; Y. Jiang <sup>1</sup>; H. Dai <sup>1,2</sup>; L. Meng <sup>1,2</sup>; F. Xia <sup>1,2</sup>; E. Schmitt <sup>1</sup>; S. Peacock <sup>1</sup>; G. Wang <sup>1</sup>; C. Eng <sup>1,2</sup>; J. Zhang <sup>1,2</sup>; Z. Chen <sup>1,2</sup>; Y. Wang <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor Genetics, Houston, TX.; 2) Baylor College of Medicine, Houston, TX

---

**Background:** Hereditary cancer is a group of inherited disorders featured by increased cancer risks. Multiple cancer types are included such as, hereditary breast and ovarian cancer syndrome, Cowden syndrome, Lynch syndrome and others. Understanding one's genetic risk of cancer is critical for prevention and treatment, especially to those with family history of cancers. An NGS panel test has been developed in house to help evaluate the individual risk of hereditary cancer. **Methods:** 27 genes that are well-known to play important roles in cancer (*APC, ATM, BARD1, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDK4, CDKN2A, CHEK2, EPCAM, MLH1, MRE11A, MSH2, MSH6, MUTYH, NBN, PALB2, PMS2, PTEN, RAD50, RAD51C, RAD51D, SMAD4, STK11 and TP53*) are evaluated by the NGS panel test. NGS sequencing results were analyzed using NextGene software to detect sequence variants, while NGS read depth analysis is used to help with the assessment of possible copy number changes. MLPA is incorporated in this test for the examination of duplication/deletion of the PMS2 gene to exclude the influence of the pseudogenes which are highly homologous to PMS2. Sequencing variants are curated following the ACMG guidelines by experienced genomic scientists. Pathogenic, likely pathogenic and variants of uncertain significance are reported. **Results:** 164 samples were involved in this study. Majority of the patients in this cohort had either personal or family history of malignant cancers. Analysis found 122 reportable variants in 77 samples. Among these variants, 15 of them from the ATM, BRCA1, BRCA2, CHEK2, MUTYH and PALB2 genes are pathogenic, including 9 stop-gain, frameshift and splicing site variants. 3 samples had likely pathogenic variants in the CHEK2 gene. Among 59 samples, there were 104 variants of unknown significance in 25 genes. No variants have been detected in PTEN and SMAD4 so far. No copy number changes have been detected with this test. **Conclusion:** Our recent NGS panel for hereditary cancer study is a useful tool to evaluate one's risk of cancer. useful tool for assessing cancer predisposition. Due to present limitations on variant classification, VUS are identified at a rate of ~ 6 times that for pathogenic or likely pathogenic variants. Further functional studies and large-scale correlation studies are needed to better classify the VUSs.

## PgmNr 978: The first study evaluating the distribution of gBRCA1/2 mutations within the ovarian cancer cluster region in ovarian cancer patients using Japan CHARLOTTE study data (Characterizing the cross-sectional approach to ovarian cancer: Genetic testing of BRCA).

### Authors:

K. Yoshihara <sup>1</sup>; T. Enomoto <sup>1</sup>; D. Aoki <sup>2</sup>; Y. Watanabe <sup>3</sup>; J. Kigawa <sup>4</sup>; N. Takeshima <sup>5</sup>; H. Inomata <sup>6</sup>; K. Hattori <sup>6</sup>; H. Tsuda <sup>7</sup>; T. Sugiyama <sup>8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Obstetrics and Gynecology, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Niigata, Japan; 2) Department of Obstetrics and Gynecology, Keio University School of Medicine; 3) Department of Obstetrics and Gynecology, Tohoku Medical and Pharmaceutical University; 4) Department of Obstetrics and Gynecology, Matsue City Hospital; 5) Department of Gynecology, Cancer Institute Hospital; 6) Department of Medical, AstraZeneca K.K; 7) Department of Basic Pathology, National defense Medical College; 8) Woman's Cancer Center, Takagi Hospital

**Background and objectives:** The CHARLOTTE study showed the prevalence of germline *BRCA1/2* mutations (g*BRCA1/2m*) in Japanese ovarian cancer patients and a slightly higher prevalence of g*BRCAm* in patients with high-grade serous carcinoma and family history of ovarian cancer. In previous report, ovarian cancer cluster region (OCCR) has been attributed to higher ovarian cancer risk and lower breast cancer risk. However, the prevalence of g*BRCA1/2m* within the OCCR has never been evaluated in Japanese ovarian cancer patients. Using the CHARLOTTE study data, we described locations of g*BRCA1/2m* within the OCCR and characteristics of the relevant populations.

**Methods:** The CHARLOTTE study is a collaborative cross-sectional study(NCT03229122)involving numerous study sites throughout Japanwith histologically confirmed FIGO stage I–IV ovarian cancer. g*BRCA*testing was performed using blood samples. Only deleterious or suspected deleterious mutations were defined as g*BRCAm*. Based on previously reported OCCR (g*BRCA1* 1380~4062 bp, g*BRCA2* 3249~7471 bp), distribution of the location of g*BRCA1/2m*were identified and characteristics of these patients were evaluated.

**Results:** A total of 666 patients were enrolled and 634 were evaluated. The overall prevalence of g*BRCA1m* and g*BRCA2m* were 9.9% and 4.7%, respectively. Ovarian cancer patients with g*BRCA1* founder mutation (L63X) were excluded from the subsequent analysis because this founder mutation is known to relate with increases in both breast and ovarian cancer risk. In patients with g*BRCA1m*, 59.6% of ovarian cancer patients were distributed within the OCCR. On the other hand, 31.9% of ovarian cancer patients with g*BRCA1m* were distributed outside of the OCCR. The ovarian cancer patients with g*BRCA2m* were 53.3% within the OCCR, whereas 46.7% of ovarian cancer patients with g*BRCA2m* distributed outside of the OCCR. As for g*BRCA1m*, 39.3% and 32.1% of patients within the OCCR had family history of ovarian cancer and breast cancer, respectively, whereas 13.3% and 60% of patients outside of the OCCR has family history of ovarian cancer and breast cancer, respectively.

**Conclusions:** In Japanese ovarian cancer patients, g*BRCA1/2m* of more than half of study population were distributed within the OCCR and have higher family history of ovarian cancer and lower family

history of breast cancer compared to those outside of the OCCR in g*BRCA1*m. These results suggest future possibility of early detection of ovarian cancer in g*BRC*Am carriers.



## PgmNr 979: Genetic variations in long non-coding RNAs are associated with oral squamous cell carcinoma survival.

**Authors:**

H. Yuan; B. Zhu; Y. Yu; L. Miao

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Oral and Maxillofacial Surgery, Affiliated Hospital of Stomatology, Nanjing Medical University, Nanjing, China

---

Oral squamous cell carcinoma (OSCC), the subtype of head and neck cancers, is notorious for its high incidence and death rate. The role of long non-coding RNAs (lncRNAs) is discovered to be significant for the cancerization and cancer progression. lncRNA urothelial carcinoma-associated 1 (UCA1) and nuclear paraspeckle assembly transcript 1 (NEAT1) were identified as a oncogene in multiple carcinomas, but there has been no research on the association of those two lncRNAs with OSCC prognosis. Therefore, we hypothesized that genetic variants of those two lncRNAs might predict the survival of OSCC. We selected six functional SNPs on those two lncRNAs to explore the relationship between the variations in these lncRNAs and the prognosis of OSCC. By using the Kaplan-Meier analyses and Cox proportional hazards regression models, we found that a predictive role of NEAT1 rs3741384 and UCA1 rs7255437 in the prognosis of OSCC. In addition, a remarkable increased risk of death was observed in the patients with an increasing number of unfavorable genotypes (NUG). As a result, our findings indicated that genetic variants in UCA1 and NEAT1 may influence the survival of OSCC patients.

## PgmNr 980: Browse, visualize and analyze clinical and whole-genome data of 3006 pediatric cancer survivors on the SJLIFE portal.

### Authors:

Zhou <sup>1</sup>; S. Lei <sup>1</sup>; K. Shelton <sup>2</sup>; N. Bhakta <sup>3</sup>; J. Wang <sup>1</sup>; E. Sioson <sup>1</sup>; J. Patel <sup>1</sup>; Z. Wang <sup>2</sup>; C. Wilson <sup>2</sup>; Y. Yasui <sup>2</sup>; A. Gout <sup>1</sup>; M. Hudson <sup>4</sup>; L. Robison <sup>2</sup>; J. Zhang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computational Biology, St. Jude Children's research hospital, Memphis, TN.; 2) Epidemiology & Cancer Control, St. Jude Children's research hospital, Memphis, TN.; 3) Global Pediatric Medicine, St. Jude Children's research hospital, Memphis, TN.; 4) Oncology, St. Jude Children's research hospital, Memphis, TN.

---

To accelerate the rate of discovery in survivorship research we have developed the SJLIFE Portal (<https://survivorship.stjude.cloud>), a data-sharing platform on the St. Jude Cloud. The Portal hosts 3006 clinically-assessed long-term survivors of childhood cancer with whole-genome sequencing (WGS) and detailed demographic, diagnostic, treatment and outcome information. The Portal features the Clinical Dictionary Browser and GenomePaint, a genetic variant browser, for visualizing and analyzing clinical and genetic data integratively. The Clinical Dictionary Browser enables interactive exploration of 1) Cancer-related variables such as diagnosis and treatment; 2) Demographic variables; and 3) Outcomes, including severity-graded chronic health conditions using a modified version of the Common Terminology Classification for Adverse Events, and subsequent cancers. Dictionary terms from these categories are arranged hierarchically from general to specific, allowing a user to traverse the dictionary tree and view patient distributions by user selected terms as customizable barcharts, or perform cross-tabulation between two terms. GenomePaint displays SNV, indel, and copy number variants computed from WGS, allowing to display and filter variants based on a variety of functional annotations and population allele frequencies. Importantly, GenomePaint provides on-the-fly locus-specific association testing on any coding or non-coding locus to identify variants associated with user's choice of clinical terms. For example, by choosing terms from the categories of cancer-related variables in the Clinical Dictionary Browser, a user can compare leukemia and non-leukemia survivors of a specific ethnic group, or survivors diagnosed within a specific age range (e.g.  $\leq 4$  years), against a control population such as gnomAD. The resulting variants can be filtered using  $r^2$  values for nearby variants that are in linkage disequilibrium. Non-coding variants of interest can be further studied for their effect on altering transcription factor binding motifs, thereby providing an in-silico assessment of their potential impact on gene regulation. Participants enrolled in the SJLIFE Study will be followed longitudinally across their lifespan. Therefore, the clinical and genetic information of the SJLIFE Portal will be updated and enriched continuously, providing an increasingly rich resource to facilitate cancer and survivorship research over time.

## PgmNr 981: Characterization of *BRCA1* and *BRCA2* genetic variants in Guatemala population with breast cancer.

### Authors:

C. Carranza <sup>1</sup>; V. Zamora <sup>1</sup>; S. Torselli <sup>2</sup>; M. Chivalan <sup>2</sup>; E. Gharzouzi <sup>3</sup>; M. Guerra <sup>1</sup>; C. Osorio <sup>1</sup>; L. Linares <sup>3</sup>; M. Garces <sup>3</sup>; M. Herrera <sup>1</sup>; N. Escobar <sup>1</sup>; N. Marroquin <sup>1</sup>; V. Alvarado <sup>1</sup>; L. Rosales <sup>1</sup>; C. Rangel <sup>1</sup>; J. Rozas-Botran <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Invegem, Guatemala; 2) Roosevelt National Hospital, Guatemala; 3) Integra Cancer Institute, Guatemala

The worldwide incidence of breast cancer is increasing. It has become the most common cancer in women in almost all the countries. In Guatemala, it is the second most frequent cancer, only exceeded by cervical cancer. Around 10% of these tumors are hereditary. The hereditary breast-ovarian cancer syndrome is produced in the majority of the cases, by the presence of *BRCA1* and *BRCA2* pathogenic variants.

In Guatemala, there are not previously reports about the variants type of *BRCA1* and *BRCA2* gene. It is important to characterize our population, because our mayan influence.

The main objective of this study was to detect the most frequent *BRCA1* and *BRCA2* variants in Guatemalan population.

To do this, we evaluated 118 patients with hereditary and familial breast-ovarian cancer. We selected patients with breast or ovarian cancer with familial history or/and with age less than 40 years.

We obtained a peripheral blood sample of each patient. And performed DNA extraction. Both genes were evaluated by next generation sequencing (Miniseq, Illumina), and were confirmed by Sanger sequencing.

The mean age in all the positive patients were 41 years old. Twenty-two cases of one hundred eighteen patients, 22/118 (19%) were positive for *BRCA1* or *BRCA2* mutations. Eight cases have *BRCA1* pathogenic variants and fourteen cases have *BRCA2* pathogenic variants.

The most frequent variants in *BRCA1* gene were: c.798\_799delTT (p. Ser267Lysfs) (five cases) and a splice site mutation: c.212+1G>A (four cases). Followed by c.2748delT (p. Asn916Lysfs) found in two cases.

The most frequent variant in *BRCA2* gene was c.9235delG (p. Val3079Phefs) (two cases). In the following table, we showed the pathogenic variants found in our population.

*BRCA1* and *BRCA2* pathogenic variants in guatemalan population

Gene	Nucleotide	protein	N
<i>BRCA1</i>	c.212+1G>A	----	4
	c.798_799delTT	p.Ser267Lysfs	5
	c.2748delT	p.Asn916Lysfs	2
	c.4547G>A	p.Arg1516Lys	1
	c.6436G>T	p.Glu215Ter	1
	c.2596C>T	p.Arg866Cys	1

BRCA2	c.9235delG	p.Val3079Phefs	2
	c.8186delA	p.Lys2729Argfs	1
	c.4936-4939delGAAA	p.Glu1646Glnfs	1
	c.1813delA/c.9097delA	p.Ille605Tyrfs/p.Thr3033Leufs	1
	c.6560C>T	p.Pro2187Leu	1
	c.223G>C	p.Ala75Pro	1
	c.8174_8185delGGTATGCTGTTAinsTT	p.Trp2725Phefs	1

This study showed the most common *BRCA1* and *BRCA2* variants in Guatemalan population, being the first report published in our population. And we demonstrated that our population most frequent variants are different than other countries reports.

## PgmNr 982: BRCA Exchange: Global data sharing and integration for variant interpretation.

### Authors:

M. Cline <sup>1</sup>; A. Coffin <sup>1</sup>; Z. Fischmann <sup>1</sup>; B. Paten <sup>1</sup>; G. Ratsch <sup>2</sup>; M. Zimmerman <sup>2</sup>; F. Alquodmami <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UC Santa Cruz Genomics Institute, Santa Cruz, California.; 2) ETH Zurich, Rämistrasse 101. 8092 Zurich. Switzerland

---

As genetic testing grows in prevalence, the rate at which new variants are discovered greatly outpaces the rate at which they can be interpreted. To some degree, this is a data sharing problem. Most genetic variants require case-level data for robust interpretation. However, most unclassified variants are rare, such that no single institution may have a sufficient set for variant interpretation, and case-level data is difficult to share for numerous legal, regulatory and ethical reasons. To address this challenge, the Global Alliance for Genomics and Health (GA4GH) launched the BRCA Challenge. The goal of BRCA Challenge is to accelerate and improve clinical interpretation of germline *BRCA1* and *BRCA2* variants in hereditary breast and ovarian cancer, through global data integration and analysis, and to disseminate variant knowledge. Further, the goal of the BRCA Challenge is to analyze what types of approaches prove to be most effective in advancing variant interpretation, given the resources of a focused and active consortium, to serve as a model for other genes, cancers and syndromes.

BRCA Exchange (<https://brcaexchange.org/>), the first work product of the BRCA Challenge, integrates variant-level data from sources including ClinVar, the Leiden Open Variation Database (LOVD), population databases, selected functional assays, and others, in a single point of access. BRCA Exchange is the largest public source on BRCA variation data, with data on close to 25,000 distinct variants. These data are publicly available for analysis on the web portal, for download, and for programmatic access via a GA4GH-compliant API. The BRCA Challenge works closely with the Evidence-based Network for the Interpretation of Germline Mutant Alleles (ENIGMA) consortium and other variant curators to identify and prioritize data resources for variant interpretation.

While BRCA Exchange has integrated data from the major public data sources, most variant data is held in private silos. To integrate these data, BRCA Exchange is developing federated analysis procedures, to 'bring the code to the data' via methods such as Docker, analyze the protected, patient-level data in its secure home repository, and generate aggregated, variant-level statistics that contain no protected health information and can be shared more broadly. BRCA Exchange is working closely with variant curators, regulatory experts and data scientists to ensure that this sharing is safe and informative.

## **PgmNr 983: Deleterious germline mutations in the *BRCA1* gene are associated with increased risk for cancers of the female reproductive system other than breast and ovarian as well as other cancers.**

### **Authors:**

C. Middlebrooks <sup>1</sup>; K. Pantin <sup>2</sup>; M. Stacey <sup>3</sup>; C. Snyder <sup>3</sup>; T. Shaw <sup>3</sup>; M.J. Casey <sup>3,4</sup>; J.E. Bailey-Wilson <sup>1</sup>; H.T. Lynch <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) National Human Genome Research Institute, Baltimore, Maryland.; 2) University Department of Psychology, of Illinois at Urbana-Champaign, Champaign, IL; 3) Hereditary Cancer Center, Creighton University, Omaha, NE; 4) Department of Obstetrics and Gynecology, Creighton University, Omaha, NE

---

**Introduction:** Mutations within the *BRCA1* gene have been linked to up to an 80% lifetime risk of breast cancer as well as increased risk for ovarian, pancreatic and melanoma cancers. In this study we examined families with known germline mutations in *BRCA1* after long-term follow-up to determine whether carriers experience higher rates of other cancers that have not yet been associated with germline mutations in the *BRCA1* gene.

**Methods:** We studied 127 Hereditary Breast and Ovarian Cancer (HBOC) syndrome families (N = 23,078 individuals who have been followed at Creighton University) in which a causal mutation in the *BRCA1* gene was identified. We performed survival analysis and a mixed effects cox regression with age at follow-up or cancer event as our time variable and presence or absence of *BRCA1*-related or other cancers (separate analyses) as our indicator variable.

**Results:** The survival curves showed a significant age effect with carriers having a younger age at cancer onset for *BRCA1*-related (as expected) as well as other cancers than that of non-carriers. The cox regression models were also highly significant ( $P = 1.77E-37$  and  $P = 1.04E-07$  for the *BRCA1*-related and other cancers, respectively). Of the cancers with enough samples to do stratified analyses, cervix, uterine, skin, lymphoma and colon cancers occurred at higher rates and at earlier ages in mutation carriers.

**Conclusions:** These analyses support the hypothesis that the *BRCA1* mutations carriers of HBOC syndrome have increased risk for early onset of several additional cancer types, especially cancers that arise in estrogen-influenced tissues.

## PgmNr 984: Uterine fibroids in Senegal: Polymorphism of *MED12* gene and correlation with epidemiological factors.

### Authors:

B. Keneme <sup>1</sup>; D. Ciss <sup>2</sup>; S. Ka <sup>3</sup>; F. Mbaye <sup>1</sup>; M. Sembene <sup>1</sup>; A. Dem <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Animal Biology, Cheikh Anta Diop University, Dakar, Senegal; 2) Service de Maternité et Gynécologie Obstétrique/Hôpital Général Grand Yoff; 3) Institut Joliot Curie, Faculté de Médecine de Pharmacie et d'Odontologie, Université Cheikh Anta Diop, Dakar, Sénégal

---

**Background:** Mediator complex subunit 12 (*MED12*) is a part of the mediator complex, which is believed to regulate transcription. Our recent study showed that *MED12* is mutated at high frequency and with different mutation frequencies in Senegalese women with uterine fibroids. However, the status of mutations has not been correlated to the epidemiological factors that are implicated in uterine fibroids.

**Methods:** This study aimed to analyze status of *MED12* mutations in Senegalese population. *MED12* was sequenced in tumoral tissues and blood samples of 54 Senegalese women with uterine fibroids. Clinical and pathological data were obtained from the patient's records and other parameters were recorded. Mutation Surveyor software version 5.0.1, DnaSP version 5.10, MEGA version 7.0.26 and Arlequin version 3.5.1.3 were used to determine the level of mutations and genetics parameters. To estimate the genetic variation according to the epidemiological parameters, the index of genetic differentiation (*F<sub>st</sub>*) and the genetic structure like analysis of molecular variance (AMOVA) were determined with Arlequin software version 3.5.1.3. The significance level (P-value) was 0.05.

**Results:** Our results showed that *MED12* is mutated at 88.89% (48/54) only in tumor tissues. The variants frequencies were not similar to those found in the Finnish populations. The Chi2 test indicates a statistically significant difference for the variants c.130G>C, c.130G>A, c.131G>A and for the Intronic Variant Site (*p* < 0.05). This variable expression of the *MED12* gene is further confirmed by the amino acid frequency between blood and tumor tissue with a statistically significant difference for alanine, glutamic acid, lysine, methionine, threonine, valine, tryptophan and tyrosine (*p* < 0.05). In addition, the selection test indicates that codon 44 is under positive selection (*p* = 0.0243) in cases of uterine fibroids. Genetic diversity according to risk factors such as parity and diet was observed in uterine fibroids in Senegal (*p* < 0.05).

**Conclusion:** Significant genetic diversity has been noted in uterine fibroids in Senegal. The codon 44 being under positive selection could be considered as a biomarker in uterine fibroids. Depending on the epidemiological parameters studied, parity and diet seem to be the risk factors most implicated in uterine fibroids in Senegal.

## PgmNr 985: Genome wide copy number variant analysis to identify regions associated with epithelial ovarian risk cancer risk.

### Authors:

A.A. DeVries<sup>1</sup>; J. Dennis<sup>2</sup>; S.G. Coetzee<sup>1</sup>; J. Tyrer<sup>2</sup>; A.L. Reyes<sup>1</sup>; J. Plummer<sup>1</sup>; R.I. Corona de la Fuente<sup>1,2</sup>; K. Lawrenson<sup>3</sup>; J. Seo<sup>4</sup>; M.L. Freedman<sup>4</sup>; D.J. Hazelett<sup>1</sup>; B.P. Berman<sup>1</sup>; P. Pharoah<sup>2</sup>; S.A. Gayther<sup>1</sup>; M.R. Jones<sup>1</sup>; OCAC

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Bioinformatics and Functional Genomics, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California, USA; 2) CR-UK Department of Oncology, University of Cambridge, Strangeways Research Laboratory Cambridge, UK; 3) Women's Cancer Program at the Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Suite 290W, Los Angeles, CA, USA; 4) Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

---

The genetic risk variants uncovered by GWAS and other genetic studies account for only 30% of the heritability for epithelial ovarian cancer (EOC). To date, no studies have identified germline copy number variants (CNVs; regions of deletion and duplication) that are associated with EOC. Germline CNVs have been associated with risk for several traits (breast and pancreatic cancer) and a single CNV has been reported as associated with survival in ovarian cancer. As germline CNVs are an inherited form of genetic variation, we hypothesized that germline CNVs contribute to EOC risk. Using intensity data from >25,000 EOC cases and >40,000 controls from OCAC genotyped on the OncoArray, we identified CNVs using PennCNV. Subsequent association testing identified CNVs significantly associated with risk for high grade serous ovarian cancer both within previously studied ovarian cancer associated genes (dup:*BRCA1*  $P=9.52 \times 10^{-10}$ , del:*RAD51C*  $P<0.0001$ ) and within several novel loci (dup:*SFI1*  $P=2.0 \times 10^{-5}$ , *GALNT18*  $P=0.0008$ , *CDYL2*  $P=0.0015$ ). We also identified CNVs in high linkage disequilibrium ( $LD \geq 0.8$  in CEU) with tag SNPs on the OncoArray. Four significant loci were identified ( $P<5.0 \times 10^{-8}$ ) in an analysis of all non-mucinous EOC, one of which was also significant in endometrioid and high grade serous histotypes (9p22.2). We then performed annotation of each associated CNV with a large set of epigenomic data generated in >40 ovarian cancer related cell types that includes H3K27Ac, H3K4me1, H3K4me3, RNA-Seq, CTCF, and WGBS, combined to create chromatin state calls. We identified cell type specific enhancers that are disrupted by non-coding CNVs and are validating these with additional genotyping methods and functional experiments. Additionally, we found that regions of H3K27Ac were enriched in a cell-type specific manner for the candidate significant CNVs across the genome, indicating these CNVs may indirectly alter expression of genes via dosage of non-coding regulatory biofeatures. Additionally, deletions and duplications within coding regions of known EOC risk genes such as *BRCA1* and *RAD51C* implicate additional potential pathogenic risk variants not currently included in clinical testing for EOC. Our combination of CNV discovery and annotation analysis has provided functional information about risk variants and identified novel loci not previously found in other ovarian cancer SNP-only GWAS, expanding our knowledge of the heritability of ovarian cancer.



## PgmNr 986: Germline deleterious rare variants and their association with lung cancer.

### Authors:

Y. Liu <sup>1</sup>; J. McKay <sup>2</sup>; S. Tsavachidis <sup>1</sup>; X. Xiao <sup>1</sup>; Y. Li <sup>1</sup>; C. Cheng <sup>1</sup>; J. Field <sup>3</sup>; M. Davies <sup>3</sup>; P. Brennan <sup>2</sup>; G. Liu <sup>4</sup>; R. Hung <sup>5</sup>; D. Christiani <sup>6</sup>; C. Amos <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor College Medicine, Houston, Texas.; 2) International Agency for Research on Cancer, Lyon, France; 3) The University of Liverpool, Liverpool, UK; 4) Princess Margaret Cancer Center, Toronto, ON, M5G 2M9, Canada; 5) Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, M5G 1X5 Canada; 6) Harvard University School of Public Health, Boston, MA 02115, USA

**Background:** Despite previous family-based linkage studies and intensive population-based genome-wide association studies analyses, a large proportion of the heritability of lung cancer remains unexplained. Recent studies suggest that rare variants exhibit stronger effect sizes than common variants, and could account for missing heritability unexplained by common variants.

**Methods:** Germline DNA from 1045 lung cancer cases and 885 controls from the Transdisciplinary Research in Cancer of the Lung and International Lung Cancer Consortium studies were sequenced, utilizing the Agilent SureSelect<sup>XT</sup> Custom ELID and Whole Exome v5 captures. To identify rare (minor allele frequencies <1%) variants, including single nucleotide variants (SNVs) and insertion/deletions (Indels), we focused on frameshift, nonsense, stop-gain/loss, and splice donor/acceptor variants. To further prioritized candidates that are highly enriched in cases (risk-conferring) or controls (protective), we conducted allelic association analyses of single variant and gene-based collapsing tests of multiple variants.

**Results:** We identified 78 candidate variants of which 28 were SNVs and 50 were Indels, including 18 reported to be pathogenic in ClinVar database (i.e., *ATM* missense p.V2716A and *CHEK2* p.S471F frameshift). The top five risk-conferring candidates (allelic association  $P < 0.005$ ), including the p.P173R frameshift in *CRIPAK*, in-frame deletion p.A324del in *NOVA2*, in-frame deletion p.R85del in *RBM10*, p.G629R frameshift in *SYCP1*, stop gain p.Q7971\* in *NEB*; whereas the top five protective candidates (allelic association  $P < 0.005$ ), including the transcription factor binding site (rs1252601936) in *KYAT1*, frameshift p.N295K in *MED8*, p.L94H frameshift in *FAM111A*, p.E9G frameshift in *TXNDC15*, and transcription end site (8:69699853) in *C8orf34*. We also identified candidates mapped to known lung cancer susceptibility genes, i.e., *HLA*, *GPC*, *TP63*, *CDKN2A/B*, *CHRNA6*, and *LAMC1*. Among the candidate genes with multiple rare deleterious variants, the top five genes with strong association ( $P < 0.001$  from the Combined Multivariate and Collapsing test and the Kernel-Based Adaptive Cluster test) are *CYP24A1*, *KYAT1*, *C8orf34*, *CDKN2B*, and *SYCP1*.

**Conclusion:** Our whole exome sequencing analyses led to identification of 78 rare and deleterious inherited variants associated with lung cancer susceptibility. Further in-depth functional follow-up studies are needed to evaluate the pathogenicity of the candidates.

## PgmNr 987: Highly aggregated lung cancer families show significant linkage to chromosome 12q23.3 for cancer risk.

### Authors:

A. Musolf<sup>1</sup>; B. Moiz<sup>1</sup>; H. Sun<sup>1,2</sup>; C. Pikielny<sup>3</sup>; D. Mandal<sup>4</sup>; M. de Andrade<sup>5</sup>; C. Gaba<sup>6</sup>; P. Yang<sup>5</sup>; Y. Li<sup>7</sup>; M. You<sup>8</sup>; R. Govindan<sup>9</sup>; R. Wilson<sup>10</sup>; E. Kupert<sup>8</sup>; M. Anderson<sup>8</sup>; A. Schwartz<sup>11</sup>; S. Pinney<sup>12</sup>; C. Amos<sup>7</sup>; J. Bailey-Wilson<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 2) Laboratory of Medical Genetics, Harbin Medical University, Harbin, China; 3) Geisel School of Medicine, Dartmouth College, Lebanon, NH; 4) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 5) Mayo Clinic, Rochester, MN; 6) Department of Medicine, University of Toledo Dana Cancer Center, Toledo, OH; 7) Baylor College of Medicine, Houston, TX; 8) Medical College of Wisconsin, Milwaukee, WI; 9) Division of Oncology, Washington University School of Medicine, St. Louis, MO; 10) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; 11) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 12) Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH

Lung cancer (LC) kills more people than any other cancer in the United States. It is known that lung cancer is a complex trait caused by both environmental and genetic factors, yet the genetic etiology of lung cancer remains poorly understood. In this study, we have performed whole exome sequencing (WES) on 262 individuals from 28 extended families that have a strong history of LC and are highly aggregated for the phenotype. The WES was recalled with PICARD/GATK and standard quality controls were performed, leaving approximately 500,000 SNVs and indels for analysis.

Parametric genetic linkage analysis was performed on these families using two distinct models – the lung cancer only (LCO) model, where only lung cancer patients were coded as affected, and the all inherited cancers (AIC) model, where other inherited cancers were coded as affected as well (pedigrees averaged 1-2 additional non-lung cancer patients). All unaffected individuals were given an “unknown” phenotype. Both models assumed an autosomal dominant mode of inheritance with a disease allele frequency of 1% and a penetrance of 80% for carriers and 1% for non-carriers. The AIC model yielded a genome-wide significant result (HLOD = 3.3) at rs61943670 in the RNA polymerase III gene *POLR3B* at 12q23.3. *POLR3B* has been implicated somatically in lung cancer but this germline finding is novel. Interesting genome-wide suggestive haplotypes were also found within individual families, particularly near *SSPO* at 7p36.1 and a large haplotype spanning 4q21.3-28.3. The 4q haplotype contains potential causal rare variants *DSPP* at 4q22.1 and *PTPN13* at 4q21.3. *PTPN13* is a known tumor suppressor in lung cancer. Functional work on *POLR3B* and several of the best candidates for the individual family signals (such as *SSPO*, *DSPP*, and *PTPN13*) is planned for future work.

## PgmNr 988: Sex differences in genetic associations with Barrett's esophagus and esophageal adenocarcinoma.

### Authors:

J. Dong <sup>1</sup>; C. Maj <sup>2</sup>; A. Thrift <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine, Baylor College of Medicine, Houston, TX.; 2) University of Bonn, Germany

---

**Background and Aims:** Esophageal adenocarcinoma (EA) is characterized by a strong and yet unexplained male predominance (with a male-to-female ratio in the incidence of up to 9:1). Genome-wide association studies (GWAS) have identified more than 20 susceptibility loci for EA and its premalignant lesion, Barrett's esophagus (BE). However, the sex differences in genetic associations with BE/EA remain largely unknown.

**Methods:** Given the strong genetic overlap, BE and EA cases were combined into a single case group for analysis and compared cases with population-based controls. We performed a sex-specific GWAS of BE/EA in two separately study datasets and then meta-analyzed using fixed-effects inverse variance-weighting approach. We further conducted a genome-wide single nucleotide polymorphism (SNP)-by-sex interaction analysis of BE/EA. To correct for multiple testing, we used  $P < 5 \times 10^{-8}$  for statistical significance.

**Results:** There were 6,758 male cases, 3,488 male controls, 1,670 female cases and 2,226 female controls. Genome-wide meta-analysis of SNP-by-sex interaction identified three novel variants at chromosome 11q13.2 not previously associated with the risk of BE/EA: rs1626067 (*CARNS1*,  $P_{\text{meta}} = 1.16 \times 10^{-8}$ ), rs12789119 (*PITPNM1*,  $P_{\text{meta}} = 2.35 \times 10^{-8}$ ) and rs2276120 (*PITPNM1*,  $P_{\text{meta}} = 2.91 \times 10^{-8}$ ). These three SNPs exhibited stronger associations in males than in females. Although meta-analysis of sex-specific GWAS did not identify any association reached genome-wide significance ( $P < 5 \times 10^{-8}$ ), top-ranked SNPs at  $P_{\text{meta}} < 1.0 \times 10^{-6}$  appeared to be sex-specific.

**Conclusions:** We identified novel variants association with BE/EA. These identified differences could improve our understanding of the genetic architecture of the disease. Further studies are needed to elucidate the biological processes through which this risk is conferred.

## PgmNr 989: GWAS and fine mapping analysis identifies 16 new epithelial ovarian cancer risk loci.

### Authors:

E.O. Dareng<sup>1</sup>; J.P. Tryer<sup>2</sup>; S.A. Gayther<sup>3</sup>; P.D.P. Pharoah<sup>1,2</sup>; Ovarian Cancer Association Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom; 2) The Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom; 3) The Centre for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Centre, Los Angeles, California, United States of America

---

Genome-wide association studies have identified approximately 40 epithelial ovarian cancer (EOC) susceptibility risk loci in women of European ancestry. A large fraction of the genetic contribution to EOC risk remains unexplained. In this analysis, we imputed genotype data from 26,151 EOC cases and 64,783 controls to the Haplotype Reference Consortium panel and assessed the association of 31,995,530 genetic variants with risk of six EOC histotypes: non-mucinous (all cases of invasive EOC excluding mucinous), high-grade serous (HGSOC), low-grade serous (LGSOC), endometrioid (ENOC), clear cell (CCOC) and mucinous (MOC). We identified five new susceptibility loci for EOC risk at  $P < 5 \times 10^{-8}$ : two for HGSOC - 4q13.3 (rs4149419,  $P = 2.7 \times 10^{-8}$ ), 9p22.1 (rs7851336,  $P = 2.9 \times 10^{-10}$ ); two for non-mucinous - 5q11.2 (rs336126,  $P = 6.4 \times 10^{-11}$ ), 21q22.12 (rs2070368,  $P = 1.1 \times 10^{-8}$ ); and one for MOC - 2q14.2 (rs72827480,  $P = 2.7 \times 10^{-8}$ ). We then performed fine mapping analysis in each region that had a signal at genome wide threshold, adjusting for the most significant SNP in the region. These identified an additional 11 independent signals at a threshold of  $P < 10^{-5}$ , four of which were genome wide significant. Five of these signals were associated with HGSOC - 2q13 (rs895412,  $P = 1.3 \times 10^{-7}$ ), 5p15.33 (rs10069690,  $P = 2.1 \times 10^{-6}$ ), 8q24.21 (rs7833298,  $P = 1.2 \times 10^{-9}$ ), 9p22.1 (rs10810671,  $P = 6.5 \times 10^{-10}$ ) and 19p13.11 (rs12982058,  $P = 4.1 \times 10^{-6}$ ); two with MOC - 2q31.1 (rs2594950,  $P = 6.5 \times 10^{-6}$ ), 8q24.21 (rs6470494,  $P = 5.7 \times 10^{-6}$ ); two with non-mucinous - 5p15.33 (rs2853669,  $P = 4.9 \times 10^{-8}$ ), 8q24.21 (rs77235147,  $P = 7.2 \times 10^{-6}$ ); and two with LGSOC - 2q13 (rs1470053,  $P = 2.6 \times 10^{-6}$ ), 5p15.33 (rs2853677,  $P = 7.8 \times 10^{-10}$ ). We calculated the Bayesian False Discovery Probability (BFDP) for each of the loci identified in our analysis to assess noteworthiness of our findings, with low BFDP scores ( $< 10\%$ ) indicating true associations. We used a prior probability association of 0.0001 for regions with one signal, 0.001 for regions with more than one signal and a plausible per allele odds ratio of 1.2. Of the five loci identified in the primary GWAS analysis, the BFDP was  $< 1\%$  for three loci, and between 1% to 10% for the other two loci. For the 11 loci identified in conditional analysis, the BFDP was  $< 1\%$  for five loci, between 1% to 10% for one locus, and  $> 10\%$  for five loci. Our findings have the potential to improve EOC risk prediction and provide insight into the underlying genetic architecture of EOC.

## PgmNr 990: Leveraging tumor characteristics to predict germline variant pathogenicity in mismatch repair genes.

### Authors:

S. Li <sup>1</sup>; D. Qian <sup>1</sup>; B.A. Thompson <sup>2,3</sup>; S. Gutierrez <sup>1</sup>; T. Pesaran <sup>1</sup>; H. LaDuca <sup>1</sup>; H. Lu <sup>1</sup>; E.C. Chao <sup>1</sup>; M.H. Black <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ambry Genetics, Aliso Viejo, California.; 2) Department of Clinical Pathology, University of Melbourne, Parkville, Victoria, Australia; 3) Department of Pathology, Royal Melbourne Hospital, Parkville, Victoria, Australia

---

Pathogenic variants (PV) in mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*) increase cancer risks. Microsatellite instability (MSI) and protein immunohistochemical (IHC) changes due to MMR deficiency are often assessed to clinically evaluate cancer risk. It has been shown that these tumor characteristics may be useful in predicting germline MMR variant pathogenicity. In a large cohort of patients who underwent germline panel testing (GPT; n=4827) or paired germline/somatic testing (GST; n=1429), we derived likelihood ratios (LR=%PV carriers/%non-carriers) for those with abnormal vs. normal MSI/IHC status. For each variant, we computed the tumor characteristic LR (TCLR) as the product of carriers' MSI/IHC LRs. For GST samples, we additionally estimated *MLH1*-hypermethylation (hm)<sup>+/-</sup> LRs among samples with abnormal MSI/IHC (n=583). The MSI/IHC LRs were combined with *MLH1*-hm<sup>+/-</sup> LRs to compute overall and tumor-specific (colorectal, endometrial, ovarian, breast) TCLRs. We then compared the prediction performance of 3 approaches: TCLR alone, *in silico* priors with TCLR, and integrating TCLR evidence with *in silico* and other evidence into a Bayesian multifactorial model. Comparing carriers of germline PV vs. non-carriers, GPT samples had LRs of 2.98 for abnormal and 0.10 for normal MSI/IHC, while GST LRs were 1.57 and 0.04 for abnormal and normal MSI/IHC, respectively. The LR in favor of pathogenicity for samples with abnormal MSI/IHC was 4.47 for *MLH1*-hm<sup>-</sup> and 0.03 for *MLH1*-hm<sup>+</sup>. MSI/IHC LRs also varied by cancer type; LRs were highest in colorectal, whereas *MLH1*-hm<sup>-</sup> LRs were significantly higher in endometrial than colorectal cancers. TCLR alone accurately predicted 67.2% of 177 classified germline missense variants. Combining TCLR with *in silico* priors accurately classified >73.4% variants, depending on the *in silico* scores used for analysis. Integrating TCLR into a multifactorial model increased the proportion of accurately classified variants to 97.2%. Applying these models to the prediction of 221 variants of unknown significance (VUS), TCLR alone predicted 73% as (likely) pathogenic or (likely) benign, while the multifactorial model with TCLR reclassified 64%. However, TCLR alone yielded 5 false positives/negatives, while the multifactorial model resulted in 0 false predictions. These data highlight the importance of incorporating somatic sequencing in germline variant assessment and demonstrate the utility of TCLR evidence for VUS reclassification.

# PgmNr 991: Role of germline functional non-coding variants associated with ovarian cancer risk.

## Authors:

S.A.M. Ezquina <sup>1</sup>; E. Dicks <sup>1</sup>; R.I. Corona <sup>2</sup>; K. Lawrenson <sup>2</sup>; S. Gayther <sup>2</sup>; M.R. Jones <sup>2</sup>; M.L. Freedman <sup>3</sup>; R. Drapkin <sup>4</sup>; P. Pharoah <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Public Health and Primary Care, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom; 2) The Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, Los Angeles, CA, United States of America; 3) Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, United States of America; 4) University of Pennsylvania, Philadelphia, PA, United States of America

---

Known rare ovarian cancer susceptibility alleles target DNA coding sequences whereas most common risk alleles are in the non-coding genome; however, the functional role of non-coding variation is not well understood.

Statistical power limits the detection of individual rare susceptibility variants; power is increased by using burden testing, which has been used to identify several ovarian cancer risk genes. We hypothesize that germline non-coding variants associated with ovarian cancer risk might target important regulatory elements, such as binding sites of key transcription factors. Therefore, the aim of this study is to find rare, non-coding ovarian cancer susceptibility alleles on regulatory elements (specifically PAX8 binding sites in active chromatin regions) using burden testing.

High-grade serous ovarian cancer, the most common ovarian cancer histological subtype, is characterized by PAX8 overexpression. PAX8 is a transcription factor, also expressed in normal ovarian and fallopian tube surface epithelial cells (FTSECs), which are the putative precursor cells of high-grade serous ovarian cancer.

We used disease-specific epigenetic marks; active chromatin annotation (H3K27ac ChIP-seq) from 5 primary HGSOCS and PAX8 binding sites (PAX8 ChIP-seq) from 5 cell lines, two FTSEC (FT194, FT246) and three ovarian cancer cell lines (JHSO4, Kuramochi, OVSAHO). We selected PAX8 binding sites present in all 5 cell lines that overlap the consensus set (3 out of 5 H3K27ac ChIP-seq HGSOCS) of active chromatin regions.

We called variants in these regions in germline genomes of a partial dataset of 247 ovarian cancer patients and 1102 non-cancer controls sequenced for the main programme version 4 of the UK 100k Genomes Project. Variants with a frequency > 0.01 in the gnomAD dataset and in our dataset were excluded. Variants were manually filtered for quality using IGV. In a preliminary analysis, we found 56 cases and 173 controls with at least one rare variant at PAX8 binding sites.

We suggest that burden testing across non-contiguous regions of the non-coding genome is a promising approach to the identification for uncommon and rare disease susceptibility alleles.



# PgmNr 992: A framework for transcriptome-wide association studies in breast cancer in diverse study populations.

## Authors:

A. Bhattacharya<sup>1</sup>; M. Garcia-Closas<sup>2,3</sup>; A. Olshan<sup>4,5</sup>; C.M. Perou<sup>5,6,7</sup>; M. Troester<sup>4,7</sup>; M.I. Love<sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NORTH CAROLINA; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; 3) Division of Genetics and Epidemiology, Institute of Cancer Research, London, United Kingdom; 4) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 5) Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 6) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 7) Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina,

The relationship between germline genetic variation and breast cancer survival is largely unknown, especially in understudied minority populations. Genome-wide association studies (GWAS) have interrogated survival associations but are often underpowered to find small effect sizes due to subtype heterogeneity in breast cancer and a wide range of clinical covariates. GWAS also tend to detect loci in non-coding regions, which require follow-up studies or colocalization analyses to interpret functionally. Recent work in transcriptome-wide association studies (TWAS) has shown increased power in detecting functionally-relevant, trait-associated loci by leveraging information from expression quantitative trait loci (eQTLs) in external reference panels in relevant tissues. However, race-specific reference panels for TWAS may be needed to draw correct inference in large, racially-heterogeneous cohorts, and such panels for breast cancer are lacking.

Here, we provide a framework for TWAS for breast cancer in diverse populations, using data from the Carolina Breast Cancer Study (CBCS), a population-based cohort that oversampled for women self-identifying as African American. Using Nanostring expression data from CBCS, we perform an eQTL analysis for 417 breast cancer-related genes to train race-stratified predictive models of tumor expression from germline genotypes. We use these models to impute expression in held-out samples from the CBCS and in TCGA, using a permutation method to assess predictive performance accounting for sampling variability. We find that these race-stratified expression models are not always applicable across race, depending on the imputation cohort. Furthermore, their predictive performance varies by breast cancer subtypes. Lastly, we conduct a small-scale TWAS for breast cancer mortality in CBCS (N = 3,828), controlling for covariates used in previous GWAS for breast cancer survival. At an FDR-adjusted *P* value less than 0.1, we find hazardous associations near *CAPN13* (2p23.1), *VAV3* (1p13.3), and *BLK* (8p23.1) and a protective association near *SERPINB5* (18q21.33) in TWAS that are underpowered in GWAS. This approach shows increased power for detection of survival-associated genomic loci, demonstrating the relative strength of TWAS over GWAS.

A carefully implemented TWAS is an efficient alternative to GWAS for understanding the genetics underpinning breast cancer outcomes in diverse human populations and across biologically distinct

tumor types.



## PgmNr 993: Characterization of candidate functional variants and susceptibility genes from a melanoma and multi-cancer risk locus at chr2q33.1 near *CASP8*.

### Authors:

H. Kong<sup>1</sup>; J. Choi<sup>1</sup>; T. Zhang<sup>1</sup>; M. Kovacs<sup>1</sup>; A. Vu<sup>1</sup>; M. Xu<sup>1</sup>; C. Gräwe<sup>2</sup>; A. Chessi<sup>3</sup>; M. Johnson<sup>3</sup>; A. Wells<sup>3</sup>; M. Law<sup>4</sup>; J. Shi<sup>1</sup>; D. Bishop<sup>5</sup>; A. Stratigos<sup>6</sup>; P. Ghiorzo<sup>7</sup>; K. Peris<sup>8</sup>; G. Mann<sup>9,10,11</sup>; G. Radford-Smith<sup>12</sup>; N. Martin<sup>13</sup>; S. Puig<sup>14,15</sup>; D. Duffy<sup>13</sup>; F. Demenais<sup>16</sup>; E. Nagore<sup>17</sup>; D. Whiteman<sup>18</sup>; S. MacGregor<sup>4</sup>; M. Iles<sup>5</sup>; M. Landi<sup>1</sup>; S. Grant<sup>3</sup>; M. Vermeulen<sup>2</sup>; K. Brown<sup>1</sup>; Melanoma Meta-Analysis Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cancer Epidemiology & Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA; 2) Department of Molecular Biology, Oncode Institute, Radboud University, Nijmegen, Netherlands; 3) Center for Spatial and Functional Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 4) Statistical Genetics, The QIMR Berghofer Medical Research Institute, Brisbane, Australia; 5) Division of Haematology and Immunology, Leeds Institute of Medical Research, University of Leeds, Leeds, UK; 6) 1st Department of Dermatology – Venereology, National and Kapodistrian University of Athens School of Medicine, Andreas Sygros Hospital, Athens, Greece; 7) Department of Internal Medicine and Medical Specialties, University of Genoa and Genetics of Rare Cancers, Ospedale Policlinico San Martino, Genoa, Italy; 8) Institute of Dermatology, Catholic University, Rome, Italy; 9) Centre for Cancer Research, Westmead Institute for Medical Research, Westmead, Australia; 10) Melanoma Institute Australia, Wollstonecraft, Australia; 11) University of Sydney, Camperdown, Australia; 12) Inflammatory Bowel Diseases, QIMR Berghofer Medical Research Institute, Brisbane, Australia; 13) Genetic Epidemiology, The QIMR Berghofer Medical Research Institute, Brisbane, Australia; 14) Dermatology Department, Melanoma Unit, Hospital Clínic de Barcelona, IDIBAPS, Universitat de Barcelona, Barcelona, Spain; 15) Centro de Investigación Biomédica en Red en Enfermedades Raras (CIBERER), València, Spain; 16) Team of Genetic Epidemiology and Functional Genomics of Multifactorial Diseases, UMR-1124, INSERM, Université de Paris, Paris, France; 17) Department of Dermatology, Instituto Valenciano de Oncología, València, Spain; 18) Population Health, QIMR Berghofer Medical Research Institute, Brisbane, Australia

---

GWAS have identified a locus at 2q33.1 as associated with melanoma (rs10931936-T,  $P=2.12 \times 10^{-12}$ ) and multiple other cancers, including breast cancer. In order to identify potential causal genes mediating melanoma risk at this locus, we generated an eQTL resource from primary human melanocytes ( $n=106$ ) and assessed candidate genes by both eQTL colocalization (HyPrColoc) and TWAS (FUSION). *CASP8* was the only gene identified by both TWAS and colocalization for this locus, where lower *CASP8* levels are associated with risk (TWAS  $P=5.57 \times 10^{-12}$ ; rs3769821 eQTL  $P=6.14 \times 10^{-12}$ ; colocalization posterior probability=0.98). *CASP8* was also a significant TWAS gene for 29 GTEx tissues, 23 of which showed the same direction of effect as melanocytes. *ALS2CR12* was also a colocalizing eQTL in melanocytes (rs3769821, eQTL  $P=2.42 \times 10^{-5}$ ; colocalization PP=0.37) and TWAS gene in 37 GTEx tissues (marginal TWAS significance in melanocytes), with the risk allele associated with higher *ALS2CR12* expression in all tissues. To identify functional risk-associated variants, we performed a multi-locus massively parallel reporter assay (MPRA) in melanoma cells. Seven high-LD

variants at this locus located within promoter/enhancer-relevant annotations in melanocytes or melanoma cells (Roadmap Epigenomics) were tested. Two variants showed significant allele-specific transcriptional activity (rs3769823, MPRA FDR=1.86x10<sup>-12</sup>, PAINTOR PP=0.76; rs3769821, FDR=0.04, PAINTOR PP=0.04), with the risk allele of rs3769823 demonstrating higher reporter activity (consistent with the *ALS2CR12* eQTL), but the risk allele of rs3769821 showing lower reporter expression. Individual reporter assays revealed that rs3769823 exhibited the strongest and most consistent effect on reporter expression, and appeared to play the largest role on transcriptional activity when tested as a part of common haplotypes for a region including rs3769823 and rs3769821. Quantitative mass spectrometry for rs3769823 using melanoma and breast cancer nuclear extract identified E4F1 and IRF2 as binding preferentially to risk-associated rs3769823-A. Promoter-focused capture-C in primary melanocytes revealed an interaction between rs3769823 and the *ALS2CR12* promoter, a finding we are presently validating via 3C. Our data suggest that the top credible causal set variant rs3769823 likely influences expression of both *CASP8* and *ALS2CR12* in melanocytes, and may be a relevant functional variant for other cancers associated with this locus.

## PgmNr 994: Integrative analysis of epistasis involving oncogenesis-related genes in lung cancer risk development.

### Authors:

Y. Li <sup>1</sup>; X. Xiao <sup>1</sup>; Y. Bossé <sup>2</sup>; Y. Han <sup>1</sup>; C. Amos <sup>1</sup>; Integral-ILLCO group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor College of Medicine, Houston, Texas.; 2) Laval University, Quebec City, QC/Canada

**Background** Epistasis plays important role in cancer development. Our previous study identified significant genetic interactions within oncogenesis-related genes in lung cancer risk development. More genetic interactions may exist between oncogenesis-related genes and outside regions in the genome. An integrative genetic, eQTL gene expression and functional annotation analysis will help us identify the functional epistasis and provide insights about the complicated mechanism in lung cancer.

**Method** The genotypes from two independent lung cancer GWAS studies including a total of 23,351 lung cancer patients and 19,657 health controls with European ancestry were collected for the analysis. Pairwise epistasis was conducted between 27,722 SNPs, from 2,027 oncogenesis-related genes, and 317,624 SNPs from the rest of the genome. Additional genotyping and gene expression data from 409 independent individuals with European ancestry were used to evaluate the effect of identified epistasis on gene expression levels in lung tissues. The epistasis-involved genes, were submitted to DAVID and Reactome for gene functional annotation and pathway analysis.

**Result** Significant genetic interactions were identified between SNPs in gene pairs ATR-GALNT18 (Interaction OR=0.76, p value=7.98x10<sup>-13</sup>) and MET-DPF3 (Interaction OR=0.76, p value=1.62x10<sup>-12</sup>) in lung adenocarcinoma; and PICALM-PDZRN4 (Interaction OR=1.47, p value=1.67x10<sup>-12</sup>) in lung squamous carcinoma. None of these genes have been identified from previous main effect association studies in lung cancer. Further lung eQTL gene expression analysis revealed the significant association in expression levels between joint genotypes at rs637304:rs285581 and the PICALM gene expression (p=0.009). A total of 12 unique genes were submitted to functional annotation and pathway analysis. Three of them (ATR, MET and FHIT) are shown to be related with lung cancer, and six of them (RAD51B, FHIT, CALNT18, RGL1, SYNE1 and TSPAN8) are involved in tobacco-use disorders. The top 10 pathways include TP53 regulates transcription of DNA repair genes (FDR=1.67x10<sup>-2</sup>), homologous DNA pairing and strand exchange (FDR=2.57x10<sup>-2</sup>), and Meiotic synapsis (3.08x10<sup>-2</sup>).

**Conclusion** We identified novel genes in lung cancer risk development by interacting with other genetic variants. The study provides evidence that epistasis explains part of the missing heritability in lung cancer; and complex gene network and pathways contribute to lung carcinogenesis.

## PgmNr 995: Combining multiethnic data improves polygenic risk prediction for breast cancer in East Asians.

### Authors:

Y. Ji<sup>1,2</sup>; R. Tao<sup>1,2</sup>; Y. Yang<sup>4</sup>; J. Long<sup>4</sup>; W. Zheng<sup>4</sup>; B. Li<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetics Institute, Nashville, TN; 2) Vanderbilt Biostatistics Department, Nashville, TN; 3) Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN, USA; 4) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA

**INTRODUCTION:** Both European and East Asian studies have shown that polygenic risk score (PRS) is an important predictor for breast cancer risk. However, it was observed in various traits that PRS predict individual risk more accurately in Europeans than non-Europeans due to much larger sample size and better statistical power. How to best construct PRS for non-European population is still an open question. As many SNP associations should be generalized across populations, we reason incorporating information from higher-powered European study may help increase accuracy in non-Europeans.

**OBJECTIVES:** The aim of this study is to evaluate whether we can improve breast cancer PRS prediction in East Asians through incorporating European information.

**METHODS:** We use the largest GWAS to date for breast cancer in European (n=228,951) and East Asian (n=44,835). A simple approach incorporating European information is to use fixed effect meta analysis for European and Asian GWAS (**FE-META**). However, FE-META approach couldn't take into account linkage disequilibrium (LD) differences between populations. Therefore we proposed a meta analysis approach that leverage LD scores to adjust for effect size differences and tailor the resulting effect size estimate closer to target East Asian population (**ADJ-META**). We construct PRSs using the single population summary (European only, East Asian only), FE-META and ADJ-META by commonly used methods: LD-clumping and pvalue thresholding (P+T) and LDpred. We evaluate prediction performance by area under ROC curve (AUC) in an independent East Asian sample (n=4646).

**RESULTS:** In the East Asian test sample, we observe the AUC of prediction to be 0.608 (European), 0.599 (East Asian), 0.619 (FE-META) and 0.620 (ADJ-META). We also assess the correlation of causal variant effect sizes between European and East Asian in GWAS and observe highly similar effects ( $r_g=0.81$  by POPCORN), supporting the use of joint analysis.

**CONCLUSION:** We have shown that combining European and East Asian data improved prediction accuracy for breast cancer in East Asian than using training data from a single population. These relative improvements help reduce the gap compared to documented risk prediction accuracy in European (0.63-0.68). However, the purposed ADJ-META show negligible improvement over FE-META, potentially related to highly similar LD score estimates and large sample size discrepancies between the two populations.

## PgmNr 996: An overview of causal associations between risk factors' genetic signals and breast cancer survival.

### Authors:

M. Escala Garcia <sup>1</sup>; S. Canisius <sup>1,2</sup>; W. Zheng <sup>3</sup>; S.E. Bojesen <sup>4,5,6</sup>; S. Kar <sup>7</sup>; D.F. Easton <sup>7,8</sup>; P.D. Pharoah <sup>7,8</sup>; M.K. Schmidt <sup>1,9</sup>; Breast Cancer Association Consortium (BCAC)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Division of Molecular Pathology, Amsterdam, The Netherlands; 2) The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Division of Molecular Carcinogenesis, Amsterdam, The Netherlands; 3) Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America; 4) Copenhagen University Hospital, Copenhagen General Population Study, Herlev and Gentofte Hospital, Herlev, Denmark; 5) Copenhagen University Hospital, Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Herlev, Denmark; 6) University of Copenhagen, Faculty of Health and Medical Sciences, Copenhagen, Denmark; 7) University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Oncology, Cambridge, UK; 8) University of Cambridge, Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, Cambridge, UK; 9) The Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Division of Psychosocial Research and Epidemiology, Amsterdam, The Netherlands

Mendelian randomization (MR) offers an attractive approach to test hypotheses of causality of breast cancer survival and assumed breast cancer risk factors. In this study, we provide an overview of causal relationships between breast cancer survival and 15 factors for which genetic variants have been identified: breast density, telomere length, chronotype, body mass index, age at menarche, age at menopause, age at onset of alcohol consumption, ever/never smoked, years of education, celiac disease, depression, schizophrenia and type 2 diabetes mellitus (T2DM).

We performed a two-sample MR using 15-year survival summary estimates of 88,793 patients of European origin with 7,189 breast cancer-specific and 13,962 all-cause mortality events of the BCAC cohort. As instrumental variables for the risk factors, we used genome-wide significant variants from the GWAS catalog. We ran the analyses for overall, breast cancer-specific mortality and for estrogen receptor (ER)-status specific subgroups. We estimated the causal relationships between each of the factors and outcome using an inverse-variance weighted (IVW) method. To test for horizontal pleiotropy effects, we used MR-Egger regression. For those exposures with a significant causal association (FDR < 0.05) we run further sensitivity analyses and applied other MR methods (simple mode, weighted median and weighted mode) to explore the effect under different assumptions.

Genetically increased risk of T2DM (Hazard Ratio [HR] = 1.07, 95% Confidence Interval [CI] = 1.02-1.12, P value [P] = 0.006) and fewer years of education attainment ([HR] = 0.69, [CI] = 0.52-0.91, [P] = 0.010) were significantly associated with worse breast cancer-specific survival. For all-cause mortality, later age at onset for alcohol consumption was associated with better survival ([HR] = 0.96, [CI] = 0.93- 0.99, [P] = 0.011). There were no significant effects for any of the risk

factors in the ER-status subgroups. For all significant associations the sensitivity analyses did not show evidence for violation of the assumptions.

The strongest risk factors for prognosis after a diagnosis of breast cancer relate to the tumor characteristics and the extent of disease at the time of diagnosis. Our data suggest that other patient and lifestyle characteristics may also contribute to variation in outcomes. Such predictors could be included in prognostic models aimed at identifying women most likely to benefit from adjuvant therapies.

## PgmNr 997: TP53 R249S mutation in circulating tumor DNA predicts prognosis of hepatocellular carcinoma patients with or without hepatectomy.

### Authors:

D. Jiang <sup>1</sup>; T. Shen <sup>1</sup>; J. Wang <sup>1</sup>; T. Zhang <sup>2</sup>; S. Zhang <sup>3</sup>; Q. Xiao <sup>4</sup>; P. Lu <sup>5</sup>; T. Chen <sup>5</sup>; L. Yu <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Infectious Diseases and Hepatology Unit, Nanfang Hospital, Southern Medical University, Guangzhou, Guangdong Province, China.; 2) Institute of Cancer, Affiliated Hospital of Jiangnan University, Wuxi, Jiangsu Province, China.; 3) State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai, China.; 4) School of public health, Fudan University, Shanghai, China.; 5) Qidong Liver Cancer Institute, Qidong people's hospital, Qidong, China.

**Purpose:** Hepatocellular carcinoma (HCC) accounts for 80%-90% primary liver cancer, which is the second most lethal tumor worldwide. Due to the lack of effective biomarkers for early diagnosis, the majority of HCC patients are diagnosed at advanced stage, and can only receive conservative treatment rather than surgical resection. In HCC, the tumor suppressor gene *TP53* is frequently mutated with a hotspot mutation at codon 249 (R249S) accounting for >50% of all *TP53* mutations. For HCC patients with hepatectomy, previous studies have indicated that *TP53* mutations detected in tissue specimens were associated with worse clinical outcomes, though the prognostic significance of specific individual *TP53* mutations remains to be clarified. For HCC patients without hepatectomy, however, no study has been conducted to investigate the association of *TP53* mutations with HCC prognosis. In the present study, we aimed to assess the relationship between *TP53* R249S mutation in circulating tumor DNA (ctDNA) and prognosis of HCC patients with or without hepatectomy.

**Methods:** We analyzed 895 HCC patients included in three cohorts. The mutation spectrum of *TP53* was examined by direct sequencing of tissue specimens in HCC patients with hepatectomy (Cohort 1, n=260). R249S or other recurrent mutations were assessed for their associations with overall survival (OS) and progression free survival (PRS) of HCC patients in Cohort 1. R249S harbored in ctDNA was detected through droplet digital polymerase chain reaction (ddPCR) and its association with OS and PRS was analyzed in HCC patients with (Cohort 2, n=275) or without (Cohort 3, n=360) hepatectomy.

**Results:** In Cohort 1, R249S occupied 60.28% of all *TP53* mutations. In addition, three other missense mutations, i.e., R110L, V157F and A159P also occurred in more than two HCC patients. R249S, but not other three recurrent mutations or other *TP53* missense mutations in combination, was significantly associated with both worse OS ( $P=0.006$ ) and PRS ( $P=0.01$ ) of HCC patients. Consistent with the results in Cohort 1, HCC patients in Cohort 2 and 3 with R249S mutation bore worse OS ( $P=8.291 \times 10^{-7}$  and  $2.608 \times 10^{-7}$  in Cohort 2 and 3, respectively) and PRS ( $P=5.115 \times 10^{-7}$  and  $5.900 \times 10^{-13}$  in Cohort 2 and 3, respectively) compared to those without this *TP53* defect.

**Conclusion:** We provided an evidence for the first time that *TP53* R249S mutation in ctDNA may serve as a promising prognosis biomarker for HCC patients regardless of receiving hepatectomy.

# PgmNr 998: Screening of cervical smears, prevalence of hr-HPV sub-types in cervical pathologies and promoter methylation of RAR $\beta$ 2 gene.

## Authors:

S. Thumaju <sup>1,2</sup>; V. Mohan <sup>1</sup>; Y.R. Ahuja <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Vasavi Medical and Research Centre, Dept of Genetics and Molecular Medicine, Hyderabad, India. 040-23210251, Hyderabad, Telangana, India; 2) Osmania University, Dept of Genetics, Hyderabad, India

## Introduction:

Though persistent infection with oncogenic Human Papillomavirus (HPV) types is considered the most important risk factor for cervical cancer development, less than 5% of women with HPV will eventually develop cervical cancer supporting that other molecular event, like gene methylation may co-contribute to cervical carcinogenesis. In India, due to lack of awareness of screening and non-availability of appropriate infrastructure, cervical cancers are detected at advanced stages where cure is difficult.

Identifying women who are sexually active for more than 3 years, with an abnormal papanicolaou (Pap) smear or cervical lesions seen on visual inspection.

Testing PAP DNA samples for the presence of high-risk (12) HPV subtypes to see their prevalence in the population.

Evaluate promoter methylation status of RAR $\beta$ 2 gene to establish the association of epigenetic changes with pathologies of the cervix.

## Method:

Organized screening programs were conducted in different rural and urban areas to identify women with abnormal PAP cytology. Complete information about the screening program was given to the women participants. Pap samples were collected along with the personal and medical history.

## Results:

Pap smears of 530 women were obtained *via* screening programs. 66% of cases were found to show an abnormal Pap cytology, mostly inflammatory smears. Upon exclusion of inflammatory smears, 32% were found to be abnormal smears. Of all abnormal cases 31.8% cases showed RAR $\beta$ 2 methylation. Of the cases detected with presence of HPV, 14.1% showed HR-HPV subtypes. Of these 81.8%, 54.5% cases showed hMLH1 & RAR $\beta$ 2 methylation respectively.

## Conclusion:

RAR $\beta$ 2 gene product inhibits transcription of viral oncogenes (E6 & E7) in HR-HPV immortalized cells. RAR $\beta$ 2 promoter methylation, resulting in loss of expression, may affect this function, giving scope for viral oncogene transcription & carcinogenesis. RAR $\beta$ 2 methylation testing along with screening for HR-HPV subtypes could prove to be biomarkers for potential cancer risk in these patients. It is also important that the prevalence of specific HR-HPV subtypes has to determine in our population before we go ahead with administering the cervical vaccines.



## **PgmNr 999: An empirical comparison of approaches to discovering novel cancer predisposition genes by rare variant analysis using public available summary counts or full genotype data.**

### **Authors:**

W. Chen; S. Wang; D.W. Ellison; G. Wu

[View Session](#) [Add to Schedule](#)

**Affiliation:** St Jude Children's Research Hospital, Memphis, Tennessee.

---

Childhood brain tumors are the second commonest pediatric cancer after leukemia, but they are still rare, with an incidence rate of 0.53% in United States. Therefore, finding novel germline cancer predisposition genes in pediatric brain tumors is challenging. Most published datasets involved a simple Fisher exact enrichment analysis using public summary counts from “healthy” individuals (e.g. gnomAD databases) as a control cohort. The advantage is the large number of the controls. However, the heterogeneity of the sequencing platforms, capture designs and coverage variations will inevitably introduce false-positive and false-negative findings. Moreover, without properly adjusting for the genetic background, the results from a simple Fisher’s exact test are unreliable. A more formal approach is to use raw sequencing data through the same pipeline for the analysis, progressing through mapping, variant calling, variant QC to generate the full genotype data. By rigorously controlling the analytical approach, systematic errors can be avoided and standard rare variant analysis can be applied.

In this study, we compare the two approaches described above. We downloaded whole exome sequencing data for ~5,700 healthy individuals from the Alzheimer's disease Sequencing Project and ~2,600 individuals from the 1000 Genomes Project. We performed rigorous QC to remove regions without good coverage, false variants due to paralogous mapping or sequencing artifacts. Top principal components were included to adjust for population stratification in the rare variant analysis. For the “summary count” approach we contrast the quality controlled genotype counts from cases with those from gnomAD. We also evaluated different population frequency thresholds to define rare variants, as well as different *in silico* functional impact measurements. Using previously classified pathogenic and likely pathogenic variants in Clinvar, we chose REVEL score as the most informative feature.

We chose ~150 brain cancer patients from the Pediatric Cancer Genome Project (PCGP) as study cases. In our analysis, when properly quality controlled, we observed no obvious inflation in QQ plot for both approaches using samples from European ancestry. Some genes are consistently ranked top in both approaches. However, some genes show large discrepancies. We validated *NF1* as a cancer predisposition gene. We also identified a few candidate ‘brain tumor risk’ genes (e.g. *PTPN11*) for further investigation.

# PgmNr 1000: Application of two joint model methods to the CSP380 longitudinal colorectal cancer screening cohort in SNP discovery for advanced neoplasia.

## Authors:

E.R. Hauser<sup>1,2,3</sup>; X. Qin<sup>1,2</sup>; T.B. Hyslop<sup>1,3</sup>; T.S. Redding<sup>1</sup>; K.J. Sims<sup>1</sup>; A. Stone<sup>4</sup>; D. Lieberman<sup>5,6</sup>; D. Provenzale<sup>1,7</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Cooperative Studies Program Epidemiology Center-Durham, Durham VA Health Care System, Durham, NC; 2) Duke Molecular Physiology Institute, Duke University, Durham, NC; 3) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; 4) Cooperative Studies Program Pharmacogenomics Analysis Lab, Central Arkansas Veterans Health System, Little Rock, AR; 5) VA Portland Health Care System, Portland, OR; 6) Oregon Health & Science University, Portland, OR; 7) Department of Medicine, Duke University, Durham, NC

---

Cohort studies following participants over time and making multiple phenotypic measurements may provide information about the presence of germline susceptibility variants for complex traits, however, such studies also suffer from loss to follow-up and death of participants. Two R packages, JMBayes and rstanarm, join the mixed model for longitudinal binary outcomes and the survival model to address repeated measurements and censored data, respectively. Our goal in this work is to evaluate these two packages in the setting of genetic variant discovery. The Department of Veteran Affairs (VA) Cooperative Study 380 is a longitudinal study of 3121 asymptomatic and otherwise healthy participants aged 50-75 from 13 VA Medical Centers who had a screening colonoscopy between 1994-1997. Participants were followed over 10 years capturing the histopathological outcomes for all colonoscopies or other colon-related procedures along with ascertainment of vital status from the National Death Index and the VA Electronic Medical Record (EMR). Advanced neoplasia (AN) at each colonoscopy was defined using pathology reports indicating at least one polyp with tubular adenoma  $\geq 10$  mm, adenoma with villous histology, adenoma with high-grade dysplasia or invasive cancer. Time to death from all causes or time to last visit was recorded. A total of 612 individuals were genotyped on the Illumina Infinium Omni2.5-8 v1.3 Beadchip. We analyzed the 500 European American participants, randomly selected 150 SNPs to evaluate the results under the null hypothesis and compared the 50 top SNPs from logistic regression of AN to check consistency. We fit a random effects model to all colonoscopies including an additive genetic term, age and two genetic principal components (PCs). We also fit a Cox proportional hazards model to time to death including baseline screening outcomes along with age, comorbidity, and follow-up group. The models are linked by the random effect. We fit the same model to both packages (JMBayes and rstanarm): 127 random SNPs ran successfully with rstanarm and the QQ plot gave a  $\lambda=0.92$ ; 84 random SNPs ran successfully with JMBayes and the QQ plot gave a  $\lambda=2.3$ . Of the 50 top SNPs, 36 ran successfully in both packages, the correlation coefficient is 0.498,  $p$ -value= 0.0019. From these comparisons we conclude that JMBayes will require adjustment of  $p$ -values if used in a GWAS context. Both packages require substantial tuning to run large numbers of SNPs.

## PgmNr 1001: Variable selection for SNP-SNP interaction models: AA9Lasso approach.

**Authors:**

H. Tung; H. Lin

[View Session](#) [Add to Schedule](#)**Affiliation:** Biostatistics, LSUHSC New Orleans, New Orleans, LA.

---

The ideal goal of genetic association studies is to improve the prediction of clinical outcomes. It has been shown that SNP-SNP interactions can explain complex diseases better than individual SNP effects. The AA9int method is a powerful method to test SNP-SNP interactions associated with an outcome by considering the non-hierarchical structure. It is beneficial to build a prediction model using AA9int SNP interaction findings. For SNP interaction analyses, the number of predictors increases dramatically due to consider all possible pairwise interactions. An effective preliminary selection can help on selecting a subset of SNP pairs for building a multivariable model. For a small sample size, SNP pairs with the main effect term(s) are often neglected during the selection process. We proposed AA9Lasso, a modified Group Lasso approach with proper weights of the penalty terms used for variable selection for AA9int identified SNP pairs. The goal of the new method is to increase detection rates for the pairs with complex interaction structures, such as Full and One-Main structure. We conducted a simulation study to compare AA9Lasso with univariate approaches and regular Group Lasso. We observed AA9Lasso outperformed other testing methods. Its detection rates of all the designed pairs were higher than 90% under a large sample size of 5000. For  $>0.2$  effect size and 0.45 of the minor allele frequency (MAF) for both SNPs, AA9Lasso had almost 100% detection rate with a sample size of 5000. AA9Lasso achieved the goal to increase detection rates of SNP pairs with a complexed structure; meanwhile, the pairs with the interaction only structure also had excellent detection rates. When the sample size reduced to 3000 and 1000, the proposed method still performed better than others. Both MAF and effect size impacted the performances of selection, but the proposed method maintained its advance. This study demonstrates that AA9Lasso can effectively select a subset of SNP pairs associated with the outcome. Although the proposed AA9Lasso already ruled out  $> 97\%$  of the false positive findings, it remained thousands of predictors. Thus, another variable selection may be needed to build the final prediction model.

## PgmNr 1002: DOT: Gene-set analysis by combining decorrelated association statistics.

**Authors:**

O. Vsevolozhskaya <sup>1</sup>; V. Zaykin <sup>2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Biostatistics, University of Kentucky, Lexington, Kentucky.; 2) Biostatistics and Computational Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC 27709, USA.

---

Historically, the majority of statistical association methods have been designed assuming availability of SNP-level information. However, modern genetic and sequencing data present new challenges to access and sharing of genotype-phenotype datasets, including cost management, difficulties in consolidation of records across research groups, etc. These issues make methods based on SNP-level summary statistics for a joint analysis of variants in a group particularly appealing. The most common form of combining statistics is a sum of SNP-level squared scores, possibly weighted, as in burden tests for rare variants. The overall significance of the resulting statistic is evaluated using its distribution under the null hypothesis. Here, we demonstrate that this basic approach can be substantially improved by decorrelating scores prior to their addition, resulting in remarkable power gains in situations that are most commonly encountered in practice; namely, under heterogeneity of effect sizes and diversity between pairwise LD. In these situations, the power of the traditional test, based on the added squared scores, quickly reaches a ceiling, as the number of variants increases. Thus, the traditional approach does not benefit from information potentially contained in any additional SNPs, while our decorrelation by orthogonal transformation (DOT) method yields steady gain in power. We present theoretical and computational analyses of both approaches, and reveal causes behind sometimes dramatic difference in their respective powers. We showcase DOT by analyzing breast cancer data, in which our method strengthened levels of previously reported associations and implied the possibility of multiple new alleles that jointly confer breast cancer risk.

## PgmNr 1003: Search for additional high-risk genes for familial cutaneous malignant melanoma using unified linkage and association analyses of whole exome sequencing data.

### Authors:

S. Yepes Torres <sup>1</sup>; M. Tucker <sup>1</sup>; H. Koka <sup>1</sup>; K. Jones <sup>2</sup>; A. Vogt <sup>2</sup>; L. Burdette <sup>2</sup>; W. Luo <sup>2</sup>; B. Zhu <sup>2</sup>; M. Yeager <sup>2</sup>; B. Hicks <sup>2</sup>; N. Freedman <sup>1</sup>; S. Chanock <sup>1</sup>; A. Goldstein <sup>1</sup>; X. Yang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) DCEG, NCI, NIH. Rockville, MD; 2) CGR, Leidos Biomedical Research, Inc. Rockville, MD

---

Approximately 10% of cutaneous malignant melanoma (CMM) cases occur in a familial setting but known high-risk CMM genes account for melanoma risk in less than 40% of melanoma-prone families. Even with the advance of sequencing technologies, finding novel high-risk CMM genes has been challenging, largely due to genetic heterogeneity, variable penetrance, and difficulty in variant interpretation. New analytical approaches with the design of integrating multiple sources of data may have the potential to address some of these challenges. Here, we performed Whole Exome Sequencing in germline DNA of 89 CMM patients from 28 families without known mutations to identify novel high-risk genes for familial CMM. We used pedigree-Variant Annotation Analysis and Search Tool (pVAAST), which uses a composite likelihood ratio test incorporating linkage signal in families, external controls, and functional prediction of variants to identify rare variants that are statistically associated with disease. pVAAST analysis identified 20 genes with statistical significance (p-value < 0.05 after  $1.0^6$  permutations) that were recurrently altered in  $\geq 3$  families. Among them, ASAP3, SRCAP, DDX60, GORASP1, MC1R and ATM showed high quality reads and had variants that were rare or absent in public databases. MC1R and ATM have been previously described as CMM susceptibility genes. GORASP1 encodes a protein required for the maintenance of the Golgi apparatus integrity. In addition to GORASP1, variants in several other genes associated with Golgi apparatus, including the previously described CMM high-risk gene GOLM1, were found cosegregating in six families, further supporting the role of Golgi complex genes in CMM susceptibility. Several families carried variants in more than one candidate gene, highlighting some of the complexities in disease gene discovery. In future work, we plan to investigate additional approaches including Rare Variant Sharing (RSV), a test to assess association and linkage between rare genetic variants and disease phenotype in pedigrees, rare variant burden test, and outlier test using gene expression data. We will compare the results from the analysis using pVAAST with results from the other complementary approaches to increase opportunities for identifying high-risk genes for CMM.

## PgmNr 1004: Transcriptome-wide association study of endometrial cancer.

### Authors:

P.F. Kho; A.B. Spurdle; T.A. O'Mara; D.M. Glubb; Endometrial Cancer Association Consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** QIMR Berghofer Medical Research Institute, Brisbane, Australia

---

Endometrial cancer is the most commonly diagnosed gynaecological cancer. We have recently performed a meta-analysis of endometrial cancer genome-wide association studies (GWAS; total 12,906 cases and 108,979 controls) and identified 16 genetic loci that robustly associate with endometrial cancer. However, 99% of endometrial cancer risk-associated variants are located in non-coding regions. Further, because of enhancer-promoter chromatin looping, the genes involved in disease aetiology are not necessarily the most proximal genes to the GWAS variants. Thus, elucidating the causal genes underlying GWAS association remains challenging. To address this issue, we have used transcriptome-wide association study (TWAS) techniques to identify genes whose expression levels associate with endometrial cancer. Using findings from the largest eQTL reference panel for blood (N = 31,684) and the endometrial cancer GWAS meta-analysis, Summary data-based Mendelian Randomization (SMR) analysis prioritised *EVI2A* as a candidate gene for endometrial cancer. We also conducted a TWAS using Fine-mapping of CaUsal gene Sets (FOCUS) on the same GWAS of endometrial cancer to prioritise credible genes by fine-mapping TWAS associations identified from the GTEx Project eQTL reference panel. FOCUS analysis prioritised *BHLHE41* (Posterior Inclusion Probability; PIP 0.99), *SRCIN1* (PIP 0.99), *SNX11* (PIP 0.92), and *RNF217* (PIP 0.85) as candidate genes. Notably, chromatin looping data from endometrial cell lines supports the targeting of *BHLHE41*, *SRCIN1* and *SNX11* by endometrial cancer risk-associated variants. Notably, *BHLHE41* and *SRCIN1* play a role in the epithelial-mesenchymal transition in cancer and *BHLHE41* is upregulated in endometrial tumours. *SNX11* is involved in intracellular trafficking but has not been previously related to cancer. However, reporter gene analysis of a *SNX11* promoter endometrial cancer risk-associated variant was consistent with the TWAS finding that increased *SNX11* expression was associated with increased risk. In summary, using TWAS and supportive functional genetic evidence, we have prioritised several candidate genes for assessment of their effects in cellular studies of endometrial cancer.

## PgmNr 1005: Deconstructing myeloma tumor heterogeneity: Transcriptome dimension signatures.

### Authors:

R.G. Waller; M.J. Madsen; J. Gardner; D. Sborov; N.J. Camp

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of Utah School of Medicine

---

Multiple Myeloma (MM), the 2nd most common hematologic malignancy, is a complex cancer of plasma cells with 32,110 new cases expected in 2019 and poor prognosis (52.2% survive 5 years). Tumor heterogeneity hinders MM clinical management and research. A standard approach is to partition tumors into mutually exclusive, categorical subtypes. However, categorical subtypes are unidimensional and may fail to capture important variation.

An alternate approach is quantitative expression dimensions. Tumor heterogeneity is deconstructed using principal component analysis (PCA) to establish tumor traits that capture important global transcriptome variation. Each dimension is an independent tumor characteristic – a unique weighted average of key genes – which alone could uncover associations with clinically-relevant endpoints. In addition, together multi-dimension signatures may uncover biologically relevant characteristics and reflect specific molecular liabilities or therapeutic vulnerabilities.

MM tumor dimensions were derived from RNA sequencing on treatment-naïve, CD138 sorted, tumor cells on 768 patients in the Clinical Outcomes on MM Genetic Profiles Assessment (CoMMpass) study. Gene-based SALMON expression counts were normalized for gene length, library size, and RNA composition. Multi-stage PCA was performed on the normalized counts to derive orthogonal, quantitative tumor dimensions. Distinct dimensions and multi-dimension signatures were investigated for associations with demographic, clinical, and genetic characteristics using penalized linear regression modeling.

We identified 28 dimensions of interest in the MM tumors and characterized multi-dimension signatures for each of the 768 patients. After correcting for multiple testing, 29 associations between a dimension and a clinical variable were significant. Dimensions 1-8 were associated with one or more somatic alterations, known to correlate with myeloma prognosis, including t(11;14), t(14;16), t(4;14) and amp(1q) (max  $p = 8.22 \times 10^{-5}$ , min  $p = 1.61 \times 10^{-23}$ ). Interestingly, dimensions 11 and 24 were significantly associated with African heritage ( $p = 2.86 \times 10^{-7}$  and  $8.04 \times 10^{-7}$ ), indicating the possibility of molecular disparities by race.

We present a new quantitative framework to represent transcriptome variability in MM tumors that provides more flexibility for statistical modeling with clinically relevant endpoints and ultimately, the potential to improve precision cancer care.



## PgmNr 1006: Investigation of novel transcript splicing in multiple myeloma reveals an ultra-high risk subgroup and potential novel therapeutic targets.

### Authors:

M.A. Bauer <sup>1,2</sup>; A. Ashby <sup>1,2</sup>; C. Wardell <sup>1,2</sup>; E. Boyle <sup>3</sup>; G. Morgan <sup>3</sup>; B. Walker <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, AR; 3) NYULangone Medical Center, Perlmutter Cancer Center, NYU Langone System, NYC, NY

Disruption of the normal splicing patterns of RNA is a major factor in the pathogenesis of a number of cancers. Increasingly research has shown the strong influence that splicing patterns can have on cancer progression. Multiple Myeloma is a molecularly heterogeneous disease classified by the presence of key translocations, gene expression profiles and mutations. The role of alternative splicing in MM remains largely unexplored.

Working with RNA-Seq data from the CoMMpass dataset, we split the newly diagnosed MM samples ( $n=598$ ) into three groups. The initial groups consisted of the top and bottom 20% (both  $n=120$ ) and middle 60% ( $n=358$ ) in number of novel splice loci. There was a significant difference in progression free survival (PFS) and overall survival (OS) between the low novel splice group and the high splice group ( $p=0.03$  and  $0.013$ ). Surprisingly, a comparison of the groups with known MM genomic characteristics revealed significantly less novel splicing in the  $t(4;14)$  group ( $p\text{-value} = 0.002$ ), a poor prognosis MM subgroup and an increase in the  $t(11;14)$  group ( $p\text{-value} < 0.001$ ), a better prognosis subgroup. Survival analysis for just the  $t(4;14)$  subgroup showed a significant adverse survival in the high splice group within  $t(4;14)$  versus  $t(4;14)$  with low number of novel splice groups (PFS:  $p=0.003$  and OS:  $p<0.001$ ). This finding suggests that there is a previously undescribed ultra-high risk group of  $t(4;14)$ .

Differential gene expression analysis identified 619 upregulated genes and 697 down regulated genes ( $\log_2$  fold-change  $> 2.0$  and adjusted  $p\text{-value} < 0.05$ ). A number of MAGE genes were over expressed in the high splice group. Gene set enrichment analysis revealed an enrichment in genes involved in the G2/M checkpoint and E2F (family of transcription factors) target pathways and a decrease of the p53 DNA repair and unfolded protein response pathways. Protein-protein analysis resulted in a highly connected network with significant enrichment for several protein domains including MAGE family, GAGE protein, and SSXRD motif. These present potential therapeutic targets with fewer side effects due to their limited expression in somatic tissue.

Alternative splicing is an important pathogenic disease mechanism in MM that affects important pathways. In this study, we show that a high number of novel splice loci is associated with adverse survival and an ultra-high risk group. We also identify a number of genes that may represent novel therapeutic targets.



## PgmNr 1007: An enhanced genetic model of colorectal cancer progression history.

**Authors:**

L. Yang

[View Session](#) [Add to Schedule](#)**Affiliation:** Ben May Department for Cancer Research, Univ Chicago, Chicago, Illinois.

---

The classical genetic model of colorectal cancer presents somatic *APC* mutations as the earliest genomic alterations, followed by *KRAS* and *TP53* mutations. However, the timing and relative order of clonal expansion and other types of somatic genome alterations such as genomic rearrangements are still unclear. Here, we performed detailed analysis of somatic genetic alterations, including point mutations, copy number alterations and genomic rearrangements, in 63 whole-genome sequenced colorectal cancers from the Cancer Genome Atlas Research Network. The relative order of these alterations occurring during tumorigenesis was inferred from variant allele fractions. We found that driver point mutations, gene fusions, and arm level copy losses typically arise early. Copy-neutral loss of heterozygosity is often a two-step event: a deletion followed by a duplication. Different mechanisms act in different genomic regions to optimize the DNA dosage. Chromothripsis, clustered genomic rearrangements previously thought to occur as a single catastrophic event, is frequent and may occur multiple times independently in the same tumor through different mechanisms. In contrast to recent studies reporting neutral growth of tumors, selection is often present on subclones. Our results suggest that different evolutionary models can operate in a single tumor at different stages. Combining these results, we present a refined tumor progression model for human colorectal cancer. Our enhanced genetic model significantly expands our understanding of tumorigenesis process.

## PgmNr 1008: Comprehensive assessment of gene fusion detection, prioritization, annotation, and characterization.

### Authors:

V. Chakraborty; S. Ramachandrula

[View Session](#) [Add to Schedule](#)

**Affiliation:** Strand Life Sciences Private Limited, Bengaluru, Karnataka, India

---

Fusion events are commonly observed in cancer and are prevalent in both hematological and solid tumors. Though these events seem to be random but they show classic non-random behavior both structurally and functionally. For example, hematological tumors mostly results in juxtaposition of a normal gene under the regulation of a new gene, resulting in abnormal expression of the normal gene. In case of solid tumors most of the known gene fusions are either Transcription control genes or Tyrosine kinases, which are potential targets for drug development. Hence, to understand the behavior of both the groups of tumors, as well as the tissue and lineage specificity of the events, we chose targeted RNA-Seq data in addition to whole transcriptome data from various types of blood, lung and prostate cancers from publicly available datasets: PRJNA484669 and GSE80126.

In this study, we used Strand NGS to perform an end to end analysis starting from aligning the raw data, filtering reads based on quality, perform gene fusion detection, filter the fusion events to remove false positives, annotate against already known fusion events, quantify the data to understand the change in expression levels caused due to these events and lastly link the biological effect of the fused genes.

We utilized the ability of Strand NGS to align raw reads from RNA data to the exonic (known transcriptome) as well as intronic parts of the genome. Apart from gene fusions, we also looked for any co-occurrence of other significant mutational events like SNP or MNP.

Usually, the trickiest part after mutational detection is to assess these events; this was readily possible by viewing the fusion events in the elastic genome browser of Strand NGS which is unique in its nature as among other properties it gives a one shot view of even the inter-chromosomal gene fusion events up to nucleotide level resolution. Finally, we tie all the sequence changes in the data with expression level changes and their functional significance via GO and Pathway analysis.

Testing for the gene fusion events helps in disease diagnosis, identifying therapeutic treatment, prognosis information and works as a guide to take targeted therapy decisions. A comprehensive GUI-based tool like Strand NGS allows biologists to analyze gene fusion events and also correlate it with the biological effects caused by these events.

## PgmNr 1009: Optimized computational genotyping of known structural variation for clinical diagnoses.

### Authors:

V. Chander <sup>1,2</sup>; R. Gibbs <sup>1,2</sup>; F. Sedlazeck <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas

---

Structural Variation (SV) is a major contributor to genomic variation and associated with many genetic diseases. Despite the importance of this class of variation, there are shortcomings in the usual *de novo* methods for their assessment using short DNA sequencing. The short reads alone provide low sensitivity and a high false discovery rate, significantly hindering applications for routine SV screening and limiting their proper characterization for clinical diagnoses. These shortcomings can be partially ameliorated with the use of long-reads, however, the current cost, DNA requirements and time-intensive nature preclude their routine application for clinical diagnostics.

We investigated SV genotyping as a cost-effective diagnostic tool in the clinic. SV genotyping enables computational detection of known annotated SVs directly from short read DNA sequence data in patient samples, which dramatically reduces the false discovery rate and thus, the potential misdiagnosis of patients. Further, the validation of methods for detection of known events provides an easier challenge for regulatory and compliance purposes than *de novo* detection.

We assessed five state-of-the-art SV genotyping software methods (Delly, Genome STRiP, STIX, SVtyper, SV2) based on short-read sequence data and investigated their potential applicability for clinical diagnosis. These methods were characterized based on their ability to genotype SV types across a range of simulated and real data. We determined which genotype software optimally performs for each type and influence of size of SVs on sensitivity. Optimal tuning of calling parameters markedly improved sensitivity and reduced initial false discovery calls. Among the SV genotypers, STIX performed best on both simulated and GiaB based SVs calls, demonstrating a good balance of high sensitivity versus reduced false discovery. In aggregate, our results indicate SV genotypers have better performance than SV callers. Our approach can be integrated to routinely scan for known pathogenic SVs into existing analysis pipelines, thereby representing a computationally efficient and rapid way to diagnose SVs in clinical diagnostic settings.

## PgmNr 1010: TACCO, a database, connects transcriptome alterations, pathway alterations, and clinical outcomes in cancers.

### Authors:

T. Chen <sup>1,2,3</sup>; P. Chou <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of bioinformatics and system biology, National Chiao Tung University, Hsinchu, Taiwan; 2) Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, 30068, Taiwan; 3) Center For Intelligent Drug Systems and Smart Bio-devices (IDS2B), National Chiao Tung University, Hsinchu, Taiwan; 4) Molecular Medicine Research Center, Chang Gung University, Taoyuan, Taiwan.

---

While numerous cancer sequencing data is already publicly available, to systematically dig out meaningful correlations from those data is still challenging for cancer biologists who do not possess related computer skills. Previously, we have downloaded all the expression levels of miRNA and mRNAs from firehose and build a database, Transcriptome Alteration in CanCer Omnibus (TACCO), which aims to link the transcription alterations and transcriptome regulatory network with downstream pathway alterations and clinical outcomes in 22 cancer types. TACCO provides: 1. Volcano plots for selection of significantly differentially expressed miRNA or mRNA (DEGs) 2. Correlations between the expression levels of miRNA and their targets 3. Gene Set Enrichment Analysis for significantly altered pathways 4. KEGG pathway enrichment analysis for selected or uploaded gene list 5. Gene Ontology (GO) enrichment analysis for selected or uploaded gene list 6. Survival-related signature selection from transcriptome. DEGs selected on TACCO or a gene list uploaded by user can be used for the pathway analysis and survival prediction model construction. These signatures have great potentialities for patient stratification and treatment decisions in later clinical applications. Based on TACCO, we include extra genomic features in the prediction models and found further improvement. As an integrated transcriptome database, our tool comes with a user-friendly query interface will be invaluable for cancer biologist to investigate the transcriptome regulatory networks alterations and the following clinical outcome in cancers systematically.

## PgmNr 1011: Representing cancer-relevant variants in variation graphs.

### Authors:

E.T. Dawson <sup>1,2</sup>; D.M. Karyadi <sup>1</sup>; S.J. Chanock <sup>1</sup>; R. Durbin <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; 2) Department of Genetics, University of Cambridge, Cambridge, UK

---

Variation graphs provide a single data structure that can represent both a reference genome and reference-relative variation. Compared to a linear reference, graphs improve read mapping and reduce reference bias when mapping against common population variation from sources such as the 1000 Genomes project. This has been shown to improve genotyping of single-base variants and indels. Graphs have also been shown to benefit structural variant calling and genotyping.

Tumors possess both a germline background and somatic variation. Nearly 200 genes have been implicated as contributing to cancer predisposition, and these genes may each have many specific predisposition variants. There are over 4,000 known germline predisposition variants among nearly 25,000 total catalogued mutations in the BRCA1 and BRCA2 genes for instance. There is significant interest in quickly and accurately detecting such variants in whole-genome sequencing data. We demonstrate a pipeline for building graphs of many thousands of predisposition variants, implemented using the variation graph toolkit (vg), and show that reads can be directly mapped to these variants.

We also build graphs of somatic variants from the Catalog of Somatic Mutations in Cancer: one containing more than 4 million coding mutations and one with tens of thousands of structural variants. Somatic calling is often confounded by low tumor purity, alternative ploidies, and tumor subclonality, making improved read mappings especially valuable as a small amount of additional evidence for a variant. However, one must carefully choose the variants included in such a graph and their representation, especially when incorporating structural variation. We discuss specific decisions that must be made when selecting variants to include and their representation, as well as how these factors impact the performance of downstream analyses. Lastly, we demonstrate that mapping reads from a sample with known somatic variants (such as BRAF V600E) is feasible with vg and compare our results to traditional variant callers.

## PgmNr 1012: A combinatorial approach for single-cell variant detection via phylogenetic inference.

**Authors:**

M. Edrisi<sup>1</sup>; H. Zafar<sup>2</sup>; L. Nakhleh<sup>1</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Computer Science, Rice University, Houston, Texas.; 2) Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania

---

Single-cell sequencing provides a powerful approach for elucidating intratumor heterogeneity by resolving cell-to-cell variability. However, it also poses additional challenges including elevated error rates, allelic dropout, and non-uniform coverage. A recently introduced single-cell-specific mutation detection algorithm leverages the evolutionary relationship between cells for denoising the data. However, due to its probabilistic nature, this method does not scale well with the number of cells. Here, we report on a novel combinatorial approach for utilizing the genealogical relationship of cells in detecting mutations from noisy single-cell sequencing data. Our method, called scVILP, jointly detects mutations in individual cells and reconstructs a perfect phylogeny among these cells. The method employs a novel Integer Linear Program algorithm for deterministically and efficiently solving the joint inference problem. scVILP achieves similar or better accuracy but significantly better runtime over existing methods on simulated data. We also applied scVILP to an empirical human cancer dataset from a high grade serous ovarian cancer patient.

## PgmNr 1013: Benchmarking copy number aberration detection tools on single-cell DNA sequencing data.

### Authors:

Fan <sup>1</sup>; M. Edrisi <sup>1</sup>; N. Navin <sup>1</sup>; L. Nakhleh <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, Rice University, Houston, Texas.; 2) Department of Genetics, MD Anderson Cancer Center, Houston, Texas

---

Copy number aberrations (CNA) have been shown to cause genetic diseases such as cancer. Accurate detection of CNAs can help with diagnosis, prognosis and treatment of cancer. Different from bulk sequencing, single-cell DNA sequencing provides a way of looking into the CNAs one cell at a time, making it possible to detect CNAs occurring to small clones and understand the heterogeneity of cancer. However, most CNA detection methods that have been applied to single-cell DNA sequencing data are not specifically designed for single cells. While single-cell DNA sequencing has its unique error profile such as low and uneven coverage, a benchmark study on the performance of these methods when applied to single cell DNA data is currently lacking. We designed experiments on both simulated and real data to measure the recall and precision of three methods that have been used to detect CNAs in single-cell DNA sequencing data--Ginkgo, HMMcopy and CopyNumber. Specifically, we designed a single-cell DNA simulator on which the CNA rate, CNA size and read coverages are all stochastically sampled from their corresponding distributions that are estimated from and thus consistent with the real data. We looked into four aspects of the three methods in the simulation experiments: the scalability and the affect of ploidy, coverage and doublets. Furthermore, using real data, we investigated the consistency among the three methods, and analyzing their inferences with respect to the evolution of copy numbers. The study highlights the strengths and weaknesses of the existing methods, and identifies areas of improvement.

## PgmNr 1014: Imputation of the *GSTM1* deletion in the 1000 Genomes Project populations to develop a reference dataset for epidemiological studies.

### Authors:

O. Florez-Vargas; W. Yan; L. Prokunina-Olsson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland.

---

The deletion of the *GSTM1* gene on chromosome 1p13.3 is a common germline polymorphism. The homozygous loss of *GSTM1* (*GSTM1*-null) is likely to be functional because *GSTM1* detoxifies various environmental carcinogens and toxins and multiple associations have been reported for susceptibility to inflammatory and metabolic pathologies and cancer risk in carriers of this genotype. Large-scale genotyping of this deletion is challenging due to the complex genomic structure of the region. Using a custom assay, we genotyped *GSTM1* deletion in 1840 individuals representing 19 populations from the 1000 Genomes Project (1KGP). *GSTM1*-null status was found in 39.5% samples, with strong inter-population differences ( $p < 0.001$ ). The *GSTM1* deletion was most strongly linked with rs36209093 ( $r^2=0.43$ ,  $D'=0.81$  in the whole set). However, we found that rs36209093 is a pseudo-SNP that results from a near-complete similarity within two duplicated regions. Considering that *GSTM1* deletion is not tagged by any easy-to-genotype proxies, we aimed to explore whether it can be imputed to facilitate its association analysis with relevant outcomes in GWAS.

We randomly partitioned the 1840 *GSTM1*-genotyped 1KGP samples into a reference panel ( $n=846$ ) and a test panel ( $n=852$ ), with each panel containing ~50% of samples from each population. Haplotype phasing within a 2-Mb window around *GSTM1* deletion was done with ShapeIT software, and *GSTM1* deletion imputation was done with IMPUTE2 software with and without pre-phasing of the test panel. *GSTM1* deletion was scored in samples with a post-probability dosage of more than 0.6. Post-imputation quality parameters were calculated based on genotyped vs. imputed *GSTM1* deletion. We observed lower concordance rate (85% vs. 90%) and imputation quality score (0.76 vs. 0.84) but a higher rate of imputable samples (96.7% vs. 93.1) and shorter computation time (4 vs. 16 hrs) for imputation with vs. without pre-phasing of the test panel. In conclusion, our results demonstrate that to explore its associations with relevant phenotypes, *GSTM1* deletion can be imputed with reasonable quality in existing GWAS datasets using the modified 1KGP or custom reference panels.



## PgmNr 1015: Identification of synthetic cytotoxic gene pairs associated with a therapeutic response for the anticancer drug etoposide based on cell line-drug database.

### Authors:

J. Goh <sup>1,2</sup>; Y.J. Kim <sup>3</sup>; K. Kim <sup>4</sup>; K. Lee <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea, Republic of; 2) Department of Chemical Engineering, Pohang University of Science and Technology(POSTECH), Pohang, Korea, Republic of; 3) Center for Precision Medicine, Seoul National University Hospital, Seoul, Korea, Republic of; 4) Division of Clinical Bioinformatics, Seoul National University Hospital, Seoul, Korea, Republic of

In cancer genomics profiling, the most important clinical question was whether the patients show treatment responsiveness as well as low toxicity. Although traditional chemotherapeutic agents form the basis of treatment for many cancers at different stages, resistance and toxicity often result in therapeutic failure. In order to evaluate the genomic signature associated with the chemotherapeutic response, the synthetic cytotoxic interaction (SCI) of the anticancer drug etoposide was analyzed based on the cell line-drug database, GDSC. To find SCI gene pairs for etoposide, we defined significantly damaging mutations if a gene have a more than one loss of function mutations and/or nonsynonymous mutations with SIFT score<0.05 and/or copy number deletion. IC50 value of etoposide for each cell line was used as drug responsiveness marker to decide whether cell line showed greater sensitivity to etoposide when both genes were damaged compared with when one or none of the genes have a damaging mutation. There were 871 cell lines having etoposide IC50 values as well as whole exome sequencing data in GDSC database. To compare the SCI effect for etoposide, we grouped cell lines into 4 groups according to mutation profile for each gene pair: PP, PN, NN and NP, and P represent gene have damaged mutation and N represent no mutation, respectively. There were 3,804 genes have at least 5 damaging mutations (7,233,306 gene-gene pairs), and Student's t-test was used to identify gene pair showed significantly lower IC50 of PP group compare to other three groups adjusting multiple testing correction.

Total of 14 gene pairs containing 24 genes were identified to have significant SCI for etoposide (p value<0.05). After functional enrichment using GO terms, five genes (*CRB2*, *RYR2*, *LRP1B*, *CACNA1E*, and *DNAH7*) were enriched to calcium ion binding (GO:0005509). We analyzed correlation between the number of mutations in five calcium ion binding genes and IC50 values for all cell lines and identified mild but significant decreasing correlation (p value<0.05).

We identified that cancer cell lines showed interactive gene-gene paired synergistic effect to etoposide especially enriched in calcium ion binding which was biologically related to a well-known etoposide side effect, cardiac toxicity, and this method have great potential which could be broadly applied to other drugs and genes as well to identify sensitive genomic profiles in cancer cell.

# PgmNr 1016: Interpretable deep learning model to integrate genomic profiles and histopathological images for cancer survival prediction.

## Authors:

J. Hao <sup>1</sup>; S.C. Kosaraju <sup>2</sup>; N.Z. Tsaku <sup>2</sup>; H.M. Koh <sup>3</sup>; D.H. Song <sup>3</sup>; J.H. Oh <sup>4</sup>; M. Kang <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Analytics and Data Science Institute, Kennesaw State University, Kennesaw, GA; 2) Department of Computer Science, Kennesaw State University, Marietta, GA; 3) Department of Pathology, Gyeongsang National University Changwon Hospital, Changwon, Republic of Korea; 4) Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY

---

Integration of heterogeneous types of biological data promises an improvement in predicting cancer survival. Specifically, the integration of genomic data and histopathological images enhances survival predictions and personalized treatments in cancer study, while providing an in-depth understanding of genetic mechanisms and phenotypic patterns of cancer. Histopathology, as a clinical gold-standard tool in diagnosis and prognosis for many cancers, allows clinicians to make a better decision on therapies. Whereas, the genetic mechanism of cancer can be unveiled with high-throughput genomic data. We propose a biologically interpretable deep learning model that integrates genomic data and pathological images, not only for improving survival predictive performance but also identifying genetic and pathological patterns that cause different survival of patients. Our model consists of genome/pathology/demography-specific layers, each of which provides biological interpretability. In particular, we adapt our previous pathway-based sparse deep neural network, named Cox-PASNet, for the genome-specific layers. We propose a novel patch-wise texture-based convolutional neural network with a patch aggregation strategy to extract survival-discriminative features without manual annotation for pathology-specific layers. Survival-discriminative features are extracted from a pre-trained deep learning model with uncensored data, and the features are aggregated from multiple patches of a whole slide image. Finally, the high-level representations of genomic and pathological data along with age are introduced to a shared layer that estimates Prognostic Index (PI) in a Cox proportional-hazards regression model. We examined gene expression data and pathological images of Glioblastoma Multiforme (GBM), downloaded from The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA), to assess our proposed model. The proposed model achieved 0.67 of C-index, which is higher than the results of recent reports with only pathological images of TCGA GBM data (0.64 of C-index) and Cox-PASNet (0.63 of C-index). More importantly, our proposed integrative deep learning model can identify genetic and pathological prognostic factors simultaneously associated with patients' survival. We will demonstrate how our integrative model can represent a complex biological system with multiple types of data and discuss a new finding of genetic and pathological patterns inferred from our model.

## PgmNr 1017: The Reactome Pathway Knowledgebase and ReactomeFIViz app.

### Authors:

R. Haw <sup>1</sup>; G... Wu <sup>2</sup>; H. Hermjakob <sup>3</sup>; P. D'Eustachio <sup>4</sup>; L. Stein <sup>1</sup>; Reactome Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) OICR, Toronto, ON, Canada; 2) OHSU, Portland, OR, USA; 3) EMBL-EBI, Hinxton, UK; 4) NYU School of Medicine, New York, NY, USA

---

Reactome (<https://reactome.org>) is an open-access, open-source, open-data, open-graphics, curated and peer-reviewed knowledgebase that stores full descriptions of pathways and biological processes. Geneticists, clinicians, molecular biologists, bioinformaticians and systems biologists use Reactome to interpret high-throughput experimental datasets, develop novel algorithms for data mining and visualization, and build predictive models of normal and abnormal pathways. Recent extensions of our data model accommodate the annotation of drug interactions and their function in disease processes. Reactome pathways currently cover over half of the translated portion of the genome, and are available through our web site for browsing, downloading, and manipulation by in-house and third-party analysis tools.

The ReactomeFIViz Cytoscape app provides several analytical tools built upon the Reactome Functional Interaction network. A recent extension of the Reactome FI network, with updated protein-protein interactions and other pair-wise protein or gene relationships, has increased the FI network to covering just over 13,000 human proteins and over 431,000 interactions. We have expanded our current Probabilistic Graphical Models (PGM) for Reactome pathways to support an inference of pathway impacts of somatic mutations affecting protein structures in tumours. Moreover, we have established an approach to convert Reactome pathways into Boolean networks (BN) and have adapted a fuzzy logic model to perform pathway simulation to infer drug and somatic mutation impact. Both PGM- and BN-based pathway modeling approaches have been incorporated into the ReactomeFIViz app. to help users to perform PGM- and BN-based pathway impact analysis.

## **PgmNr 1018: iCOALAA-lite: An R-Shiny application for interactive correlation and liquid association analysis of multi-omics pan-cancer data.**

### **Authors:**

H. Ho <sup>1</sup>; K. Li <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, Taiwan; 2) Department of Statistics, University of California, Los Angeles, USA

---

In recent years, pan-cancer analysis has demonstrated its powerful use in identifying the commonality and dissimilarities among the alternation in the key biological processes across tumor types. The Cancer Genome Atlas (TCGA) and Pan-Cancer Atlas have generated a rich resource of multi-omics data processed by the standardized procedure with batch corrections. Due to the substantial case-to-case heterogeneity, some associations between molecular features and clinical features are expected to be preserved only in a subgroup of cases. An interactive analysis and visualization tool guided by real-time/precomputed statistics, and biological knowledgebases will help investigators better explore and interpret the complex association in massive multi-omics pan-cancer databases. We have developed the R-shiny application iCOALAA-lite to serve this purpose. This application allows users to select subgroups of cases of their interest based on the clinical and molecular features for the subsequent analysis. Precomputed statistics within each tumor type are provided for the preliminary screening of molecular features and the signatures derived from Pan-Cancer Atlas. Liquid association, a general statistical notion depicting the change in the covariation of two variables as a third variable varies, together with the conventional correlation will be computed in real-time for the second-stage feature selection. It allows users to perform modern visualization methods to visualize the case-to-case and feature-to-feature similarities in an interactive manner. This R-Shiny application and the precomputed statistics will be hosted as a web-based application. The R-Shiny application will also be released individually as a package for users to analysis their in-house data locally.

## PgmNr 1019: Cross-gene interactions of somatic mutations in transcriptomic system of cancer pathways.

**Authors:**

S. Jang; B. Han

[View Session](#) [Add to Schedule](#)**Affiliation:** Department of Biomedical Sciences, Seoul National University, Korea

---

Genetic mutations are known to affect gene expressions of downstream genes in the cancer pathways. However, the comprehensive view of regulatory systems of cancers is still yet to be uncovered. Recently, there were efforts to interpret the interplays between the cancer somatic mutations and the expression levels of genes in cancer pathways. These studies showed that somatic mutations can affect the gene expressions of the gene itself (cis) and other genes (trans), constructing a graph of regulatory effects which describe genomic effects on the transcriptome. Here, we found there exist strong cross-gene interactions in the transcriptomic system of cancer pathways. Specifically, we found that, if we consider two genes connected in the pathway (signal-sending gene and signal-receiving gene), there were interplays between the trans-effects of somatic mutations of signal-sending gene and the cis mutations of the signal-receiving genes. We analyzed key signaling pathways of cancer using the TCGA data to show the existence of cross-gene interactions. Our observations suggested that the regulatory systems of cancers can hardly be explained by the standard additive model.

# PgmNr 1020: An integrative pathway activity inference method using directed random walks for predicting survival in bladder cancer.

## Authors:

S.Y. Kim <sup>1,2</sup>; E.K. Choe <sup>1</sup>; M. Shivakumar <sup>1</sup>; D. Kim <sup>1,3</sup>; K.A. Sohn <sup>2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Department of Software and Computer Engineering, Ajou University, Suwon 16499, South Korea; 3) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA, USA

---

Urologic cancers include prostate, kidney, testicular, and bladder cancer with common genetic architecture in different types. To better understand the molecular features of urologic cancers, a comprehensive analysis using multi-omics data has been conducted. Additionally, a pathway activity inference method has been developed to facilitate the integrative effects of multiple genes. However, one of the limitations of the existing pathway activity inference approaches is to target a single genomic profile alone. In this respect, we have recently proposed a novel integrative analysis approach using directed random walk-based pathway activity inference method and demonstrated that it not only contributed to a higher survival group classification performance and also successfully reflected the combined effect of genes. In this study, we integrated multi-omics data to infer pathway activities for predicting survival outcome in bladder cancer. To reflect the interaction effects of genes, we designed a directed gene-gene graph using pathway information by assigning interactions between genes in multiple layers of networks. The proposed method selects cooperative driver pathways and predicts survival outcome using Lasso-Cox model. As a proof-of-concept study, multi-omics datasets, including RNA-Seq, DNA methylation, and copy number data, for bladder cancer were obtained from the Cancer Genome Atlas (TCGA) data. In the experiments, the proposed integrative method on three genomic profiles was evaluated with respect to the survival outcome prediction performance and the prioritized genes and pathways for survival of bladder cancer patients. In the results, the proposed integrative approach achieved average 7.7% improvements (C-index) of survival prediction performance than that of a single genomic profile. The integrative approach also identified 13 pathways as predictive prognostic features in bladder cancer, which are not found from a single genomic profile. They include 4 amino acid metabolism related pathways, 3 DNA repair pathways, and 6 cell proliferation, and growth related pathways. Our results showed that the integrative approach guided by pathway information not only improves survival outcome prediction performance, but also provides better biological insights into the pathways and genes prioritized by the model in an integrated view. This study warrants further exploration in other urologic cancer types to identify common/cancer-specific driver pathways.

## PgmNr 1021: Enabling big data analysis through cloud-based tools in the Cancer Research Data Commons.

**Authors:**

E. Kim; T. Davidsen; J. Klemm; A. Dearry; T. Kerlavage

[View Session](#) [Add to Schedule](#)

**Affiliation:** Cancer Informatics Branch, CBIIT/NCI/NIH, Rockville, Maryland.

---

Basic and clinical research in human disease is increasingly being focused on generation of rich datasets to identify molecular underpinnings of disease. NIH and NCI have supported numerous programs including The Cancer Genome Atlas (TCGA) and Clinical Proteomic Tumor Atlas Consortium (CPTAC) to generate a wealth of data to be leveraged by the research and clinical communities to better understand diseases and ultimately inform development of better treatments and prevention tools. However, we are still limited in our ability to draw insights and meaningful interpretations from these rich datasets, which include multi-omics, imaging, and clinical data, by challenges in integration across disparate datasets.

To progress towards this goal, the research community will need to access, integrate, and analyze multi-modal data, including genomics, proteomics, imaging data, clinical treatment/outcomes, and population-based data. Investment in the informatics infrastructure to fully leverage these diverse data types is imperative and was called out as a priority by the Cancer Moonshot<sup>SM</sup> Blue Ribbon Panel. To this end, NCI has initiated development of a Cancer Research Data Commons (CRDC) to provide access to interoperable data repositories, analysis tools, and workspaces. The vision for the CRDC is a virtual, expandable infrastructure that provides secure access to diverse data types, allowing users to analyze, share, and store results, leveraging the storage and elastic compute of the cloud. The CRDC is built upon Data Commons Framework (DCF) which provides a core set of modular, reusable components to rapidly develop data nodes for the CRDC. Currently the CRDC hosts the Genomics Data Commons (GDC) and the Proteomics Data Commons (PDC), while additional nodes (e.g., Imaging Data Commons, Integrated Canine Data Commons) are in development. In addition, the CRDC offers three NCI Cloud Resources that provide innovative methods to query, visualize, and analyze data by making best practice tools and pipelines readily available. These cloud-based systems eliminate the need for researchers to download and store large-scale datasets by allowing them to bring bioinformatics tools to the data either by integrating with CRs or by connecting via DCF. They also offer the computational capacity to enable big data analysis. The NCI CRDC nodes, DCF, and Cloud Resources are currently available to the scientific community to accelerate research and promote new discoveries.

## PgmNr 1022: An investigation of germline variants as the determinants of somatic mutations' patterns on cancers.

### Authors:

S. Lee <sup>1</sup>; C. Lee <sup>2</sup>; S. Song <sup>2</sup>; Y. Koh <sup>2,3</sup>; H. Yun <sup>2</sup>; S.-S. Yoon <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Medicine, Seoul National University Hospital, Seoul, Republic of Korea; 2) Cancer Research Institute, Seoul National University College of Medicine, Seoul, Republic of Korea; 3) Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea

---

With advances of sequencing technologies, in-depth studies of the somatic mutations for various cancers have been realized and identified some germline mutations that are related to the somatic mutations. However, a detailed study on such relations needs to overcome their sparsity and insufficient effect.

Here, we conducted an extensive study of the large-scale cancer dataset with both somatic and germline mutations, using two concepts to overcome the limitations: the mutational signature and novel pathway-based method that integrates biological information and statistical model. Using the mutational signatures from the somatic mutations and their corresponding germline mutations from 2,588 whole genomes, we conducted association studies between those mutational signatures and germline mutations.

Based on cancer-wise and signature-wise approaches, our association studies identified statistically significant signals between five genes and multiple signatures. Literature searches support their relationship of the corresponding cancer types.

Additional studies on those genes identified strong functional relationship on epigenetic regulation, which suggests the potential evidence on the mutational signatures with no proposed etiology.

In summary, we successfully conducted our preliminary study by identifying associations between germline and somatic mutations, with the novel methods that can overcome the current limitations.



## PgmNr 1023: QC measurements of exome chip sequence data in a family-based study.

### Authors:

Q. Li <sup>1</sup>; S. Wank <sup>2</sup>; J. Bailey-Wilson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, Maryland, United States of America; 2) Digestive Diseases Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland, United States of America

---

Quality control (QC) is an important step in sequence data analysis. In VCF files, many genotype calling quality measurements are reported at the variant level and the variant-and-sample level. Common practice is to drop variants based on pre-set thresholds of several key QC measurements, including QD, MQ, FS, SOR, MQRankSum, and ReadPosRankSum at the variant level. In addition, genotypes at specific loci should be set to missing if the variant-and-sample level measurements are poor. Pedigree information can be used as the third step of data cleaning, eliminating genotypes causing Mendelian inconsistencies.

In a linkage analysis of small intestinal carcinoid tumors, we performed QC on exome sequence data from a pedigree of 34 individuals. In this work, we reported the summary statistics on the genotype calling QC measurements of 219,742 variants. We found that the QC measurements have a wide range of variation across different chromosomes. Using hard-filtering based on recommended thresholds, 182,132 (~83%) variants were kept. We found that the current QC thresholds cannot remove all the poorly typed genotypes. Among the retained variants, 21% incurred at least one Mendelian error. Given the 18 informative child-parent trios in the pedigree and the 28,114 variants, this is a rate of ~1.9% per trio, per variant. Extensive and iterative cleaning of Mendelian errors are needed after data filtering by QC measures alone. In conclusion, the current hard filtering thresholds are inadequate and are improved by Mendelian inconsistency checks.

## PgmNr 1024: *Ab initio* spillover compensation in CyTOF data.

### Authors:

Q. Miao <sup>1,2</sup>; F. Wang <sup>1</sup>; R. Basar <sup>3</sup>; M. Muftuoglu <sup>3</sup>; L. Li <sup>3</sup>; K. Rezvani <sup>3</sup>; K. Chen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Bioinformatics and Computational Biology, Division of Quantitative Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 3) Department of Stem Cell Transplantation, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX

---

The mass cytometry (CyTOF) technology, which employs metal-isotope-tagged monoclonal antibodies to measure proteomic marker expressions in single cells, has been widely used to characterize cellular heterogeneity. Due to technological limitation, the intensity measured in a channel can be affected by neighboring channels. Although generally minor, such spillover effects can substantially limit the accuracy of clustering and lineage tracing. The CATALYST approach (PMID:29605184) can reduce spillover, but it requires the use of additional control beads, which increases the cost.

Here, we present a novel computational method that autonomously compensates the spillover effects in a CyTOF dataset without using any control beads. Our method utilizes knowledge-guided modeling and statistical techniques such as finite mixture modeling and sequential quadratic programming to infer the optimal compensation matrix and perform the correction.

We evaluated our method using four CyTOF datasets obtained from C57BL/6J mouse bone marrow, healthy human bone marrow, human peripheral blood, and chronic lymphocytic leukemia patient samples, respectively. In two datasets, our method resulted in improved clustering consistency with manual gating results. In the other two datasets, our method led to the discovery of novel subpopulations expressing functionally meaningful, cluster-specific markers.

The tool we developed in R will greatly facilitate large-scale cellular profiling of tumor immune microenvironment and development of novel immunotherapy.

# PgmNr 1025: A comparison of eVai, CADD and VVP variant prediction results on the ICR639 hereditary cancer dataset.

**Authors:**

G. Nicora <sup>1</sup>; I. Limongelli <sup>2</sup>; S. Zucca <sup>2</sup>; R. Santolisier <sup>1</sup>; P. Magni <sup>1</sup>; R. Bellazzi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) University of Pavia, Via Ferrata 5, 27100 Pavia, Italy; 2) enGenome S.r.l., Via Ferrata 5, 27100 Pavia, Italy

---

**Background**

Genomic variant interpretation for the diagnosis of hereditary diseases is a complex process and several bioinformatics software to support geneticists have been developed in the last few years. Strategies generally include variant filtering algorithms based on information from reference populations and other omics databases, computation of a score of deleteriousness assigned to each variant, or a combination of them.

The Expert VARIant Interpreter (eVai.engenome.com, PMID: 30298955) classifies variants based on ACMG/AMP guidelines and assigns each variant a score for pathogenicity prioritization. Performances in variant prioritization have been compared to two currently available hypothesis-free tools: CADD (PMID: 30371827) and VVP (PMID:29463208). Benchmark analysis was performed exploiting ICR639 validation set of cancer predisposition (PMID:30175241).

**Methods**

ICR639 hereditary cancer dataset was downloaded from European Genome-phenome Archive. It includes 639 VCF files from individuals tested for cancer predisposition genes (Illumina TruSight Cancer panel); 645 pathogenic variants in total (339 indels, including 16 complex indels and 24 indels longer than 5 bps), 232 SNVs and 74 exon copy number variants (excluded from this study). VCF files of each samples were analyzed with eVai, CADD and VVP tools as suggested by the guidelines and manuals. For each tool, a threshold was established according to specific tools guidelines to consider a variant as pathogenic (eVai score >3, CADD score >20, VVP score > 57). Commonly classification metrics were then applied for each tool to test classification performances with respect to known pathogenic variants such as accuracy (ACC), F1, ROC and PRC. Moreover, the Normalized Discounted Cumulative Gain (NDCG) ratio was used to identify the most performing tool in prioritizing pathogenic variants in the top 20 variant list: the more this score tends to one, the better the prioritization is among the first 20 positions.

**Results**

eVai resulted to be the best performing tool, with classification metrics higher than CADD and VVP. Classification accuracy and F1 score were higher for eVai (ACC=0.998, F1=0.416) respect to VAAST (ACC=0.988, F1=0.065) and VVP (ACC=0.860, F1=0.006). ROC and PRC AUCs were higher for eVai (ROC-AUC=0.999, PRC-AUC=0.803) compared with CADD (ROC-AUC=0.996, PRC-AUC=0.266) and VVP (ROC-AUC=0.985, PRC-AUC=0.334). Finally, NDCG-20 value was 0.956 for eVai, 0.619 for CADD and 0.436 for VVP.

## PgmNr 1026: Allelic-imbalance in DNA inferred from RNA-seq data using a haplotype-based method.

### Authors:

Z. Ozcan <sup>1,2</sup>; F.A. San Lucas <sup>1</sup>; J. Fowler <sup>1</sup>; J. Gu <sup>1,2</sup>; Y. Jakubek <sup>1</sup>; P. Scheet <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

- 1) Department of Epidemiology, The University of Texas M.D. Anderson Cancer Center, Houston, TX;
- 2) Graduate School of Biomedical Sciences, The University of Texas Health Science Center, Houston, TX

---

There is a growing interest in using RNA-seq data to identify somatic DNA mutations, particularly for scenarios when analyses of DNA are not feasible. Recently, this approach was used to characterize somatic mutations in The Genotype-Tissue Expression (GTEx) data for a comprehensive view of mosaicism across human tissues, primarily focusing on point mutations. Mutation detection in such settings is challenging due to the non-random distribution of expressed genes across tissues, and differential expression rates resulting in uneven coverage. In this work we sought to overcome some of these challenges for inference of acquired chromosomal alterations (allelic imbalance; AI) from RNA-seq by using a haplotype-based approach.

We obtained SNP-array, whole-exome sequencing (WES), and RNA-seq data from 279 lung adenocarcinoma (LUAD) participants in The Cancer Genome Atlas (TCGA). For AI detection, we applied hapLOHseq to WES and RNA-seq data, and hapLOH to SNP arrays. We considered AI events detected in the SNP array data as a gold standard, for the purposes of scoring our approach. Overall, our method tailored to RNA-seq data performed nearly as well as inference of AI from WES. Evaluating concordance at each SNP marker, we achieve sensitivity estimates of 86% and 83% in WES and RNA-seq, respectively, and corresponding specificity estimates of 87% and 77%.

AI calls in the RNA set that are not present in the gold standard set ("RNA-only events") could be false positive calls in the RNA, false negative calls in the gold standard set, or could represent AI patterns exclusive to RNA. To more confidently identify RNA-only AI candidates, we explicitly tested these specific regions for AI in the WES and SNP array data using a binomial test and excluding any segments with p-values < .10. After this filter, RNA-only events comprised 8% of the RNA-seq derived calls. We observed an apparent enrichment of these, in absolute counts and by application of GRIN ( $q < 1e-9$ ), on chromosomes 2, 7, and 10.

Our results suggest that detection of AI in RNA-seq may constitute a useful surrogate for detection of DNA-level AI in resource-limited settings. Currently, we are benchmarking AI detection in WES and RNA-seq in additional cancer sites from the TCGA, and further investigating technical and biological factors underlying the inferred RNA-only events, as these may offer insights into mechanisms of long-range epigenetic regulation.

## PgmNr 1027: Development of a new microsatellite instability defining tool using whole exome sequencing data.

### Authors:

H. Park <sup>1,2,3</sup>; S. Jung <sup>3</sup>; Y. Chung <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedicine & Health Sciences, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 2) Department of Microbiology, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 3) Integrated Research Center for Genome Polymorphism, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

---

**Background:** Microsatellite instability (MSI) is an important marker estimating the prognosis and determining the direction of therapy in gastrointestinal cancer patients. Polymerase chain reaction (PCR) is currently a gold standard to define MSI, but it takes time and has a possibility of misreading. We intended to develop a new tool to define MSI using next-generation sequencing (NGS) data.

**Methods:** To develop the MSI defining tool, we used whole exome sequencing data of 517 colorectal cancers in The Cancer Genome Atlas (TCGA) with MSI information as a training set. Through the analysis of diverse features such as single variant and copy number variants called from the colorectal cancer data set, we developed a NGS-based MSI defining tool based on the decision tree, one of the machine learning methods.

**Results:** When we applied our system to the 517 TCGA colorectal cancers (94 MSI and 423 MSS), the performances of defining MSI are as follows: sensitivity: 0.947, specificity: 0.991, accuracy: 0.983. We validated this tool with 23 colorectal cancer whole exome sequencing data generated in our lab, and their MSI status was confirmed by pentaplex PCR (8 MSI and 15 MSS). As a result, our tool discriminated the 8 MSI and 15 MSS cases exactly as the pentaplex PCR results.

**Conclusions:** Our NGS-based MSI defining tool is possible to execute by a simple command in the Linux OS. This tool showed good performance to discriminate MSI and MSS in both TCGA and our colorectal cancer data. Considering that NGS analysis became an ordinary test in the clinical field for gastrointestinal cancer diagnosis, this system would be a useful tool for defining MSI status without PCR analysis.

## PgmNr 1028: Detecting and refining low frequency somatic variants from gene panel-based next-generation sequencing.

### Authors:

S. Tian <sup>1</sup>; H. Yan <sup>1</sup>; E. Braggio <sup>2</sup>; M. Bockol <sup>3</sup>; Y. Liu <sup>1</sup>; S. Baheti <sup>1</sup>; Y. Asmann <sup>4</sup>; J.P. Kocher <sup>1</sup>; S. Slager <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905; 2) Division of Hematology/Oncology, Department of Medicine, Mayo Clinic, Phoenix, AZ 85054; 3) Division of Research and Education Support Systems, Department of Information Technology, Mayo Clinic, Rochester, MN 55905; 4) Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Jacksonville, FL 32224

---

Targeted cancer gene panel sequencing is the most prevalent NGS implementation in clinic. Despite its significance, major challenges exist in accurate detection of somatic mutations, especially of those with low allele frequency (LAF). As LAF variants often occur in a small proportion of the measured cells, the allele frequency can't be theatrically modelled. Also, cancer gene panel sequencing usually lacks matched normal in practice, which increases the error rate in defining variants as somatic origin. To improve the detection accuracy, we systematically evaluated four open source somatic SNV callers in both pair- and tumor-only mode and another two in either mode; multiple filtering strategies were further assessed to refine variants. We first computationally assembled benchmarking gene panels from AML31 sequencing data (dbGAP phs00159) with 268 cancer genes, including a primary tumor, its subsequent relapse, and a matched normal sample. The gold standard mutations (PMID: 26645048) were used as the reference set, including 25 somatic mutations in primary tumor (VAF: 0.08-49.5%) and 18 in relapse tumor (VAF: 0.16-22.4%). Using pair-mode in primary tumors, Varscan identified ten (MAF  $\geq$  0.39%), vardict identified seven (MAF  $\geq$  1.49%), and Lofreq, mutect2 and bassovar each identified six (MAF  $\geq$  11.8%) mutations, which all dropped to six after applying the common filters based on variant quality and MAF in public germline databases (1000G, ExAC and ESP). In relapse tumor that has lower sequencing depth, Varscan identified ten (MAF  $\geq$  0.24%), vardict identified five (MAF  $\geq$  2.84%), and the other three each identified 3-4 (MAF  $\geq$  6.8%) mutations, which were reduced to four for Varscan, verdict and lofreq (MAF  $\geq$  6.8%) and 2-3 for other two (MAF  $\geq$  14.3%) after filtering. Using single-mode in primary tumors, where bassovar was replaced with GATK UnifiedGenotyper, all five callers identified six mutations (VAF: 11.8-50.43%) and missed LAF variants. In relapse tumors, the five callers identified 4-5 mutations; mutect2 missed two and verdict missed one variant after the filtering. However, mutect2 reported less total variants than other callers. We further checked the possible source of false positives and found that over 60% and 20% of them were located in genes with certain characteristics of domains or having homologous events, respectively. Our analysis argues for the need to develop a more optimal strategy to reduce false calls, such as machine learning approach.

## PgmNr 1029: BioMethyl: An R package for biological interpretation of DNA methylation data.

### Authors:

Y. Wang<sup>1</sup>; J. Franks<sup>1,2</sup>; M. Whitfield<sup>1,2</sup>; C. Cheng<sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, New Hampshire; 2) Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Lebanon, New Hampshire; 3) Norris Cotton Cancer Center, Lebanon, New Hampshire; 4) Department of Medicine, Baylor College of Medicine, Houston, TX

---

**Background:** The accumulation of publicly available DNA methylation data sets has resulted in the need for tools to interpret the specific cellular phenotypes in bulk tissue data. Current approaches use either single differentially methylated CpG sites or differentially methylated regions that map to genes. However, these approaches may introduce biases in downstream analyses of biological interpretation, because of the variability in gene length. There is a lack of approaches to interpret DNA methylation effectively. Therefore, we have developed computational models to provide biological interpretation of DNA methylation (BioMethyl) in the context of The Cancer Genome Atlas (TCGA).

**Method:** We downloaded RNA-seq and DNA methylation data for all 37 TCGA cancer types from Firehose (<https://gdac.broadinstitute.org/>, Nov, 2016). We fit models to the cancer types that contained more than 50 samples with both RNA-seq and DNA methylation profiles. We trained each BioMethyl model using linear regression to capture the association between gene expression and DNA methylation for each cancer types. To automatically identify enriched pathways, GSEA R script method “GSEA.1.0.R” was deployed in BioMethyl package to perform enrichment analysis with default settings

**Results:** Using breast cancer as an example, BioMethyl shows high consistency in the identification of enriched biological pathways from DNA methylation data compared to the results calculated from RNA sequencing data. We find that 12 out of 14 pathways identified by BioMethyl are shared with those by using RNA-seq data, with a Jaccard score 0.8 for estrogen receptor (ER) positive samples. For ER negative samples, three pathways are shared in the two enrichments with a slight lower similarity (Jaccard score=0.6). Using BioMethyl, we can successfully identify those hidden biological pathways in DNA methylation data when gene expression profile is lacking. We also shown that BioMethyl is able to apply to non-cancer diseases.

**Conclusion:** We illustrate that BioMethyl utilizes the complete DNA methylation data for a given cancer type to reflect corresponding gene expression profiles and performs pathway enrichment analyses, providing unique biological insight. BioMethyl R package is freely available in the GitHub repository (<https://github.com/yuewangpanda/BioMethyl>).



## PgmNr 1030: Transcriptomic profiling and deconvolution for biomarker identification and immune cell infiltration estimation in colon mucosa of Lynch syndrome patients from a phase Ib chemo-preventive trial of naproxen.

### Authors:

W. Wu<sup>1</sup>; L. Reyes<sup>1</sup>; K. Chang<sup>1</sup>; E.M. Stoffel<sup>2</sup>; R. Lim<sup>3</sup>; P. Kanth<sup>4</sup>; P.M. Lynch<sup>1</sup>; N.J. Samadder<sup>5</sup>; G.L. Milne<sup>6</sup>; L.J. Marnett<sup>6</sup>; V. Sepeda<sup>1</sup>; Y.N. You<sup>1</sup>; L.A. Vornik<sup>1</sup>; J.J. Lee<sup>1</sup>; E. Richmond<sup>8</sup>; M. Perloff<sup>8</sup>; P.H. Brown<sup>1</sup>; S.M. Lipkin<sup>7</sup>; E. Vilar<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Departments of Clinical Cancer Prevention, Epidemiology, Gastroenterology, Hepatology & Nutrition, Surgical Oncology, Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX.; 2) Division of Gastroenterology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI.; 3) Department of Gastroenterology, Dana-Farber/Brigham and Women's Cancer Center, Boston, MA.; 4) Division of Gastroenterology, Cancer Control and Population Sciences, Huntsman Cancer Institute, Salt Lake City, UT.; 5) Department of Gastroenterology and Hepatology, Mayo Clinic, Phoenix, AZ.; 6) Departments of Biochemistry, Chemistry, Medicine, and Pharmacology, Vanderbilt University School of Medicine, Nashville, TN.; 7) Weill Cornell Medical College, New York, NJ.; 8) Division of Cancer Prevention, National Cancer Institute, Bethesda, MD.

---

**Background** Lynch Syndrome (LS) is one of the most prevalent cancer-prone hereditary conditions. It is characterized by predisposition to a wide spectrum of cancers, primarily colorectal cancer (CRC) and endometrial cancer (EC). Patients with LS are more likely to be diagnosed with cancer in a young age and have up to 70-80% risk of developing CRC during their life time. Naproxen is a non-steroidal anti-inflammatory drugs (NSAID) that is widely used for treatment of pain and has an excellent safety profile. Preclinical study using a genetically engineered LS mouse model (VC-MSH2-LoxP) has shown that Naproxen is the most effective agent among a selected panel of NSAIDs in preventing colorectal tumors. The present phase Ib clinical trial was designed to assess the safety profile and tolerability of long-term chemo-preventive usage of Naproxen, as well as to identify noncanonical molecular biomarkers and their associated possible functions via transcriptomic profiling and immune deconvolution. **Methods** Patients were randomly assigned to Naproxen 220 mg qd (LD), 440 mg qd (HD) or placebo (PL) once daily administered for 6 months. Biopsy of colon mucosa was obtained at the time of enrollment (pretreatment) and after 6 months of Naproxen administration (posttreatment). RNA was extracted and sequenced at Sequencing and Microarray Facility (SMF) at MD Anderson. Sequencing reads was aligned to UCSC hg19 reference genome using spliced aligner STAR. Whole transcriptome was quantified using *rsem-calculate-expression*. Differential expression, pathway analysis and immune deconvolution were performed using R. **Results** From 80 enrolled patients, 23/28 in placebo group, 18/27 in LD Naproxen group, and 17/25 in HD Naproxen group completed the study, and their RNASeq data were used for bioinformatics analyses. 0, 70 and 120 differentially expressed genes were found in PL, LD and HD posttreatment vs pretreatment at FDR=0.1 respectively. Pathway analysis revealed significant enrichment of immune/chemokine related pathways in both LD and HD Naproxen groups. Differential infiltration fractions of various immune cell types were found in LD and HD posttreatment groups. **Conclusion** Long-term



Chemoprevention with Naproxen for a total of 6 months was generally well-tolerated. Naproxen may exert chemo-preventive effect by modulating infiltration of certain immune cell types into colon mucosa.

## PgmNr 1031: gcMECM: Graph-based clustering of mutual exclusivity of cancer mutations.

**Authors:**

C. Yan; Y. Hu; Q. Chen; D. Meerzaman

[View Session](#) [Add to Schedule](#)

**Affiliation:** CBIIT, NCI, Rockville, Maryland.

---

Advancement of next-generation sequencing technology has transformed the study of cancer genome. Whole-genome or whole-exome measurements of somatic mutations in large numbers of cancer genomes are now a reality. With the comprehensive mutation landscape, tremendous progress has been made to identify cancer driver genes through detection of highly recurrent mutations called “significantly mutated genes”, which are limited to the highly mutated genes. To provide therapeutic options for most patients, it will, therefore, be critical to identify and understand the pathway-level implications of all genes mutated at low frequencies. The reason is that cancer genes tend to be altered in a limited number of pathways related to cell division, differentiation, survival, and genome maintenance. Here we present a novel tool that combines Louvain Method for community detection, mutation association, and gene interaction to identify the mutually exclusively mutated gene sub-networks. The sub-networks reveal crucial genes in the canonical pathway and discover new cancer-relevant genes, which are used to build better prediction models of response and survival. We analyzed mutation data in thirty-three cancer types from The Cancer Genome Atlas (TCGA) and identified sub-networks common in multiple cancer types with improved survival prediction comparing to the single gene method. gcMECM also provides informative visualization functionality of mutual exclusivity and networks.

## PgmNr 1032: Reducing false positive somatic variant calls arising from breakpoint-associated alignment artifacts.

### Authors:

Q. Zeng; Y. Cao; L. Iyer; N. Russell; A. Kenyon; D. Boles; S. Letovsky

[View Session](#) [Add to Schedule](#)

**Affiliation:** LabCorp, 3400 Computer Drive, Westborough, MA 01581

---

Somatic variant calling in cancer samples is routinely done using short read next generation sequencing (NGS) technologies. False positive somatic variant calls from tumor samples can arise from a wide range of artifacts in NGS reads and read alignments. These include spurious calls from homopolymer regions, low-complexity regions, microsatellite, regions with pseudogene counterparts, reads with extreme strand bias. Of particular interest in tumor sample analysis are the false positive caused by breakpoint-associated short read alignment artifacts when a duplicated sequence is inserted or deleted with respect to the reference genome. In these cases, if the end of a read does not have enough flanking sequence from the neighboring non-duplicated region, the read ends can be incorrectly aligned to the reference genome, leading to spurious somatic variant calls.

In tumor sample analysis, FMS-like tyrosine kinase 3 (FLT3, a proto-oncogene) internal tandem duplications (ITDs) are detected in a quarter of the patients with acute myeloid leukemia (AML). FLT3 ITDs are strongly linked to bad prognosis, higher risk of relapse and lower survival rate. In samples with FLT3 ITDs, the reads ending with a short sequence stretch in the ITDs often aligns with the nearby genomic region, leading to false positive calls. False positives can also arise for several other tumor-associated genes with duplicated segments (e.g., CEBPA, CALR), where reads with the deletion of one of the duplicated segments often align incorrectly. A computational protocol has been developed to address this issue, resulting in significantly reduced the false positive calls due to breakpoint-associated artifacts in NGS-based somatic variant calls.

This approach performs multiple iterations of read realignment by the ABRA tool in regions with INDELs, DUPs and structural variants ((e.g., for FLT3 ITDs), where imperfect read alignment is more likely to happen. Use of the re-aligned BAM files not only reduces the false positive calls due to misalignment near the ends of the reads, but also improves sensitivity for detection of true somatic mutations since more reads harboring the duplicated region are correctly aligned. The protocol for tumor sample analysis has been implemented and it has eliminated almost all false positives due to breakpoint-associated false positive calls.

## PgmNr 1033: Construction of circRNA-miRNA-mRNA networks in the pathology of gastric cancer based on integrated bioinformatics strategy.

### Authors:

M. Ahmadi <sup>1</sup>; F. Haji Esmaeili <sup>1</sup>; S.M. Kalantar <sup>2</sup>; R. Fotuhi Ardakani <sup>3</sup>; S. Houshmand <sup>4</sup>; S. Nouri <sup>4</sup>; S.A. Momeni <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Booali Office of USERN, Universal Scientific Education and Research Network (USERN), Qom, Iran; 2) Abortion Research Centre, Yazd Reproductive Sciences Institute, Shahid Sadoughi University of Medical Science, Yazd, Iran.; 3) Clinical Laboratory Science Department, Qom University of Medical Sciences; 4) Research and development division, Booali medical diagnostic laboratory, Qom, Iran

---

Gastric cancer (GC) is a prevalence malignant neoplasm of the digestive system. The aim of the present study was to identify key genes, micro RNAs (miRNA) and circular RNAs (circRNAs) in GC and understanding the potential molecular mechanism of GC by constructing a circRNA-miRNA-mRNA network. With the threshold of the adjusted false discovery rate (FDR) P-value of  $<0.01$  and a log fold-change ( $\log FC$ )  $>1$ , 751, 481 and 3,340 mRNAs from GSE63089, GSE27342 and STAD (from TCGA), 111, 233 and 130 miRNAs from GSE63089, GSE27342 and STAD and with predefined criteria (a log fold-change ( $\log FC$ )  $>2$ ), 33 circRNAs from GSE78092 were differentially expressed in GC. By intersecting the 51 overlapping miRNA target genes and the 233 overlapping DE mRNA between GSE63089, GSE27342 and STAD and 12 target genes of miRNA sponged by DE circRNAs, we acquired 290 overlapped genes. Using the DAVID database (v. 6.8), it demonstrates that the 290 genes were mainly enriched in the extracellular space and proteinaceous extracellular matrix and were involved in the mitotic nuclear division, collagen catabolic process, and participated in the regulation of the Cell cycle pathway. Interactions between differentially expressed miRNAs, mRNAs and circRNA were predicted via miRWalk3.0 and CircInteractome. Co-expression of miRNAs, circRNA, and mRNAs was selected to construct circRNA-miRNA-mRNA interaction networks. A protein-protein interaction network based on the 290 genes was established, 12 hub genes (*CDK1*, *CCNA2*, *CCNB1*, *ESR1*, *BUB1*, *IGF1*, *KIF11*, *TOP2A*, *CHEK1*, *PLK1*, *AURKA*, *MKI67*, *KIF20A*, and *CDC6*) determined from the network. The results were visualized in Cytoscape software (v. 3.7.1). DGIdb database predicted 39 small molecules as the possible drugs for treating GC, the identified target network was visualized using STITCH (version 5.0). In summary, we identified key genes, pathways, and constructed circRNA-miRNA-mRNA networks in GC by integrated bioinformatics analysis. These data may provide new insights into GC pathogenesis and useful information for the utility of personalized therapy in the future.

## PgmNr 1034: Paired DNA and RNA sequencing illuminates the extent of allele-specific expression in multiple myeloma and identifies impacted pathways.

### Authors:

C. Ashby <sup>1,2</sup>; M.A. Bauer <sup>1,2</sup>; C.P. Wardell <sup>1</sup>; Y. Wang <sup>2</sup>; R.G. Tytarenko <sup>2</sup>; P. Patel <sup>2</sup>; S. Deshpande <sup>2</sup>; F. van Rhee <sup>2</sup>; M. Zangari <sup>2</sup>; S. Thanendrarajan <sup>2</sup>; C.D. Shinke <sup>2</sup>; F.E. Davies <sup>3</sup>; G.J. Morgan <sup>3</sup>; B.A. Walker <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Informatics, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Myeloma Center, Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, AR; 3) Myeloma Program, NYU Langone Medical Center, New York City, NY

In the human genome, allele-specific expression (ASE) is the differential expression of two alleles. Further, it has been shown that when ASE occurs in cancer it is often linked to driver genes. However, very little work has been performed in this area as they relate to certain specific cancer types. In multiple myeloma (MM), a hematological malignancy, ASE has not been fully investigated. In this work, we look at a large dataset of MM samples with paired DNA and RNA data to characterize the ASE present.

The dataset was composed of 589 newly diagnosed MM samples from the Multiple Myeloma Research Foundation CoMMpass study. Each sample had paired whole-exome sequencing and RNA-Seq expression data. Strelka2 was used to identify mutations. We used cisASE to identify somatic allele-specific expression at the gene level. Read positions were kept based on default cisASE recommendations. We identified a median of 67 (range 18-1937) genes containing at least one mutation per sample. Using a significance level of  $p < 0.01$  identified a median of 33 (range 4-633) genes with ASE per sample.

All major translocation subgroups in MM were identified as having some degree of ASE and were significantly associated (Fisher's exact;  $P < 0.01$ ) with the genes impacted by the translocation events in those samples. These include *FGFR3* in t(4;14) (100%,  $n=27/27$ ), *CCND1* in t(11;14) (80%,  $n=16/20$ ), *CCND3* in t(6;14) (66%,  $n=2/3$ ), *MAF* in t(14;16) (90%,  $n=9/10$ ) and *MAFB* in t(14;20) (100%,  $n=4/4$ ).

We observed 169 genes with overexpressed mutated alleles (significant and RNA variant frequency  $> 75\%$ ) that were recurrently present in two or more samples. These genes were significantly implicated in a wide variety of pathways including many cancer types and key pathways, such as the p53 and PI3K-Akt signaling pathway. Next, using published results of 63 MM driver genes, we identified overexpressed mutated alleles in 33 of them including: *CCND1* 80% ( $n=16/20$ ), *DIS3* 52% ( $n=15/29$ ), *FGFR3* 37% ( $n=10/27$ ), *NRAS* 18% ( $n=9/49$ ), *TP53* 56% ( $n=9/16$ ), *TRAF3* 64% ( $n=9/14$ ), *MAF* 80% ( $n=8/10$ ), *TENT5C* 24% ( $n=7/29$ ), *TGDS* 100% ( $n=7/7$ ), *MAX* 50% ( $n=5/10$ ), *MAFB* 100% ( $n=4/4$ ), *ATM* 27% ( $n=3/11$ ), *SAMHD1* 27% ( $n=3/11$ ), and *XBP1* 23% ( $n=3/13$ ).

Our results show that ASE is a frequent event in MM and has the potential to further increase our

understanding and identify novel therapeutic targets of this malignancy.

## PgmNr 1035: Transcriptomic deconvolution reveals cell-type specific signals across cancer types.

### Authors:

S. Cao <sup>1</sup>; R. Liu <sup>2</sup>; L. Yang <sup>3</sup>; J. Ahn <sup>4</sup>; J. Chen <sup>1</sup>; F. Gao <sup>1,2</sup>; Z. Wang <sup>1,2</sup>; F. Zhang <sup>1</sup>; E. Efstathiou <sup>6</sup>; D. Frigo <sup>6,7</sup>; S. Kopetz <sup>8</sup>; H. Zhu <sup>5</sup>; W. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Department of Statistics, Rice University, Houston, TX; 3) Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, NC; 4) Department of Biostatistics and Bioinformatics, Georgetown University, Washington, DC; 5) Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX; 6) Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX; 7) Department of Cancer Systems Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX; 8) Department of Gastrointestinal (GI) Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX

---

Most tumors consist of a variable proportion of malignant and nonmalignant cells including epithelial cells, fibroblasts, and infiltrating immune cells, which confounds biomarker studies of response to treatment. Deconvolution approaches have been developed for transcriptomes to address this heterogeneity in tumor samples.

We developed a new deconvolution framework, *DeMixT*, (Bioconductor R package) for the more accurate and efficient transcriptome deconvolution of high dimensional data from mixtures of more than two components. Besides estimating mixing proportions, *DeMixT* uniquely provides per-gene per-sample expression levels of each mixture component. To address multiple sources of variations observed in real data, we proposed a novel profile likelihood-based (PL) method to adaptively select a gene set, with a high signal-to-noise ratio, for the proportion estimation. This adaptive gene set selection substantially improves the estimation accuracy of both proportions and cell-type specific gene expression levels for the entire transcriptome.

We applied the PL-enabled *DeMixT* in two component modes (tumor and stroma) to 16 solid tumor types from The Cancer Genome Atlas (TCGA) database: BLCA, BRCA, COAD, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PAAD, PRAD, READ, STAD, THCA, UCEC. We compared results of pathway and survival analyses from the original mixed expression data with the deconvolved expression data. We also performed three-component deconvolution (tumor, stroma and immune) for 5 cancer types (BLCA, COAD, PRAD, HNSC, KIRC) without requiring data from the immune profiles. For these 16 cancer types, we obtained tumor purities, immune proportions (if available) and the deconvolved individual-level gene expression. The mean number of genes (per cancer type) recovered was 13,239 (sd=886), and the mean number of tumor deconvolved samples was 371 (sd=201). Downstream analyses were performed and biological findings from previous literature were validated using deconvolved expressions but not the original mixed expressions. Novel cell-type specific pathway activities were identified. Immune proportions were shown to be highly predictive and of good prognosis in multiple cancer types.

We provide comprehensive transcriptome deconvolution outputs for the TCGA datasets which will enable investigation of the tumor-stroma-immune microenvironment and elucidate new biological mechanisms for cancer.



## PgmNr 1036: Chromosomal allelic imbalances in lymph nodes of lung cancer patients.

### Authors:

K. Chang <sup>1</sup>; F. San Lucas <sup>2</sup>; J. Fujimoto <sup>3</sup>; B. Sanchez Espiridion <sup>3</sup>; C. McDowell <sup>1</sup>; X. Tang <sup>3</sup>; J. Stewart <sup>4</sup>; I. Wistuba <sup>3</sup>; H. Kadara <sup>3</sup>; G. Eapen <sup>5</sup>; P. Scheet <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX.; 2) Molecular Diagnostics Laboratory, The University of Texas MD Anderson Cancer Center, Houston, TX.; 3) Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX.; 4) Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX.; 5) Pulmonary Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX.

---

The lymphatic system is a component of cancer metastasis, serving as a gateway to organs distant to a primary malignancy. To attempt to understand genomic mechanisms associated with metastasis, we sought to study the DNA alterations present in lymph nodes (LNs) derived from patients with early stage lung cancer. Traditional challenges to such an endeavour include access to tissues and the low mutant cell fraction expected to be exhibited in such samples.

We obtained samples from 141 lymph nodes and 27 corresponding primary tumors from 34 non-small-cell lung cancer (NSCLC) patients from a retrospective collection of specimens at MD Anderson. The LN samples were either from surgical resection or needle aspiration during endobronchial ultrasound (EBUS), a non-surgical procedure supporting assessment of the extent of tumor involvement in LN samples. Samples were genotyped via DNA microarrays (Global Screening Array, Illumina). To overcome low aberrant cell content, we applied our haplotype-based algorithm to study somatic copy number alterations (sCNAs).

We detected 317 sCNAs in LNs with tumors cells detected by cytopathology review (LN+) and 81 alterations in LNs without tumor detected (LN-). Chromosomes 2 and 12 demonstrated the most sCNAs in LN+ samples, while chromosome 3 and 8 were most prevalent among LN-. 19 of the 42 LN+ samples had sCNAs that were shared with the corresponding primary tumors. However, only 12 of the 99 LN- samples had shared sCNAs. Additionally, we identified “mirrored” alterations in 7 patients by statistically testing whether the over-represented parental haplotypes matched between overlapping alterations in LN and tumors, indicating independent, convergent mutational events.

Our findings demonstrate a potential to discover chromosomal alterations in histopathologically confirmed LN- tissues; further, some of these sCNAs were found in the corresponding primary tumor. Further analyses of our data may help understand molecular alterations associated with, or critical for, metastasis. In the longer-term, our approach may offer the potential to advance or complement traditional forms of detection and diagnosis of NSCLC and LN metastasis.

## PgmNr 1037: Adaptive immune responses associated with the progression of premalignant lesions to colorectal cancer.

### Authors:

Z. Chen <sup>1</sup>; X. Wang <sup>1,2</sup>; W. Wen <sup>1</sup>; J. Ping <sup>1</sup>; X. Shu <sup>1</sup>; Q. Cai <sup>1</sup>; J. Long <sup>1</sup>; Y. Xia <sup>2</sup>; G. Yang <sup>1</sup>; W. Zheng <sup>1</sup>; X. Guo <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University Medical Center, Nashville TN, USA.; 2) Department of Dermatology, The Second Affiliated Hospital, School of Medicine, Xi'an Jiaotong University, Xi'an, China.

---

Colorectal cancer (CRC) progresses by the accumulation of molecular alterations, from normal epithelium to adenoma to cancer. It has been increasingly recognized that immune responses, measured by the abundance of tumor-infiltrating lymphocytes (TILs), may play a critical role in CRC etiology and prognosis and treatment response. However, immune responses in the progression of premalignant lesions to CRC is still unclear. Analyses of gene expression data can comprehensively characterize immune responses in tumor tissues. Here we systematically collected publicly available gene expression data from the Cancer Genome Atlas (TCGA) and previous literature, including the data from 118 normal colon mucosa, 243 adenoma and 2,608 cancer tissue samples. We performed bioinformatic analyses to normalize the gene expression data combining from all these tissue samples. We estimated the ImmuneScore (quantified by overall local immune activities, including B-cells, CD4+ cytotoxic T-cells, CD8+ cytotoxic T-cells and other TILs) and the abundance of each type of TIL using the xCell tool. Our results showed that the ImmuneScore consistently decreased along the continuum of carcinogenesis from the normal mucosa, to adenoma, to cancer. We also observed the same trend for B cells, CD8+ cytotoxic T-cells, CD4+ effector memory T-cells and overall T-cell densities. To explore the biological processes underlying these alterations of immune responses, we analyzed the differences of gene expression among normal colon mucosa, adenoma and cancer. Using a threshold fold change > 2 and a False Discovery Rate (FDR) < 0.05, we identified a total of 1,984 and 1,708 differentially expressed genes (DEGs) between normal mucosa and adenoma, and between adenoma and cancer, respectively. Of them, 469 genes were commonly observed in both comparisons. We further performed gene-set enrichment analysis for the above two sets of DEGs, respectively. Our results showed that immune-related signature IFN $\gamma$  response and MYC targets were significantly enriched in both gene sets, while Wnt/beta-catenin signaling and oxidative phosphorylation were specifically enriched in the former gene set, and KRAS signaling, inflammatory response and TNF- $\alpha$  signaling were significantly enriched in the latter gene set. These findings provide novel insight into the alterations of immune responses and the related biological mechanisms in the progression of premalignant lesions to CRC.

## PgmNr 1038: Modeling aneuploidy evolution in tumor single cells reveals genetic divergence and functional convergence.

### Authors:

K. Chen <sup>1</sup>; F. Wang <sup>1</sup>; Q. Wang <sup>1,2</sup>; J. Han <sup>1</sup>; S. Liang <sup>1,2</sup>; V. Mohanty <sup>1</sup>; R. Gao <sup>1</sup>; L. Ding <sup>3</sup>; N. Navin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The University of Texas M.D. Anderson Cancer Center, Houston, Texas.; 2) Rice University, Houston, TX; 3) Washington University School of Medicine, St. Louis, MO

---

Aneuploidy is one of the major mechanisms for tumor clonal evolution. It can lead to a burst of diverse genomic profiles, enabling rapid adaption and evolution. Although studies based primarily on multi-regional bulk-seq samples have yield valuable insights, how aneuploidy evolves amongst single cells remains elusive due partly to technological and analytical challenges. Particularly, single-cell analytical methods available to construct tumor cellular phylogenies cannot accurately represent the lineage, timing, diversity, scope and types of CNAs underlying the evolution.

We developed a novel single-cell evolutionary modeling method that can efficiently construct a rooted direct minimal aneuploidy event tree (MAET), from a population of thousands of tumor cells undergone single-cell DNA or RNA sequencing. Our algorithm dynamically infers the minimal set of CNA events required to evolve one genome to the next, providing an unbiased estimate on the CNA rate at various stages and lineages of the evolution. Evaluating our method using data simulated under various CNA rates, evolution dynamics and CNA mechanisms such as whole-genome doubling, breakage-fusion-bridge, etc. confirmed the superior applicability of our method over conventional phylogenetics methods on available single-cell data.

We examined longitudinal scDNA-seq data obtained from tens of chemoresistant triple-negative breast cancer patients and scRNA-seq data from a set of relapsed multiple myeloma patients. The MAETs constructed using our method indicated branching evolution in all the cases, with CNAs often accelerated in cells sampled from late stage samples and from lineages with alterations in DNA damage repair genes. Despite the increased genetic divergence, we found that CNAs tend to affect genes in the PI3K, TGFb, Nrfs, Hippo, and/or MYC pathways in an evolving tumor, indicating both convergent and competing evolution amongst the lineages.

Our study revealed structural, functional characteristics of aneuploidy evolution at single-cell resolution, implicated combinational therapeutics for treating aneuploid tumors, and provided a first-of-its-kind method that supports accurate evolutionary modeling in tumor single-cell data, which are becoming widely available.

## PgmNr 1039: Metagenomic bioinformatics tools for the study of microbial associations in human cancers and other human disease.

### Authors:

M. Dadlani <sup>1</sup>; N. Hasan <sup>1,2</sup>; R. Colwell <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) CosmosID, Rockville, MARYLAND.; 2) University of Maryland

Increasingly, relationships between microbiome composition, human disease, host immune response, and genomic characterization are being studied. Metagenomics and microbial functional genomics contributes to an understanding of how different microbiomes (e.g., oral, gut, and skin) contribute to or correlate with specific cancers or disease states. Current studies include the association between the human oral microbiome and pancreatic cancer risk, as well as changes in the microbiome observed in human cancer cases including oral squamous cell carcinoma (saliva), esophageal cancer (saliva, biopsied tissue), gallbladder cancer (bile culture), colorectal cancer (feces, biopsied tissue). In addition, many human cancers are known to be driven by microorganisms such as *Helicobacter pylori*, Human papillomavirus, Epstein-Barr virus, Human immunodeficiency virus, Hepatitis B, C viruses, Human T-cell lymphotropic virus type 1, and Human herpesvirus 8. A perpetuating bottleneck is the ability to perform the analysis of microbiome data in a seamless way without the need for large computational infrastructures. Bioinformatics tools are needed to provide comprehensive microbiome analysis that reports subspecies or strain level microbial identification, antimicrobial resistance and virulence information, and comparative microbiome analysis of samples among different geographic locations, cohorts, cases, or time series.

To address these needs, CosmosID has developed a cloud-based metagenomics bioinformatics platform based on microbial and gene databases with >150,000 reference sequences. CosmosID has also analyzed tens of thousands of biological samples. The databases have been cleaned of human and other DNA contamination commonly found in public repositories and human host DNA has no effect on microbial identification. Microbiome sequencing data can simply be uploaded to the CosmosID web application where CosmosID identifies all microorganisms (bacteria, viruses, fungi, and protists) and characterizes their attributes (e.g., antibiotic resistance, virulence) in a single, universal, rapid analysis in a secure platform with version control, periodic updates, frequent addition of new features, and the ability to do both whole genome metagenomics and amplicon analysis. Both metagenomic and amplicon analysis can also be performed with command line API access. In addition to the web application, CosmosID provides bioinformatics support for customized study needs.

## PgmNr 1040: Robust foreground detection in cancer genomic read depth data.

### Authors:

A. Deshpande<sup>1,4</sup>; T. Walradt<sup>2,4</sup>; Y. Hu<sup>4,6</sup>; A. Koren<sup>6</sup>; M. Imielinski<sup>2,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Tri Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Graduate School of Medical Sciences, New York; 2) Weill Cornell Medical College, New York; 3) Pathology and Laboratory Medicine, Weill Cornell Medicine, New York; 4) New York Genome Center, New York; 5) Institute for Computational Biomedicine, Weill Cornell Medicine, New York; 6) Department of Molecular Biology and Genetics, Cornell University, Ithaca

---

Somatic copy number alterations (SCNAs) can drive cancer and provide therapeutic, diagnostic, and prognostic value in precision oncology. The key challenge to sensitive and precise detection of SCNAs in cancer sequencing data is “waviness” and “noise” in read depth data, which has been previously attributed to technical and biological biases introduced during library preparation and sequencing. To overcome this challenge, we present dryclean, a signal processing algorithm to optimize SCNA detection in whole genome sequencing (WGS) and targeted sequencing platforms through the modeling and subtraction of a complex background. Application of dryclean to WGS demonstrates that the primary contributor to WGS background is replication timing and the intensity of this waviness correlates with biomarkers of proliferation. Through rigorous benchmarking utilizing numerous primary tumor samples, we show that the dryclean foreground signal improves the detection of biologically significant and clinically actionable SCNAs relative to state-of-the-art algorithms across multiple sequencing assays. Applying dryclean to *in silico* dilutions simulating cell-free tumor DNA detected with WGS, WES, and targeted panel (TP), we demonstrate more than 10-fold improvement in liquid biopsy residual disease detection for breast cancer and glioblastoma compared with existing methods. Dryclean’s ability to detect subtle, yet clinically actionable signals in genomic read depth data suggest this algorithm has the potential to become a vital component of the oncologist’s armamentarium.

## PgmNr 1041: *Oncogene.iobio*: A web app for real-time, integrative examination and functional prioritization of tumor mutations.

### Authors:

S.J. Georges; Y. Qiao; X. Huang; T. Di Sera; G.T. Marth

[View Session](#) [Add to Schedule](#)

**Affiliation:** USTAR Center for Genetic Discovery, Department of Human Genetics, University of Utah, Salt Lake City, UT

---

Genetic testing and clinical tumor sequencing are rapidly becoming part of the standard of care for a large subset of cancers. At present, a scattered framework of resources, often requiring expensive hardware and heavy computational skills, exists to aid clinicians and researchers in the bioinformatic analysis process. To assist in addressing the need for a comprehensive tool that operates intuitively, and at the pace of critical therapeutic timelines, we have developed *oncogene.iobio*. Utilizing our novel *iobio* analysis ecosystem (<https://iobio.io>), *oncogene.iobio* visualizes genomic, transcriptomic, and epigenetic data to aid in the discovery of variants with the highest potential for association with the patient's cancer.

*Oncogene.iobio* consumes primary sequencing data (in bam, vcf, fasta/fastq formats) sourced from local, or cloud-based repositories. In real-time, variants are identified and annotated on a gene-by-gene basis, and prioritized using functional annotations (sourced from SNPEFF, VEP, SIFT, PolyPhen) and cancer databases (COSMIC, cBioPortal). Uniquely, *oncogene.iobio* identifies compound heterozygotes between inherited cancer-implicated variants and somatically-acquired tumor mutations, dynamically filters mutations by allele frequency, and analyzes somatic variants within areas of chromosomal CNVs or LOH. Areas with large transcriptional change or allele-specific expression are highlighted when RNAseq data is present, along with hypo- or hyper-methylated sites given bisulfate-treated DNA sequencing reads. Ranked, candidate gene lists, cancer-associated variants, and subtype-associated metadata (sourced from ICGC, cBioPortal) are provided by *oncogene.iobio* to guide users in the analysis process, and to quickly elucidate suspect loci.

With extensive integration and immediate feedback, *oncogene.iobio* aims to aid the discovery of impactful variants in oncological research and diagnostic care.

## PgmNr 1042: The landscape of *BRCA1* and *BRCA2* putative protein-truncating variants from a UK Biobank population sampling.

### Authors:

Z. Ghazoui; A. Matakidou; S. Petrovski

[View Session](#) [Add to Schedule](#)

**Affiliation:** Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca. Cambridge Road, Royston SG8 6EH. England

Prior prevalence estimates of *BRCA1* and *BRCA2* protein-truncating variants (PTVs) have been generally limited to studies that have ascertained for cancer patients or individuals with a family history of cancer. As a community, access to a large population-based sampling of exome sequences linked to rich phenotypic data has not been available to perform a more real-world assessment. The aim of this study is to utilise one of the largest sequenced Biobank population to investigate the prevalence of *BRCA1* and *BRCA2* PTVs in the UK. We analysed whole-exome sequence data (WES) from an initial 100,000 UK Biobank participants (<https://www.ukbiobank.ac.uk/>), which is a prospective cohort study of 500,000 individuals that is linked to a wide variety of phenotypic data.

Among the 100K exome sequences, we identified a collection of 2,236 carriers of a predicted PTVs in *BRCA1/2*. Applying a panel of PTV *in silico* filtering to these variants, 1780 (79.6%) carriers were withdrawn based on the PTV affecting the final 5% of the gene (driven by two common SNPs that accounted for 1769 of these exclusions), 28 (1.3%) carried a PTV with a ClinVar Benign or Uncertain Significance classification, and 12 (0.5%) were restricted to a *BRCA1/2* transcript that is not among the consensus coding sequence (CCDS). The remaining 416 (18.6%) individuals are carriers of high-confidence rare protein-truncating variants (PTV) in *BRCA1* (124 PTV carriers) or *BRCA2* genes (292 PTV carriers). Of the total 416 individuals, 202 (48.6%) have been diagnosed with cancer, 125 (30%) have a reported family history of cancer but are not diagnosed with cancer in the UK Biobank self-report, doctor-diagnosed, or ICD10 cancer registry information, and approximately 89 (21.4%) have no reported history (or family history) of cancer. We also evaluated the distribution of cancer incidence in the UK Biobank among these 416 *BRCA1/BRCA2* PTV carriers and identified that ovarian (5.2% of reported), breast (1.9% of reported) and prostate (1% of reported) cancers are amongst the most commonly enriched types of cancers in this cohort.

This study provides a summary of the oncology landscape among *BRCA1/2* germline PTV carriers. Of further interest are the 21.4% of individuals with a germline *BRCA1/2* PTV that currently have no reported family history of cancer. This study further highlights the value derived from adopting a simple *in silico* screening of PTV annotated variants to reduce noise in interpretation of effects.



## PgmNr 1043: Germline associations with response to immune-oncology and adverse events in a large patient cohort.

### Authors:

S. Groha; S. Abou Alaiwi; E. LePisto; M. Manos; O. Rahma; T. Choueiri; M.L. Freedman; D. Schrag; K. Kehl; A. Gusev

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dana Farber Cancer Institute, Boston, MA.

---

### Background:

Immuno-oncology (IO) has led to an unprecedented improvement in durable responses in cancer therapy but only a fraction of patients derive lasting benefit, motivating research into IO response mechanisms and biomarkers. Current studies have primarily focused on rare somatic mutations or nonspecific features such as total tumor mutational burden. In contrast, only few studies have focused on germline factors.

### Methods:

Using a large cohort of IO patients in a clinical setting (N=600 non-small-cell lung cancer (NSCLC); N=1200 pan-cancer), we probe the influence of germline genetic factors on clinical outcomes of IO, including overall & progression free survival (OS/PFS), as well as multi-event progression trajectories abstracted from medical imaging reports with deep neural networks. We impute common SNP genotypes, germline HLA alleles, genetic ancestry, and GWAS polygenic risk scores (PRS) from panel sequenced tumors in addition to traditional somatic and clinical features. We perform multivariate survival analysis, considering covariates representing the heterogeneity of our data and ascertainment of patients. Using competing risk models we investigate germline and somatic influences on adverse events. We integrate published datasets to replicate and meta-analyze our findings in a combined pan-cancer cohort of ~2500 IO patients.

### Results:

We study the reproducibility of known biomarkers and look for novel associations to germline factors. In the NSCLC cohort we find significant associations between OS and a PRS for pan-cancer risk ( $q < 0.05$ ) and nominal associations between OS and PRS's for specific cancers, and auto-immune-related phenotypes (both  $q < 0.1$ ). We perform an association scan across common HLA alleles and identify nominal associations ( $p < 0.05$ ) in the NSCLC patient cohort.

### Conclusions:

We are continuing to investigate associations to complex IO outcomes and toxicities with the aim of developing clinical predictors. Our initial results suggest that multiple germline factors significantly influence progression and OS.



## PgmNr 1044: A powerful method to estimate cell-type specific QTLs from bulk expression by leveraging allelic imbalance.

### Authors:

C. Kalita; A. Gusev

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medical Oncology, Dana-Farber Cancer Institute, Boston, MA.

Quantitative trait loci (QTLs) have been crucial in providing an understanding of how genetic variants in regulatory sequences can affect gene expression levels that lead to cancer and influence progression. However, tumor and normal tissues are not a pure cell type, and disease related variants have been found to exert their effects in a cell type specific manner. Existing approaches to identify cell type specific QTLs are infeasible for large scale analysis of multiple tumor types and are currently underpowered. While methods exist to identify genotype-by-cell-type interactions from deconvoluted bulk RNA sequencing data, they require massive sample sizes to achieve adequate power (particularly for cell types present at low frequency), and have largely been unexplored. However, by measuring allelic imbalance (AI) within-individuals, additional power can be leveraged from read coverage as well as sample size. eQTL-interaction and AI tests are independent, and can thus be combined to further increase power to detect cell-type specific QTLs.

We developed a new method, DeCAF (DEconvoluted Cell type Allele specific Function), to identify cell-type specific QTL effects in tumors by leveraging both allelic and total expression information. We simulated allelic read counts modeling a variety of real life conditions and real cell proportion data from The Cancer Genome Atlas (TCGA) to investigate test performance. At a sample size of N=600, our combined test achieved >50% power to detect a cell-type specific eQTL with a low allelic fraction of 0.6. In contrast, the conventional eQTL interaction test required 1,000 individuals to achieve >50% power for the same effect size. As most TCGA sites have <600 individuals, our method thus empowers many new cell-type specific discoveries. Using real immune cell proportion data estimated from gene expression signatures in kidney cancer, only the combined test was well-powered (>75%) to detect allelic effects in CD8 T cells and Dendritic cells for low eQTL effect sizes. When effects were high (allelic fraction of 0.9 and MAF=0.5), the number of individuals required to reach >75% power was, on average, 1.8x lower than the conventional eQTL interaction test. Our method could lead to a better understanding of germline mechanisms underlying the anticancer immune response as well as identify the relevant cell types and QTLs contributing to cancer risk. We are now applying our methodology to real gene expression data from TCGA.

## **PgmNr 1045: Dealing with change, the only constant: Evaluation of a novel approach to rapidly develop and deploy optimized cancer hotspot panels for clinical research.**

**Authors:**

A. Kothandaraman; D. Fox; C. Clear; J. Stewart; K. Kushiro

[View Session](#)[Add to Schedule](#)

**Affiliation:** PerkinElmer, Inc., Austin, TX

---

The interrogation of mutation hotspots in oncogenes or tumor suppressors using genetic sequencing platforms such as Next Generation Sequencing is critical for the advancement of clinical oncology research. New hotspots are being discovered at a rapid pace and opinions on clinical relevance of these hotspots has varied among experts. Laboratories have the desire to adapt their targeted sequencing panel to the constantly evolving genomic landscape but do not have ready access to the necessary development and deployment tools. To address this challenge, we developed an automated design pipeline that provides flexibility to add and/or remove hotspot targets using a core panel as the backbone. The automated design pipeline's machine learning capability allows rapid development of new panels with optimized primer designs. A universal reagent formulation was used such that new designs are tied to a previously validated library prep chemistry. The poster describes the automated design pipeline and performance metrics associated with this novel approach.

For Research Use Only. Not for Use in diagnostic procedures.

## PgmNr 1046: Expression of folate-mediated one-carbon metabolism-related genes in cancer cells is associated with response to chemotherapy.

### Authors:

J. Krushkal; S. Vural; D.-J. Min

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Cancer Treatment and Diagnosis, National Cancer Institute, National Institutes of Health, Rockville, Maryland.

---

Folate-mediated one-carbon metabolism (OCM) is important for cell growth and survival. Malignant cells are dependent on OCM reactions, and folate cycle genes have been associated with cancer risk and progression. We investigated whether the response of cancer cells to chemotherapy may be partially influenced by pre-treatment expression levels of OCM genes. We used drug response data for 251 antitumor agents, OCM gene expression information, and additional data on genome sequence variants, gene amplification, and gene fusions in 635 cancer cell lines with matching information from the Cancer Cell Line Encyclopedia and the Genomics of Drug Sensitivity in Cancer resources. We examined whether pre-treatment expression of 34 OCM-related genes was associated with drug response. Selected association results were validated in the NCI-60 cancer cell line panel. Expression of *GART*, *TYMS*, *SHMT2*, *MTR*, *ALDH2*, *BHMT*, *MAT2B*, *MTHFD2*, *NNMT*, and *SLC46A1* showed modest correlations with response to a variety of antitumor agents. Higher expression levels of *SLC46A1* were associated with resistance to multiple drugs, whereas elevated expression of *GART*, *TYMS*, *SHMT2*, *MTR*, *BHMT*, and *MAT2B* was associated with chemosensitivity to multiple agents. *NNMT* expression was bimodally distributed and showed different directions of association with various agents. Expression of *MAT2B*, *MTR*, *SHMT2*, and *SLC46A1* was associated with response to crizotinib, and these associations remained significant after accounting for the presence of genome alterations (*ALK* fusions, *ROS1* fusions, *MET* amplification, or *MET* exon 14 skipping mutations) known to affect crizotinib sensitivity. Expression of several OCM genes was strongly associated with expression of components of drug target pathways. For example, expression of *TYMS* and *GART* was strongly positively correlated with *BRCA1* expression, *NNMT* expression was associated with *EGFR* and *ABL2*, and expression of *TYMS*, *DHFR*, and *SHMT1* was positively correlated with *AURKB*. Pre-treatment expression levels of *DHFR*, *TYMS*, *ATIC*, *GART*, *AHCY*, *MTHFD1*, *SLC19A1*, and other genes were positively correlated with each other, suggesting their co-regulation, whereas *NNMT* expression was negatively correlated with expression of several OCM genes. Correlations of expression of OCM genes with drug response could be related to metabolic roles of these genes, or they may be associated with the increase in the rate of cancer cell growth and proliferation and with tumor progression.

## PgmNr 1047: Pan cancer genomic characterization of patient-matched primary and brain metastases.

### Authors:

O. Lee <sup>1</sup>; A. Mitra <sup>1</sup>; W. Lee <sup>1</sup>; K. Fukumura <sup>4</sup>; H. Beird <sup>1</sup>; J. Weinstein <sup>2</sup>; M. Davies <sup>3</sup>; J. Huse <sup>4</sup>; P. Futreal <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas; 2) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas; 3) Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas; 4) Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas

---

Brain metastases (BM) occur in 25-50% of patients with cancer. Approximately 200,000 new cases of brain metastases are diagnosed in the United States each year, with median survival after diagnosis ranging from 3-27 months. Studies have identified clinically actionable mutations in BM that are distinct from those in the primary tumor samples such as *PTEN* nonsense and *PBRM1* frameshift mutation. Additional profiling and genomic analyses will provide deeper understanding of the pathogenesis of BM and suggest new therapeutic approaches. We collected a total of 166 tumor samples from 49 patients with four different cancer types: renal cell carcinoma (n=35), breast cancer (n=51), lung cancer (n=45), and melanoma (n=35). By studying 4 different tumor types, we aim to identify genomic features of brain metastasis that are not specific to tumor type. The samples included matched BM and primary tumors from 38 patients and matched BM and extracranial metastases (ECM) from 11 patients. We performed whole-exome sequencing at 200x depth and subsequent analyses to call somatic point mutations. Our initial analysis was based on functional somatic mutations. To determine whether somatic cancer-associated mutations we identified had been previously studied and functionally validated, we compared our calls with the Database of Curated Mutations, which includes highly curated lists of disease-causing mutations. We detected 32 curated cancer-gene mutations (13 unique genes) in the cohort and found that 23 of them (71%) were shared between BM and primary tumors or BM and ECM. This suggests that BM and primary tumors may have been derived from a common ancestral clone. Among the 9 non-shared mutations, 8 were from BM-primary tumor pairs, and 1 was from a BM-ECM pair, implying that BM could accumulate additional cancer gene mutations that are not present in both primary tumors and ECM. We examined the genes that were found mutated only in BM in at least 3 patients. We uncovered mutations in 4 genes that were not previously associated with metastasis: *GABRE*, *RUNX1*, *SCML2*, and *SPEG*. *RUNX1* mutations were recurrent in breast cancer patients, and mutations in *SCML2* were detected in a majority of the tumor histology. This study aims to confirm the landscape of BM-specific mutations in tumors from diverse sites. We will perform further analyses of indels and clonal relationships for better understanding of BM-specific sub-clones and their metastatic outgrowth in the brain.

## PgmNr 1048: Roles of genome-wide biased-allelic-expression (BAE) effect in tumor development from the perspective of liver hepatocellular carcinoma (LIHC).

### Authors:

C.H. Lo <sup>1,2</sup>; H. Nakaoka <sup>1,2</sup>; I. Inoue <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics Laboratory, Department of Genomics and Evolutionary Biology, National Institute of Genetics, Japan; 2) Department of genetics, School of Life Science, The Graduate University for Advanced Studies (SOKENDAI), Japan

---

The majority of germ-line or somatically derived mutations stay in heterozygous status among patients. Besides, according to genome wide association studies (GWAS), disease associated single nucleotide polymorphisms (SNPs) are mostly located in non-coding regions. Those SNPs are suspected of regulating pathogenic gene expression in quantitative trait loci (eQTL) by changing the transcription factors binding sites or by altering CTCF binding motifs. Either in cis- or trans- regulating ways, above mentioned SNPs are supposed to cause abnormal biased-allelic-expression (BAE) effect, where BAE means the transcriptional levels between two alleles of a gene are different.

Therefore, I presume that if the BAE effect is inclined to express the pathological allele higher, it may be associated with exacerbating disease status. Particularly for the unstable genomes such as cancer cells, where BAE prevails in genome-wide. And the emergence of BAE effects in some genes can play critical roles in oncogenesis.

I have integrated high-throughput sequencing data from The Cancer Genome Atlas (TCGA) and Genotype-Tissue Expression (GTEx) projects. Above two databases enable us to compare each liver hepatocellular carcinoma (LIHC) tissue to its corresponding paired normal-like tissue (from cancer patients), and also to those control-normal tissues (from non-cancer individuals).

Firstly, differential expression analysis implies that the transcriptional changes of tumor tissues might be caused from genomic disturbs, which let new dividing cells abort expressing tissue specific genes, and stochastically increase expressing other unrelated genes.

Secondly, tumors exist higher genome-wide BAE effects, and most BAE genes are not directly caused by their allelically unbalanced copy number variations (CNV) in DNA level but by the irregularity in their transcriptional level. And those BAE occurring rates locating in non-CNV regions display positive association with the total somatic CNV length of cell, which implies BAE effect associates with the instability of chromatin.

Thirdly, there are few commonly shared BAE genes inside each sample group, which implies most BAE effects initially fall on genes stochastically without specific targets through genome-wide distribution. Finally, I demonstrated that BAE genes present special distribution patterns in tumor tissues, which might be associated to the evolutionary process of pro-tumor cells inside body.

## PgmNr 1049: Analysis of whole exome sequencing data of hereditary lung cancer families identifies germline Copy Number Variations (CNVs) in multiple genes.

### Authors:

D. Mandal <sup>1</sup>; D. Tran <sup>1</sup>; K. Termine <sup>1</sup>; A. Musolf <sup>2</sup>; C. Pikielny <sup>3</sup>; M. de Andrade <sup>4</sup>; C. Gaba <sup>5</sup>; P. Yang <sup>4</sup>; Y. Li <sup>6</sup>; M. You <sup>7</sup>; R. Govindan <sup>8</sup>; R. Wilson <sup>9</sup>; E. Kupert <sup>7</sup>; M. Anderson <sup>7</sup>; A. Schwartz <sup>10</sup>; S. Pinney <sup>11</sup>; C. Amos <sup>6</sup>; J. Bailey-Wilson <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA, USA; 2) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, USA; 3) Geisel School of Medicine, Dartmouth College, Lebanon, NH, USA; 4) Mayo Clinic, Rochester, MN, USA; 5) Department of Medicine, University of Toledo Dana Cancer Center, Toledo, OH, USA; 6) Baylor College of Medicine, Houston, TX, USA; 7) Medical College of Wisconsin, Milwaukee, WI, USA; 8) Division of Oncology, Washington University School of Medicine, St. Louis, Missouri, USA; 9) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH, USA; 10) Karmanos Cancer Institute, Wayne State University, Detroit, MI, USA; 11) University of Cincinnati College of Medicine, Cincinnati, OH, USA

---

About 10% of lung cancer (LC) cases (22,000 cases per year) in the U.S. have at least one first-degree relative affected with LC, and about 1% of cases have at least three first- or second-degree affected relatives. To identify susceptibility gene(s) for hereditary lung cancer (HLC) in families ( $\geq 3$  LC/family) collected by the Genetic Epidemiology of Lung Cancer Consortium, we conducted whole exome sequencing (WES) on eight highly aggregated families ( $\geq 4$  LC cases/family). These families included samples from sixty-five individuals, twenty of whom were affected and forty-five were unaffected at the time of data collection. We used two CNV-specific algorithms as incorporated in CANOES (<http://www.columbia.edu/~ys2411/canoes/>) and XHMM (<https://atgu.mgh.harvard.edu/xhmm/>) to identify germline copy number variations (CNVs). To confirm, CNVs were then visualized independently using Integrative Genomics Viewer (IGV).

We identified CNVs in more than 25 genes that are deleted or duplicated in two or more individuals in two or more families. Furthermore, both CNV-specific algorithms identified four cancer related genes: *GSTM1* (1p13.3), *RHD* (1p36.11), *CFHR3* (1q31.3), and *CFHR1* (1q31.3) in more than five families in two or more affected individuals/family. Reports from other studies show CNVs in those genes in somatic LC samples. Previously, no germline specific CNVs have been reported in those genes in HLC families. Comparison with the available public database (Database of Genomic Variants or DGV) also demonstrate that these alterations are rare and are not common polymorphisms found within the general population. The results from the other available database on somatic mutations (The Cancer Genome Atlas (TCGA) and Catalogue of Somatic Mutations in Cancer (COSMIC)) verify that these alterations are found in a small number of lung cancer tumor samples. The CNVs identified in our study are found to be common within these high-risk families, which are overall a small portion of the population. This observation calls for more in-depth analysis to take place on the next generation sequencing data that might be useful for the development of HLC-specific biomarkers in the future.

## **PgmNr 1050: A flexible, high-throughput software pipeline for identifying candidate variants and their effects on protein structure, starting from NGS data or VCF files.**

### **Authors:**

K. Maxfield; M. Keyser; M. Larson; S. Baldwin; D. Nash; R. Nelson; S. Darnell; T. Durfee

[View Session](#) [Add to Schedule](#)

**Affiliation:** DNASTAR, Inc., Madison, Wisconsin.

---

Advances in DNA sequencing have made identifying genetic variation from any individual routine, allowing researchers to analyze variation across populations or cohorts for candidate mutations causing diseases or traits of interest. To address the significant bioinformatic, data management and computing resource challenges inherent to this analysis, we have developed an integrated software solution for each step in the pipeline.

First, variant call profiles for each sample are either determined directly from raw NGS sequencing data or extracted from existing VCF files. For NGS data processing, reads from each sample are aligned to a reference genome using a fast, non-memory bound assembler. Gapped alignments are analyzed in-stream using a Bayesian modeled probabilistic variant caller to produce annotated single nucleotide variant (SNV) and small indel calls. To leverage existing variant information in VCF format, we have also developed a VCF annotation tool which maps the variants onto a corresponding annotated reference sequence and “decorates” each variant with information such as the affected gene(s) and impact on protein encoding regions and/or splice sites.

Next, annotated variant profiles from each sample, supplemented with their allele frequencies as well as predictions about functional impact and pathogenicity, are automatically combined into a single project for analysis. Various filtering and statistical methods then enable candidate genes and/or variants of interest to be readily identified. Further, for genes with known 3D protein structures, the effect of candidate missense mutations on the structure can be automatically predicted and analyzed using our molecular structure visualization module.

As a demonstration of the pipeline, we will present results from our reanalysis of 96 targeted resequencing samples from a Chinese cohort with lung squamous cell carcinomas (LSCC, Li et al., Sci Rep. 2015. 5:14237). We will show how the software can be used to easily identify unique mutations in numerous samples across the cohort which all lead to nonsense and frameshift mutations in the *TP53* tumor suppressor gene. We will also show how filtering on the functional impact and pathogenicity predictions can be used to identify likely deleterious *TP53* missense mutations in other members of the cohort and how in one of those cases, the predicted protein structure change indicates that the DNA binding activity of the protein is likely directly affected.



## PgmNr 1051: The benign variants in the COSMIC cancer gene census confound somatic mutation interpretation.

### Authors:

K. Park <sup>1</sup>; W. Jang <sup>2</sup>; M. Kim <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Laboratory Medicine, Myongji Hospital, Goyang, Korea, Republic of; 2) Department of Laboratory Medicine, Inha University School of Medicine, Incheon, Korea, Republic of

---

**Background:** The presence of benign variants or passenger mutations in driver genes from public databases is one of the well-documented limitations when reporting clinical actionability of somatic mutations. However, our understanding on classification and interpretation of somatic mutations is far from being perfect. The aim of this study is to investigate the frequency of benign variants from the Catalogue of Somatic Mutations in Cancer (COSMIC) Cancer Gene Census (CGC) and the interpretation concordance of OncoKb variants in depth.

**Methods:** We used a total of 339,425 somatic mutations in CGC genes registered in the COSMIC. Multiple databases including CIViC and Gene Drug Knowledge database and *in silico* tools were used to interpretate the clinical actionability of the mutations. All mutations were reanalyzed by a four-tiered system based on the consensus recommendation of the Association for Molecular Pathology (AMP), American Society of Clinical Oncology, and College of American Pathologists, and compared to OncoKb classification.

**Results:** The frequency of benign or likely variants, which were categorized as Tier IV variants according to AMP guideline, in the COSMIC CGC was 1% (4,200/339,425) of the mutations. Although 75% (3,129/4,200) of the Tier IV variants were flagged as polymorphism, while there was no warning sign in remaining 25% of benign variants. In addition, 70% (2,952/4,200) of the Tier IV variants were recurrently observed in at least two samples. Overall concordance of actionable mutations between AMP classification (Tier I or Tier II) and OncoKb (Oncogenic or likely oncogenic) were 88% (3,118/3,565) of variants studied.

**Conclusions:** Interpretation and reporting of somatic mutations in cancer databases need to use with caution. We found that significant number of somatic mutations registered in COSMIC CGC may not be clinically-actionable mutations. In addition, we revealed that there was some disagreements in interpretation of oncogenic role and actionability of somatic variation according to the databases and classification criteria. This study highlights the importance of the reinterpretation of the reported mutations and could be a good starting point for in-depth review of somatic mutation information essential for precision oncology.



## PgmNr 1052: A denoising sparse autoencoder to decipher mutational signatures (DeepMS) in human cancers.

### Authors:

G. Pei <sup>1</sup>; Y. Dai <sup>1</sup>; Z. Zhao <sup>1,2,3</sup>; P. Jia <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA

---

Most cancers are originated from somatic mutations on genome. Somatic mutations are often associated with mutational processes that generate characteristic mutational signature. However, systematic understanding of the mutational processes is still far from enough. Previous studies have applied Non-negative Matrix Factorization (NMF) or related methods to decompose mutational signatures using large-scale somatic mutation data from next-generation sequencing. In this study, we developed a deep learning approach to systematically search for mutational signatures using the most recent somatic data curated by the Pan-Cancer Analysis of Whole Genomes Project (PCAWG). We focused on a total of 52,671,908 somatic mutations from 2780 whole cancer genome of 37 most common cancer types. We proposed a deep generative model, called Denoising Sparse Auto Encoder Neural Network (DSAENN), and identified 54 single base substitution (SNV), 11 doublet base substitution (DBS) and 16 small insertion and deletion (InDel) mutational signatures. We further developed a regression model to estimate the association of each signature with previously known mutagenesis process. As a result, we found that 37/54 SNV, 5/11 DBS and 9/16 InDel signatures of our findings are newly reported. Estimation of the explained variance of each mutational signature to each cancer type revealed cancer type specific signatures (such as tobacco smoking in squamous-cell lung carcinoma and solar ultraviolet exposure in melanoma) as well as general signatures (such as DNA repair deficient related process by POLE or APOBEC). These analyses provide an important supplementary resource to the repertoire of mutational processes.

## PgmNr 1053: Associations between cancer and non-cancer-related Mendelian disease genes and specific cancer types in TCGA.

### Authors:

B. Salisbury; S.M. Bray; B.G. Parikh; M.A. Schroering; T.D. Barber; S.S. Wong

[View Session](#) [Add to Schedule](#)

**Affiliation:** LifeOmic, Indianapolis, IN

---

**Background:** We hypothesized that genes and variants established to be involved in Mendelian diseases outside of cancer might also be predisposing for certain cancer types.

**Methods:** The germline genomes of over 10,000 cancer cases in The Cancer Genome Atlas (TCGA) from whole exomes sequencing were loaded on LifeOmic's Precision Health Cloud platform and annotated using the Gnosis knowledge system. We examined the distribution of all variants in ClinVar for enrichment in the 32 cancer types. For control comparison, we conducted the same analysis for 152 cancer-predisposing genes. Enrichment was assessed in comparison with the gnomAD-exomes database of reference population frequencies.

**Results:** Consistent with expectations, ClinVar "pathogenic" and "likely pathogenic" variants in known and suspected cancer-predisposition genes were significantly more common in cancer patients than in the gnomAD-exomes population pool, whereas "benign" and "likely benign" variants were not enriched. This enrichment included striking cancer-type specific patterns in cases such as *BRAC1* and *BRCA2* with breast and ovarian cancers and *RET* with pheochromocytoma and paraganglioma. The same pattern, but less pronounced, was found for non-cancer genes: pathogenic and likely pathogenic variants were overall more common in cancer cases than in gnomAD-exomes, while benign variants were not. Some non-cancer genes showed patterns of specific cancer-type enrichment; other variants and genes showed enrichment in a pan-cancer fashion.

**Discussion:** Our results suggest that previously unrecognized relationships may exist for cancer risk and Mendelian disease genes. The etiology of cancer in these situations is unclear, perhaps involving organ system damage from the primary disease or treatment. This analysis also supports clarification of the pathogenicity of variants of uncertain significance in known cancer-predisposition genes (also enriched in cancer cases) and the extension of some genes' relevance to cancer types with which they are not traditionally associated.

## PgmNr 1054: Systematic identification of MYC-dependency in non-MYC-amplified tumors across 19 tumor types.

### Authors:

E. Schaafsma <sup>1</sup>; Y. Zhao <sup>1</sup>; C.C. Cheng <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Systems Biology, Dartmouth, Lebanon, New Hampshire.; 2) Department of Medicine, Baylor College of Medicine, Houston, TX; 3) The Institute for Clinical and Translational Research, Baylor College of Medicine, Houston TX

---

Transcription factor MYC is a well-known oncogene in many cancer types. It is estimated that 70% of all tumors exhibit MYC misregulation. Although the most common type of MYC alteration is MYC amplification, several other mechanisms can contribute to dysregulation, including: amplification of MYC family members MYCN and/or MYCL, point mutations in the MYC coding region, stabilization of MYC mRNA, reduced methylation of the MYC gene or, activation of signaling pathways that augment MYC expression. In order to establish a metric that can capture all of these MYC-related alterations, we established a tumor-specific MYC activity score. We show that our activity score captures MYC amplification status, but also additional MYC alterations, including MYCN or MYCL amplification, decreased methylation, and increased MYC expression. We identified that MYC activity scores are more clinically relevant compared MYC amplification status in several tumor types. In addition, we could risk-stratify patients without MYC amplification based on their MYC activity scores, even after adjustment for clinical variables. Lastly, we identified several oncogenic mutations and pathways that might contribute to high MYC activity in non-MYC amplified samples. In conclusion, we show that several mechanisms can activate MYC in non-MYC amplified tumor samples and that this can limit prognosis and contribute to tumor progression.

## PgmNr 1055: TOD-CUP: A gene expression rank based algorithm for tissue origin diagnosis of cancer of unknown primary across different data types.

### Authors:

Y. Shen <sup>1,2,3</sup>; W. Jiang <sup>2</sup>; K. Chen <sup>1</sup>; L. Fan <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas.; 2) Cancer Biotherapy Center, First Affiliated Hospital, School of Medicine, Zhejiang University; 3) Institute of Bioinformatics and IBM Bio-computational Laboratory, Zhejiang University

---

Carcinoma of unknown primary (CUP) represents a group of heterogeneous cancers and is defined by the presence of metastatic disease without an identifiable primary tumor site on presentation. It constitutes 3-5% of all human malignancies. This type of cancer is called a cancer of unknown primary (CUP) or occult primary tumor. Some biomarkers based on gene expression profiles had been identified to aid cancer tissue origin diagnosis of cancer of unknown primary. However, because of the batch effect and differences between different data types of RNA-seq and microarray measurements and heterogeneity of cancer patients, the predictive performances of biomarkers based on gene expression profiles often decrease greatly in inter-laboratory and different data type validation. To address this problem and assist in more precise diagnosis, we developed a general gene expression rank-based algorithm for tissue origin diagnosis for cancer of unknown primary (TOD-CUP). Massive (10553) tissue-specific RNA-Seq data sets were first obtained from The Cancer Genome Atlas (TCGA) and 538 feature genes were used to train and validate the TOD-CUP algorithm. The accuracy of the model was >97% based on cross validation by the data from TCGA. To validate the TOD-CUP method can accurately predict the tissue origin of CUP in different data types, 1,029 TCGA microarray cancer samples which generated by Agilent microarray platform and 2,277 microarray data set which generated by Affymetrix and Illumina microarray platform were tested in the TOD-CUP method using TCGA RNA-seq data as the training data sets. The accuracy of the method in these two microarray data set is 91% and 94%. Furthermore, we curated a data set which included 141 metastatic cancer tumor samples RNA-seq data from five cancer types to evaluate the performance of the TOD-CUP method in tissue of origin diagnosis of the metastatic cancer. The accuracy of the TOD-CUP method in the metastatic cancer tumor samples is 94%. Based on our findings, we believe that the TOD-CUP algorithm is a powerful robust methodology to accurately identify the tissue origin of cancer of unknown primary across different data types and an important step toward more precision-based medicine in cancer diagnosis and treatment.

## **PgmNr 1056: Mutational signatures and clinicopathological relationship extracted from clinical sequence data of 201 cases with colorectal cancer.**

### **Authors:**

S. Takeuchi <sup>1</sup>; L. Yiwei <sup>1</sup>; Y. Watanabe <sup>1</sup>; Y. Shimada <sup>2</sup>; T. Wakai <sup>2</sup>; S. Okuda <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Divisions of Cancer Genome Informatics, Graduate School of Medical and Dental Sciences, Niigata University; 2) Divisions of Digestive and General Surgery, Graduate School of Medical and Dental Sciences, Niigata University; 3) Divisions of Bioinformatics, Graduate School of Medical and Dental Sciences, Niigata University

---

The concept of mutational signatures has significantly improved our understanding of relationship between mutagenesis and carcinogenesis. Although signatures of more than 30 mutational processes have been identified in cancer genomes through thousands of experiments, many of which still have unknown etiologies. Tracing of their origins may help for the development of potential ways to reduce the risk of getting cancer. Mutational signatures may also be useful predictor of a response to different treatments. Therefore, elucidating the cause of unknown mutational signatures may open new avenues for both cancer prevention and cancer treatment. In this study, signatures were depicted using substitution classification defined the substitution type and sequence context from our archived comprehensive targeted panel sequencing data of 201 Japanese patients with non-hereditary colorectal cancer. We then performed extraction of mutational signatures, estimated the contributions of each clinicopathological factor and frequently mutated gene. We found that multiple clinicopathological factors and frequently mutated genes are significantly associated with several types of mutational signatures.

## PgmNr 1057: Integrated transcriptomic-genomic tool Texomer profiles cancer tissues.

### Authors:

F. Wang <sup>1</sup>; S. Zhang <sup>1</sup>; T. Kim <sup>1</sup>; Y. Lin <sup>1</sup>; R. Iqbal <sup>1</sup>; Z. Wang <sup>1</sup>; V. Mohanty <sup>1</sup>; K. Sircar <sup>1</sup>; J. Karam <sup>1</sup>; M. Wendl <sup>2</sup>; F. Meric-Bernstam <sup>1</sup>; J. Weinstein <sup>1</sup>; L. Ding <sup>2</sup>; G. Mills <sup>1</sup>; K. Chen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The University of Texas MD Anderson Cancer Center, Houston, TX; 2) Washington University School of Medicine, St. Louis, MO

---

Achieving precision and individualization is the key to further advance the understanding and treatment of cancer. Molecular profiling of tumor using bulk DNA sequencing are of limited power and precision. Multi-omics profiling promises comprehensive, functional readout of a tissue sample. However, novel, systematic approaches are required to leverage the increased data dimensionality, heterogeneity and complexity.

To address the issues of intra-tissue heterogeneity and integration of multi-omics in cancer, we developed a statistical approach called Texomer that enables a joint analysis of bulk whole exome (WES) and whole transcriptome sequencing (WTS) data to perform allele-specific deconvolution and quantify tumor purity and heterogeneity. Texomer first estimates tumor DNA purity, intratumor heterogeneity (ITH) and allele-specific copy numbers (ASCNs) using germline single nucleotide polymorphisms (SNPs) and somatic single nucleotide variants (SNVs) from WES data. It then estimates tumor RNA purity and allele-specific expression levels (ASELs) through the autologous WTS data, given the estimated DNA purity and ASCNs. Texomer further probabilistically classifies each SNP and SNV as copy number concordant or discordant, which can be used to reveal functionally selected variants in copy number altered regions. Finally, Texomer estimates a differential allelic cis-regulatory effect (DACRE) score to quantify the tumorigenic potential of a variant allele relative to its wildtype. Evaluation using simulated data and multiple datasets from the cancer genome atlas indicated that Texomer achieved desirable technical accuracy and outperformed existing tools. Based on Texomer-transformed profiles, we found more accurate genotype-phenotype association and breast cancers categorization, resulting in more clusters with homogeneous profiles and distinct biological properties than do the bulk data. In addition, the improved power achieved by Texomer manifested in significantly improved accuracy for functional variant prioritization. Applying DACRE score as a filter doubled the precision in predicting the known functional variants.

Our study clearly revealed the analytical challenges involved in performing joint WES and WTS profiling of patient samples, and delivered a statistically robust solution Texomer to accomplish the benefits of multi-omics profiling towards the realization of personalized genomic medicine.

## **PgmNr 1058: Expressional single nucleotide variations (eSNV) of autophagy genes are associated with poor survival of lung adenocarcinoma patients.**

### **Authors:**

J. Zhang <sup>1</sup>; Z. Yi <sup>2</sup>; W. Zhang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Bergen County Academies, Hackensack, NJ; 2) Icahn School of Medicine at Mount Sinai, New York, New York

---

Expressional single nucleotide variations (eSNV) are caused by genetic variation, somatic mutations or RNA editing and play oncogenic roles in cancer development. To comprehensively characterize the eSNVs associated with human cancers, we developed a robust computational framework for identification and quantification of non-synonymous eSNV from RNA sequencing data. The framework performs sequence alignment, variant calls, and subsequent counting of the reads that cover the variant alleles and corresponding reference alleles, respectively. The read counts for the non-synonymous variants and reference alleles are normalized and correlated with clinical outcomes after adjustment for demographic characteristics. We applied this approach to RNA sequencing data from a TCGA lung adenocarcinoma cohort (LUAD, n=459) and identified 22295 non-synonymous eSNVs occurring in at least 5% of patients. Cox regression analysis with the counts of eSNV and reference alleles identified expression levels of 579 eSNVs (but not corresponding reference alleles) associated with poor overall survival. Interestingly, the most significantly enriched functions for 460 genes harboring these prognostic eSNVs are autophagy and viral process with p values less than  $1.0 \times 10^{-8}$ . These preliminary findings might shed light on the role of autophagy in LUAD progression at the transcriptional level and will be extensively validated.

## PgmNr 1059: Comparison and evaluation of pathway and gene-level methods for cancer prognosis prediction.

### Authors:

Zheng <sup>1</sup>; C.I. Amos <sup>1,2</sup>; H.R. Frost <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Hanover, NH; 2) Department of Medicine, Baylor College of Medicine, Houston, TX

---

Cancer prognosis prediction has become an important research goal. A limited number of gene-level analyses have led to clinically useful methods like Oncotype DX but these remain very difficult to develop and implement. A promising direction for improving the performance and interpretation of expression-based predictive models involves aggregating gene-level data into biological pathways. Although a few studies have used pathway-level predictors, a comprehensive comparison of pathway-level and gene-level prognostic models has not been performed.

To address this gap, we characterized the performances of penalized Cox proportional hazard models built using either pathway or gene-level predictors for the cancers profiled in The Cancer Genome Atlas (TCGA) and pathways from the Molecular Signatures Database (MSigDB). When analyzing the TCGA data, we found that pathway-level models are more parsimonious and easier to interpret than the gene-level models without a loss of predictive performance. For example, both pathway and gene-level models have an average Cox concordance index of 0.85 for the TCGA glioma cohort, however, the gene-level model has twice as many predictors on average and the predictor composition is less stable across cross-validation evaluations. In simulations, when the correlation structure of the real data is broken, the pathway-level models have greater predictive performance and superior interpretative power relative to the gene-level models. For example, the average concordance index of the pathway-level model is 0.88 while the gene-level model falls to 0.56 for the TCGA glioma cohort.



## PgmNr 1060: Comprehensive analysis of structural variants in clinical cancer samples.

**Authors:**

E.T. Lam; A.W.C. Pang; T. Anantharaman; J. Wang; J. Velazquez-Muriel; T. Wang; D. Zhang; P. Tandon; S. Way; A.R. Hastie; M. Borodkin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bionano Genomics, San Diego, California.

---

Tumors are often comprised of heterogeneous populations of cells, with certain cancer-driving mutations at low allele fractions in early stages of cancer development. Effective detection of such variants is critical for diagnosis and targeted treatment. However, typical short sequence read sequencing is expensive at coverage depths needed for detection of variants in rare clones. Short read sequencing is also limited in its ability to span across repeats in the genome and this results in high error rates in structural variant (SV) analysis. Based on specific labeling and mapping of ultra-high molecular weight DNA, we developed a single-molecule platform that is able to detect disease-relevant SVs and give a high-resolution view of tumor heterogeneity.

We have developed a pipeline that effectively detects structural variants at low allele fractions. It includes single-molecule based SV calling and fractional copy number analysis. Preliminary analyses using simulated data and well-characterized cancer samples showed high sensitivity for variants of different types at as low as 5% allele fractions with reasonable genomic coverage easily collectable on a Bionano Saphyr Chip.

The candidate variants are annotated and further prioritized based on control data and publically available annotations such as DGV and dbVar. The SQL-based infrastructure also allows new databases to be incorporated as they become available. Also, we have developed an algorithm to detect loss of heterozygosity events, which have been reported in cancer genomes.

The data are imported into a graphical user interface tool that includes new visualization features (such as dynamic variant filtering, Circos diagrams, and report generation) for interactive visualization and curation. Together, these components allow for efficient analysis of cancer genomes of interest.

## PgmNr 1061: A general probabilistic algorithm to predict *de novo* mutations in familial diseases as demonstrated in Li-Fraumeni syndrome.

### Authors:

E.B. Dodd <sup>1</sup>; X. Pan <sup>1,2</sup>; F. Gao <sup>1,3</sup>; J. Bojadzieva <sup>4</sup>; P.L. Mai <sup>5</sup>; V.E. Johnson <sup>2</sup>; K. Zelley <sup>6</sup>; K.E. Nichols <sup>6</sup>; J.E. Garber <sup>7</sup>; S.A. Savage <sup>5</sup>; L.C. Strong <sup>4</sup>; W. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX.; 2) Department of Statistics, Texas A&M University, College Station, TX; 3) Department of Statistics, Rice University, Houston, TX; 4) Department of Genetics, The University of Texas MD Anderson Cancer Center, Houston, TX; 5) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 6) Division of Oncology, Children's Hospital of Philadelphia, Philadelphia, PA; 7) Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA

---

Germline *de novo* mutations (DNMs) have been increasingly recognized as causal factors for rare diseases. The ability to identify heterozygous DNM carriers, scarcely distributed among all mutation carriers, will allow researchers to study the likely distinct molecular mechanisms of DNMs in these diseases. However, testing the mutation status of the parents in retrospective studies is often impractical due to lack of blood samples. Using Li-Fraumeni syndrome (LFS), caused by heterozygous germline pathogenic variants in *TP53*, as a representative of a rare inherited syndrome, we developed Famdenovo to predict the *de novo* status (*de novo*/familial) of germline mutation carriers.

We collected a total 324 LFS family pedigrees with confirmed *TP53* germline mutations from four medical centers in the US. 186 pedigrees with available *TP53* genetic testing results in at least one pedigree trio served as a validation set and the remaining 138 families that did not have testing for a full trio served as a discovery set to identify potential germline *TP53 de novo* mutation carriers. In the validation set, the area under the ROC curve (AUC) of Famdenovo was 0.95 (95% confidence interval [CI] 0.92-0.98) and the ratio of the observed to expected (OE) was 1.32 suggesting excellent ability for discrimination and good concordance. In the discovery set, we predicted an additional 40 individuals (95% CI 30-50) were DNM carriers, which increased the total number of DNM carriers to 82 across the four medical institutions.

We observed similar distributions of ages-of-onset at specific cancer sites (breast, brain, leukemia, osteosarcoma, soft tissue sarcoma, lung, and adrenal cortical carcinoma) across the validation and discovery sets. Interestingly, lung cancer only occurred in female DNM carriers (n=4). Among *TP53* hotspot mutations, R248Q was more likely to be a DNM. Our new statistical method, Famdenovo, provides the probability of a person carrying a *de novo* germline mutation in rare inherited syndromes when the mutation status of both parents is not available.

Famdenovo, a freely available R package downloaded from <http://bioinformatics.mdanderson.org/main/Famdenovo>, is a general tool and can be applied to other

cancer genes where there is a good understanding of the penetrance of the associated disease phenotype. The computer-based identification of DNM carriers in *TP53* may enable future epidemiological studies of the unique cancer diagnoses in DNM carriers.

## PgmNr 1062: Systematic identification of mutational hotspots in hereditary cancer susceptibility genes.

**Authors:**

C. Ede; S. Strom

[View Session](#) [Add to Schedule](#)**Affiliation:** Dept of Variant Curation, Fulgent Genetics Inc, Temple City, CA

---

The classification of DNA-sequence variants into pathogenic, uncertain, or benign considers a broad spectrum of criteria, from over-representation in affected individuals to the specificity of gene-disease association. Classification of missense variants remains a challenge as in the majority of cases, experimental functional evidence is either not available or clinically indeterminate. Here we propose and evaluate a method used to detect areas of potential functional and clinical importance in the absence of such experimental evidence. This method aims to directly address one of the classification criteria established by the guidelines for clinical interpretation of sequence variants by the American College of Medical Genetics (ACMG): “variant is located within a mutational hotspot or functionally critical domain (PM1)”. To date, systematic approaches to predict PM1 have not been sufficiently rigorous to implement in a clinical setting. Here we demonstrate a new, binomial likelihood-based method for the detection of fragments exhibiting clinically relevant mutational clustering. The algorithm was tested using a set of published mutations in 127 genes associated with various forms of hereditary cancer predisposition. Results demonstrated the capability of detecting both potential pathogenic hotspots enriched for pathogenic variants and benign “coldspots” depleted of them. Previously established mutational hotspots were confirmed with high confidence, while other, apparently clustered regions were demonstrated to be likely a result of a randomized pattern. The simple input structure and universal calculation enables detection of mutational hotspots for a variety of purposes, including classification of disease causing variants. This method may also be a valuable tool for the recognition of patterns in the distribution of functionally relevant protein structures. It could foster a deeper understanding of the molecular basis of disease development and protein function, potentially enabling the refinement of variant classification during clinical DNA analysis and assist laboratories to improve variant classification and reduce the number of variants of uncertain significance (VUS).

## PgmNr 1063: 20x faster and more accurate UMI-aware consensus molecule creation with the Sentieon Genomics Tools.

**Authors:**

D. Freed; H. Chen; X. Liu; J. Hu; Z. Li

[View Session](#) [Add to Schedule](#)

**Affiliation:** Sentieon Inc., Mountain View, CALIFORNIA.

---

Next-generation sequencing (NGS) is increasingly used in sensitive clinical applications including liquid biopsy for monitoring of residual disease in cancer and detection of subclonal tumor populations. Due to the high accuracy required in these assays, elimination of false-positive signals due to errors introduced during library preparation and sequencing is critical. Towards this end, specialized library preparation methods have been developed where unique molecular identifiers (UMIs) are added to fragments molecules prior to amplification and sequencing. Using this UMI information, the sequence of the original molecule can be inferred computationally, significantly reducing the impact of library preparation and sequencing artifacts.

Here we present an improved pipeline for determination of molecular consensus molecules from UMI tagged NGS reads that is 20x faster than currently available pipelines. Compared to existing pipelines, our pipeline uses improved logic to identify reads derived from the same source molecule while accounting for sequence changes from library or sequencing artifacts. We also introduce a new model for SNP and INDEL errors that is calibrated directly from the read data allowing for more accurate removal of PCR artifacts and sequencing errors. Using publicly available benchmark datasets, we test our pipeline with popular variant callers and demonstrate improved variant calling accuracy.

## PgmNr 1064: Inference of tumor cell phylogenies using single cell sequencing.

**Authors:**

S. Hui <sup>1</sup>; R. Nielsen <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Center for Computational Biology, University of California, Berkeley, Berkeley, CA.; 2) Department of Integrative Biology, University of California, Berkeley, Berkeley, CA.; 3) Department of Statistics, University of California, Berkeley, Berkeley, CA.

---

Although large consortia have sequenced thousands of cancer patients' blood and tumor samples, tumor evolution remains a poorly understood process at the single cell level. Single cell whole genome sequencing lends itself well to population genetics approaches, but current methods are insufficient for application in cancer studies due to common core simplifying assumptions (such as random mating, consistent ploidy, and constant mutation rate). We present a method to better illuminate tumor evolution, by creating an inference framework to accurately call copy number changes from paired tumor/normal single cell sequencing data.

In order to account for technical noise from low coverage single cell whole genome sequencing data and to account for clonal cell growth in tumor development, we developed a Hidden Markov Model to jointly analyze read depth data from a collection of tumor cells, with matched normal cells as negative controls. By performing joint pairwise copy number calls, we are able to identify shared and singleton copy number changes in tumor cells. Furthermore, by creating a novel similarity metric based on the shared branch length in these pairwise copy number calls, we adapt existing population genetics methods to estimate cell development phylogenies. Using a combination of public datasets and simulations, we show we can accurately decode copy number profiles and reconstruct trees of tumor development. These trees can be used to obtain estimates of tumor growth rates, evidence of selection, and for inferring age of mutations and copy number changes, as well as estimates of tumor purity, tumor heterogeneity, and prevalence of tumor subpopulations.

## PgmNr 1065: Identification of a novel group of endometrial cancer patients with loss of thyroid hormone receptor expression and improved survival.

### Authors:

D. Piqué <sup>1</sup>; J. Greally <sup>2</sup>; J. Mar <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Systems and Computational Biology, Albert Einstein College of Medicine, Bronx, New York; 2) Dept. of Genetics, Albert Einstein College of Medicine, Bronx, New York; 3) Australian Institute for Bioengineering and Nanotechnology, The University of Queensland, Brisbane, Australia

---

Endometrial cancer (EC) is the most common gynecologic cancer in women, and the incidence of EC is increasing. Though most EC has a good prognosis, the loss of nuclear hormone receptor (NHR) expression defines subtypes of EC that are often associated with poor clinical outcomes. For example, estrogen receptor (ER)-negative EC typically harbors a worse prognosis compared to ER-positive EC. In addition, resistance to hormonal therapies such as progestin and aromatase inhibitors, which block the conversion of testosterone to estrogen, is a major clinical issue and may be mediated in part through the loss of NHR expression. Though several studies have identified associations between clinical outcomes, such as survival and tumor grade, and NHR expression loss, the molecular basis for and consequences of the loss of NHR expression in endometrial tumors is largely unknown. Furthermore, there is no statistical framework designed to identify tumors that lose NHR mRNA expression relative to normal tissue. The development of such an approach could aid in the classification of patients into subgroups that are associated with survival or treatment outcomes. Here, a new method, termed receptLoss, is developed for identifying hormone receptor expression loss in endometrial cancer relative to adjacent normal tissue. Using a score-based approach, we identify 6 out of the 49 NHRs that are highly expressed in normal tissue and that exhibit expression loss in a subset of endometrial tumors. Several of the NHRs previously known to lose expression in certain endometrial cancers, such as estrogen, progesterone, and androgen receptors, were among the 6 NHRs identified. A novel association was found between thyroid hormone receptor beta (*THRB*) expression loss and increased rates of 5-year survival. *THRB* expression loss occurs independently of estrogen and progesterone expression loss, suggesting that this subgroup of patients is indeed different from previously reported subgroups. Developing a quantitative approach to identify NHR expression loss in endometrial tumors could lead to the development of biomarkers of disease progression and personalized hormonal treatment strategies for patients with endometrial cancer.

## PgmNr 1066: Recovery of genetic and epigenetic heterogeneity using single-cell DNA and methylome sequencing.

### Authors:

C. Wu <sup>1</sup>; N.R. Zhang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Perelman School of Medicine , University of Pennsylvania, Philadelphia, Pennsylvania.; 2) Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania.

---

Genetic and epigenetic heterogeneity within tumor samples reflect the molecular mechanisms underlying tumor evolutionary dynamics. Intratumor heterogeneity is commonly studied at the clonal level where a “clone” is usually defined by DNA mutations such as single nucleotide variations (SNVs) and copy number variations (CNVs). Recently developed single-cell DNA-sequencing (scDNA-seq) enables mutation detection at the single cell level. In addition to DNA mutations, epigenetic heterogeneity is also prevalent in tumor samples. Single-cell methylome sequencing enable methylation detection in individual cells. These two types of sequencing data are low-coverage and plagued by dropout events, and computational methods are required to address these issues. Currently, mutation detection methods for scDNA-seq and methylation profiling methods for single-cell methylome sequencing data are lacking in accuracy and scalability. Although the two types of data reflect different genomic features, they can be treated under the same statistical and computational framework.

In this study, we develop a statistical framework to denoise scDNA-seq and single-cell methylome sequencing data using an unsupervised neural network called autoencoder. For scDNA-seq imputation, we borrow the raw allele frequency of alternative alleles/CNV states across all detected SNV/CNV sites in all cells to impute the missing genotypes and denoise all positions. Results show that the autoencoder outperforms two existing methods in terms of SNV signal estimation. We found that the autoencoder was able to impute missing SNV genotypes under different simulated scenarios. The SNV signals estimated by the proposed method allows the recovery of underlying clonal structures. A similar strategy is employed to impute the methylation states for single-cell methylome data. Our results reveal the potential of low dimensional latent space methods in the recovery of genetic and epigenetic signals for scDNA-seq and single-cell methylome sequencing data, facilitating the characterization of intratumor genetic and epigenetic heterogeneity.



## PgmNr 1067: HiC-Hiker: A probabilistic model to determine contig orientation in chromosome-length scaffolds by Hi-C.

### Authors:

R. Nakabayashi; S. Morishita

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Japan

---

Detection of structural variants (SVs) such as inversions, duplications and translocations in cancer genomes has been crucial to understand the mechanisms and consequences of cancer genome instability, demanding an accurate and inexpensive way to detect missing SVs. Recent studies demonstrated that Hi-C could be a powerful and cost-effective means to output chromosome-length scaffolds for non-model species with no genome marker resources, because the Hi-C contact frequency between a pair of two loci can be a good estimator of their genomic distance even if there is a large gap between them.

We attempted to extend this novel *de novo* assembly technique to be applicable to accurate cancer genome SV detection. However, we found it challenging to reduce errors in contig orientation because shorter contigs have less contacts with their neighboring contigs. There is a pressing need to overcome this problem because these orientation errors can cause a number of false-positive SVs and disturb the analysis.

To correct these contig orientation errors in draft genome assembly from Hi-C data, we here propose a new algorithm and its implementation named HiC-Hiker, which has a firm grounding in probabilistic theory, models Hi-C contacts across contigs rigorously, and effectively infers most probable orientations by Viterbi algorithm. We compared HiC-Hiker and 3D-DNA using a human genome contigs generated from short reads, evaluated their performance, and observed a remarkable reduction in contig orientation error rate from 3.0% (3D-DNA) to 1.0% (HiC-Hiker). Of note, HiC-Hiker outperformed 3D-DNA in repairing orientation of both short and long contigs simultaneously.

Our method achieved a higher accuracy in estimating contig orientation because it can consider long-range information between distal contigs, and estimates Hi-C read contact probabilities among contigs precisely. Our probabilistic approach to Hi-C scaffolding could be also promising for generating most probable ordering of contigs.

## PgmNr 1068: Comprehensive structural and copy-number variant detection with long reads.

**Authors:**

A.M. Wenger; A. Toepfer; Y. Li; L. Hickey

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pacific Biosciences, Menlo Park, CA

---

To comprehensively detect large variants in human genomes, we have extended pbsv – a structural variant caller for long reads – to call copy-number variants (CNVs) from read-clipping and read-depth signatures. In human germline benchmark samples, we detect more than 300 CNVs spanning around 10 Mb, and we call hundreds of additional events in re-arranged cancer samples.

Long-read sequencing of diverse humans has revealed more than 20,000 insertion, deletion, and inversion structural variants spanning more than 12 Mb in a typical human genome. Most of these variants are too large to detect with short reads and too small for array comparative genome hybridization (aCGH). While the standard approaches to calling structural variants with long reads thrive in the 50 bp to 10 kb size range, they tend to miss exactly the large (>50 kb) copy-number variants that are called more readily with aCGH and short reads. Standard algorithms rely on reference-based mapping of reads that fully span a variant or on *de novo* assembly; and copy-number variants are often too large to be spanned by a single read and frequently involve segmentally duplicated sequence that is not yet included in most *de novo* assemblies.

We extended pbsv to utilize signatures of read clipping and read depth to detect CNVs missed with the spanning read and *de novo* assembly approaches. PacBio long-read sequencing has very consistent coverage across the genome, which combined with clipped reads at CNV boundaries, provides power to detect CNVs. We apply the extended pbsv to characterize CNVs in Genome in a Bottle germline benchmark samples and a highly-rearranged breast cancer lines. Together with insertion, deletion, inversion, duplication, and translocation calling from spanning reads, this allows pbsv to comprehensively detect large variants from a single data type.

## PgmNr 1069: Mela-Immu portal: An open platform for exploring immunotherapy-related multidimensional genomic profiles in Melanoma.

### Authors:

J. Yang <sup>1,2</sup>; Q. Liu <sup>1,2</sup>; Y. Shyr <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, Tennessee.; 2) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee.

---

**Aims:** Although immune checkpoint blockade (ICB) antibodies therapies provide significant clinical benefits for advanced melanoma, predictors of response to these therapies remain incompletely characterized. There are increasing studies to identify genetic and nongenetic factors that may influence sensitivity or resistance to immunotherapy. In order to make these data available, we developed Mela-Immu, (<http://bioinfo.vanderbilt.edu:3838/Mela-Immu/>), which is an open-access resource for interactive exploration of a variety of known and novel signatures predictive of immunotherapy response across multiple datasets in melanoma.

**Methods:** We collected all public available datasets which have both genomic profiling and ICB therapy outcome. All the datasets are stored at different genetic profiling, including genomics, transcriptomics and cell subpopulations and correlated with three types of clinical outcome, immunotherapy response, overall survival and progression free survival.

**Results:** The current version of Mela-Immu contains 889 samples derived from both bulk sequencing and single cell RNA sequencing. The association between each genetic profiling and clinical outcome can be queried, visualized, and downloaded across multiple datasets to explore the consistency of signatures. Each genetic profiling provides multiple types of features, including the known or novel signatures predictive of immunotherapy response. Genomic profiling provides gene mutation, tumor mutational burden, and mutational signatures. Transcriptomic profiling includes gene/gene sets expression, gene expression relations, and immune cell components. The single cell RNA-seq contains gene expression signals of cells and cell population components. In addition, integrative analysis of multi-features is also implemented in Mela-Immu to identify potential predictors. To be noted, Mela-Immu allows users to upload their own datasets, therefore they can be analysed independently or co-analyzed with Mela-Immu's existing datasets to validate the reliability and generality of identified signatures.

**Conclusions:** With the ability of easily exploring the relationship between multidimensional genetics data and ICB therapy response across multiple datasets, Mela-Immu will facilitate discoveries in immune-mediated therapy of melanoma.

## PgmNr 1070: Aggregating multiple rare variant association tests to enhance gene discovery in rare variant association studies.

### Authors:

Y. Yu; J. Chen; R. Bohlender; C. Huff

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX

---

Gene-based association tests have been widely used in rare variant association studies to evaluate statistical evidence of disease association, yielding numerous novel discoveries of susceptibility loci. Various statistical tests have been proposed in recent years, such as burden tests, variance-component tests, and combined omnibus tests. Typically, a rare variant association study applies between one and three of these tests to evaluate all coding variants in each gene. This approach has three potential disadvantages: First, selecting the optimal test is difficult as the power of each test varies according to genetic architecture, which differs by gene and is generally unknown. Second, using more than one test increases the multiple testing burden, which may reduce overall statistical power. Third, testing all coding variants in a gene, irrespective of transcript isoform, is often approximately equivalent to a test of the largest isoform, which is not always optimal. To address these questions, we introduce the X-Method Association Test (XMAT), which is a permutation approach to conduct gene-based tests, incorporating multiple statistics and all isoforms in a given gene. XMAT summarizes the contribution of each test and each transcript to calculate a single gene-level p-value, without the need to explicitly model correlation between tests or transcripts. XMAT can be applied in conjunction with any gene-based association test and can include tests that restrict variants to particular function categories (e.g. truncation, missense, and splice-altering variants). To demonstrate the utility of XMAT, we conducted case-control studies involving 783 TCGA breast cancer cases, 272 TCGA ovarian cancer cases, and 3,607 population controls of European ancestry. We evaluated 24 rare variant association tests with different combinations to compare the power of XMAT relative to each individual test. In general, we observed that the p-values in XMAT were comparable to the most powerful individual test, with equivalent power in most scenarios (e.g. *BRCA2* in breast cancer, *BRCA1* and *RAD51C* in ovarian cancer) and a modest attenuation of association signal in some scenarios (e.g. *BRCA2* in ovarian cancer). In all examples, XMAT outperformed any combination of individual tests after correcting for multiple testing. Our results demonstrate that XMAT is a statistically robust approach for incorporating multiple gene-based tests in rare variant association studies.

## PgmNr 1071: Computational strength of evidence for clinical sequence variant classification without circular reasoning.

### Authors:

S.V. Tavtigian <sup>1</sup>; C.C. Young <sup>1</sup>; B.J. Feng <sup>2</sup>; E. Ziv <sup>3</sup>; S.L. Neuhausen <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Oncological Sciences and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT.; 2) Department of Dermatology and Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 3) Division of General Internal Medicine, Department of Medicine and Institute for Human Genetics and Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, CA; 4) Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, CA

---

The American College of Medical Genetics (ACMG) sequence variant classification guidelines include two criteria based on computational evidence, 'PP3' and 'BP4'. These equate to supporting evidence in favor of and against pathogenicity, respectively and further equate to odds in favor of pathogenicity of 2.1-4.3 (or the reciprocal range, for BP4). While some investigators in the field believe that computational evidence should be able to provide more than 'supporting' evidence, rigorous tests of strength of evidence have been difficult to perform due to hidden multiple testing, circularities between sets of sequence variants used for creation and testing of computational tools, and because tools may opaquely leverage more than one ACMG evidence code.

As a proof case that avoids these obstacles, we evaluate *CHEK2* case-control mutation screening data with the computational tool Align-GVGD. A relatively simple tool created more than a decade ago, Align-GVGD was calibrated using *BRCA1/2* data and has no circular connections to *CHEK2*. As an initial step, we performed a pooled analysis of two *CHEK2* mutation screening studies: the first a multi-ethnic study with approximately 1,300 early onset breast cancer cases and ~1,100 ethnically matched controls first reported on in 2011, and the second a series of ~1,100 Latina early onset and/or familial cases and ~1,200 matched controls described this year.

Analyzing these case-control data, which include 56 distinct rare missense substitutions, we estimate odds in favor of pathogenicity without reference to existing *CHEK2* variant classifications, further avoiding logical circularities. Using the same ordered categorical groupings that are applied to *BRCA1/2* sequence variants (thereby reducing multiple testing), we found that: the expected benign Align-GVGD grade, C0, exceeds odds of benign effect required for BP4; the C15, C25 grade yields an indeterminate result; and the logically more pathogenic C35-C55 and C65 grades both exceed odds of pathogenicity required for PP3. There was no evidence of heterogeneity between the two studies.

We will report results from identical analyses applied to order of magnitude larger series of breast cancer cases and controls, allowing us to draw firm conclusions regarding the strength of evidence that can be achieved without concern about hidden circular reasoning and, in particular, whether ACMG 'moderate' evidence of pathogenicity is within reach of simple computational tools.

## PgmNr 1072: Prioritization of variants and genes in an exome study of familial lymphoid cancers.

### Authors:

S. Ralli <sup>1,2</sup>; S.J. Jones <sup>1,3</sup>; S. Leach <sup>1</sup>; A. Williams <sup>1</sup>; J.M. Connors <sup>4</sup>; A. Brooks-Wilson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer, Vancouver, British Columbia, Canada; 2) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Burnaby, British Columbia, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 4) Centre for Lymphoid Cancer, British Columbia Cancer, Vancouver, British Columbia, Canada

---

In lymphoid cancer affected families, members often develop different types or subtypes of lymphoid malignancies rather than developing the same disease. This implies presence of shared genetic factors that affect susceptibility to more than one type or subtype of lymphoid cancer. In large multicase families, genetic factors can be identified using methods such as linkage analysis. It remains a challenge, however, to detect susceptibility factors in small families with fewer cases. We propose a method to prioritize variants and genes in collections of lymphoid cancer families for where genetic heterogeneity is likely but biological commonalities are plausible.

We are applying the method to a collection of 185 multigenerational families, with 2-10 members with Hodgkin lymphoma, non-Hodgkin lymphoma, chronic lymphocytic leukemia or myeloma. Exome data from 89 affected individuals within 40 families will be used to discover and prioritize putative genetic variants. The remaining families will be used for further study of prioritized genes and variants.

The approach involves filtering step to retain genotypes that are shared among lymphoid affected individuals in each family. Shared genotypes are prioritized by weighting them for five criteria's: 1) the number of affected individuals in a family (family weight), 2) the percentage of the genome that is shared amongst affected individuals in a family (sharing weight), 3) the minor allele frequency of the variant (population allele frequency weight), 4) *in silico* prediction of variant function (prediction weight) and 5) a Surveillance, Epidemiology, and End Results (SEER) incidence rate that combines sex, age of lymphoid cancer onset, subtype of lymphoma and ethnicity (individual weight). These weights are normalized to the range of 0 to 1 and multiplied together to produce an overall weight for each variant in each family. The overall weights are combined to generate a multifamily weight for each variant. Genes with variants that have multifamily weights above a user-defined threshold are analyzed for enrichment at the level of the gene, biological process and pathway. Identifying genetic factors that contribute to susceptibility to lymphoid cancers will allow a better understanding of the biological processes underlying these malignancies. The weighting system proposed here for searching for biological commonalities can be used in other familial diseases with small families.

## PgmNr 1073: Whole genome bisulfite sequencing of paired primary and recurrent high grade serous ovarian cancer identifies epigenomic reprogramming and modes of clonal expansion that confer chemoresistance.

### Authors:

N. Gull<sup>1</sup>; M. Jones<sup>1</sup>; S. Coetzee<sup>1</sup>; A. Reyes<sup>1</sup>; J. Plummer<sup>1</sup>; B. Davis<sup>1</sup>; S. Chen<sup>1</sup>; J.P. Govindavari<sup>2</sup>; J. Rutgers<sup>5</sup>; K. Lawrenson<sup>1,3</sup>; D. Hazelett<sup>1</sup>; B. Karlan<sup>5</sup>; S. Gayther<sup>1,4</sup>; B.P. Berman<sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Bioinformatics and Functional Genomics, Department of Biomedical Science, Cedars-Sinai Medical Center, Los Angeles, CA, 90048; 2) Dept of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, 90048; 3) Women's Cancer Program, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048; 4) Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048; 5) Department of Obstetrics and Gynecology, David Geffen School of Medicine at UCLA, Los Angeles, CA; 6) Department of Developmental Biology and Cancer Research, Faculty of Medicine, The Hebrew University of Jerusalem, Ein Kerem

---

High grade serous ovarian carcinoma (HGSOC), the most aggressive and prevalent type of epithelial ovarian cancer, continues to be the leading cause of gynecological cancer-related deaths. Several factors such as late stage diagnosis and increased metastatic potential contribute to poor prognosis. Despite many patients responding positively to debulking surgery followed by platinum-chemotherapy, over 70% of patients will experience disease relapse and chemotherapy resistance will progress to fatal disease. Recent studies of methylation in primary HGSOC tumors using microarrays demonstrated widespread differences between chemo-resistant and chemo-responsive patients at the molecular level; for instance, BRCA1 promoter hypermethylation was linked to reduced patient survival. To identify genomic landscape changes in chemotherapy resistant HGSOC tumors, we used whole genome bisulfite sequencing (WGBS) and RNA-seq to profile paired primary and recurrent HGSOC tumors (n=73) from patients who have either BRCA1/2 mutations (n=12) or are BRCA wild type (n=24). Stark differences in global DNA methylation and transcriptome signatures were observed that could potentially explain the relapsing course of the disease. We identified 11,844 (q<0.05) differentially methylated regions (DMRs) between primary and recurrent patient samples; 3,873 of these DMRs overlapped between two or more patients. Using our novel computational method ELMER (Enhancer Linking by Methylation Expression Relationships) constrained by the DMRs, we identified genes known to be perturbed in ovarian cancer such as *RADX*, *AGR2*, and *CLDN2*, as well as other novel genes. Additionally, we identified 118 transcription factors hypomethylated in primary samples and 83 hypomethylated in recurrent samples (q<0.05, OR>1.1). Interestingly, we also observed changes in epigenetic domain structure, specifically deep PMD (partially methylated domain) hypomethylation in multiple recurrent samples indicating a tumor purity as well as a high level of proliferation. These results provide strong evidence for alterations in methylation that promote chemotherapy resistance and provide a better understanding of development and a basis for therapeutic decisions and interventions. Our WGBS data represent the first whole methylome analysis of HGSOC patients to date and the combinatorial approach has provided a genome-wide cis-

regulatory epigenetic landscape in ovarian cancer as it acquires and maintains chemoresistance.



## PgmNr 1074: Targeting noncoding RNAs can improve the effectiveness of radiation therapy for cancer patients.

**Authors:**

H.S. Chiu; S. Somvanshi; E. Patel; P. Sumazin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA

---

Profiling of many tumor types revealed dramatic upregulation of long-noncoding RNAs (lncRNAs). These lncRNAs are promising therapeutic targets because inhibitors that specifically target them may be able to target cancer cells while having minimal toxicity in non-tumor cells. We developed software—including longhorn (Chiu *et al.*, 2018) and bighorn (Chiu *et al.*, 2019)—to infer disease-relevant functions for lncRNAs. Using these tools, in conjunction with data from hundreds of cell lines and thousands of patients, we identified lncRNAs that target DNA-repair pathways and are predicted to improve the effectiveness of x-ray and proton therapies. Expression and genomic alterations at the loci of these lncRNAs predict cell and patient response to radiation therapies. Focusing on pan-cancer lncRNAs, which have high expression in many cancer types, we showed that silencing these lncRNAs *in vitro* increases the susceptibility of cell lines to radiation and reduces radiation dosages required for killing cancer cells. We argue that tumor-suppressor lncRNAs that regulate the DNA repair pathway can indicate candidates for lower-dosage radiation therapy and that pan-cancer and cancer-specific lncRNAs that are upregulated in tumor cells and are predicted to disrupt DNA repair pathways are excellent candidate therapeutic biomarkers.

## PgmNr 1075: DiNeR: A nonparametric *Differential* graphical model of *Network Rewiring* to infer transcription factor co-binding alterations.

### Authors:

J. Liu <sup>1,2</sup>; J. Zhang <sup>1,2</sup>; D. Lee <sup>1,2</sup>; Y. Yan <sup>1,2</sup>; S. Lou <sup>1,2</sup>; G. Gursoy <sup>1,2</sup>; M. Gerstein <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520; 2) Computational Biology and Bioinformatics Program, Yale University, New Haven, CT 06520; 3) Department of Computer Science, Yale University, New Haven, CT 06520

---

During transcription, numerous transcription factors (TFs) are recruited in a highly coordinated manner to control the temporal and spatial expression pattern of genes. TF-target edge gain or loss in the regulatory network, also known as network rewiring, can change the co-regulation of TFs, which in turn can potentially introduce expression changes and phenotypic variations. However, quantification of changes in TF co-binding relationships between normal and disease states has not been comprehensively studied. To address this problem, we propose a computational method DiNeR to directly construct the differential TF-TF regulatory network from paired disease-to-normal ChIP-seq data. Specifically, we assembled genome-wide binding profiles of 104 common factors in K562 and GM12878 human cell lines. We then applied a non-parametric differential graphical model to estimate the TF-TF co-binding changes between chronic myeloid leukemia (CML) and normal samples. We further applied network stability analysis by sub-sampling the network to control the sparsity in the differential network. As a result, we identified 351 significantly altered edges during oncogenesis. Specifically, we found that in leukemia the co-binding of the tumor suppressor BRCA1 and RNA polymerase II, a well-known transcriptional pair in healthy cells, was disrupted in tumor cells. Further extraction of hub regulators and the discovery of well-known CML risk genes demonstrates the effectiveness of our DiNeR method in highlighting key regulators of gene expression in disease.

## PgmNr 1076: Internal normalization of RNA sequencing data using genes with stable expression.

### Authors:

B. Avery <sup>1</sup>; M.J. Madsen <sup>1</sup>; N.J. Camp <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT.; 2) School of Medicine, University of Utah, Salt Lake City, UT

---

Cancers are heterogeneous and can be difficult to characterize from any single perspective. Multi-dimensional characterization of tumors has the potential to improve our ability to classify tumors, identify germline genetic risk, and associate with clinical outcomes. Our goal is to identify biologically relevant quantitative tumor traits using transcriptome data.

Previously we used principal components analysis of the expression of PAM50 genes to identify breast tumor dimensions. We illustrated utility of these for gene mapping, prediction of prognosis, and response to paclitaxel (Madsen et al. 2018, Camp et al. 2019). The potential impact of applying this approach more broadly is high. The PAM50 panel, however, is a small set of highly curated genes. To expand the same concept to whole transcriptome RNA sequencing data will require normalization for gene size, library size, and RNA composition; the latter particularly challenging. Many normalization procedures for RNA composition exist, however, most are targeted to differential gene expression applications and implement a reference to remain robust. Our goal is a reference-free method, which is important to allow tumor traits to be calculated in a single sample and compared across different studies. To achieve this, we have developed a procedure for 'internal normalization' that uses STABLE (STeady And reliABLy Expressed) genes.

We identified STABLE genes using all samples in the TCGA and GTEx datasets. In each resource, we eliminated genes that had zero expression in any sample. The remaining genes were naively corrected for library size, and ranked by coefficient of variation within each dataset (least to most variation). Rank aggregation across TCGA and GTEx was used to identify 101 genes with expression stability across both resources; the STABLE genes. In an independent dataset, we calculate a sample-based size factor based on the median expression level of the 101 STABLE genes from within each sample. Gene counts for each sample were adjusted by gene size, library size, and STABLE size factor.

To test the effectiveness of our method, we constructed several datasets engineered to contain issues of library size and RNA composition, and normalized using established methods and our novel procedure. Internal normalization using STABLE genes performed as well or better than other methods. This is a significant step toward the development of universal tumor traits using transcriptome data.

## PgmNr 1077: Inference and characterization of copy number variants from bulk and single cell RNA-sequencing datasets.

**Authors:**

A. Harmanci; A.S. Harmanci; X. Zhou

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Biomedical Informatics, University of Texas Health Science Center, Houston, HOUSTON.

---

RNA sequencing has become the de facto standard for measuring the expression levels of genes. Each RNA sequencing experiment generates large amounts and spectrum of information about expression levels of genes including expression levels of splice variants, gene fusions, and novel non-coding and intergenic transcripts. With the recent advancement in the single cell technologies, single cell RNA-sequencing is now utilized to deconvolve complex tissues and organs with respect to changing expression patterns in development and disease. Although they are mainly used for quantifying expression levels, they contain much more biologically relevant information about the genetic variants. For example, well-known variant detection tools such as the genome analysis toolkit (GATK) has semi-standardized pipelines for detecting single nucleotide variants and small insertions and deletions from the RNA-seq datasets with decent accuracy. Among these variants, detectability of structural variants and copy number variants (CNVs) from RNA-sequencing data is not well-studied. We present CaSpER, a signal processing approach for identification, visualization, and integrative analysis of focal and large-scale CNV events in multiscale resolution using either bulk or single-cell RNA sequencing data. CaSpER utilizes a novel smoothing of the genomewide RNA sequencing signal profiles in different multiscale resolutions, identifying CNV events at different length scales. CaSpER next employs a novel methodology for generation of genomewide allelic shift signal profile from the reads and utilizes it in multiscale fashion for correction of CNVs calls. The allelic shift signal is used to quantify the loss-of-heterozygosity, LOH, that is valuable for CNV identification. The multiscale nature of CaSpER enables comprehensive analysis of focal and large-scale CNVs and LOH segments. CaSpER performs well in accuracy compared to gold standard SNP genotyping arrays. In particular, analysis of single cell GBM RNA sequencing data with CaSpER reveals novel mutually exclusive and co-occurring CNV sub-clones at different length scales. CaSpER increases the utility of RNA-sequencing datasets and complements other tools for complete characterization and visualization of the genomic and transcriptomic landscape of single cell and bulk RNA sequencing data, especially in cancer research.

## PgmNr 1078: Single cell cancer heterogeneity: Clonotype detection in renal cell carcinoma.

### Authors:

E.I. Velazquez Villarreal; D.W. Craig; J.D. Carpten

[View Session](#) [Add to Schedule](#)

**Affiliation:** Translational Genomics, University of Southern California, Los Angeles, California.

---

Among urological malignancies, one of the most lethal and responsible for around 80 percent of all primary renal neoplasms is the renal cell carcinoma (RCC). The single cell sequencing have the potential to provide a new insight about RCC cancer heterogeneity since clonotype and variant detection at the highest resolution. In order to study the extensive cancer heterogeneity of RCC, we used single cell sequencing copy-number analysis and clonotype detection in four RCC tumors. Single cell 10x Chromium<sup>TM</sup> Technology was used to reach detection of 100 Kb CNV events and calling clonotypes down to 10 of 1000 cell inputs. By processing four samples, one resulted in around 500 total with tumor cells more than 50% of the total cells, other samples had lower percentage. Variable media ploidies were calculated through genome-wide ploidy analysis of the tumor cells. We used 20 Kb cases of reads to estimate regional copy number. Different sub-clones were identified among samples. In sample number one four clusters of sub-clones characterized cancer heterogeneity. Genome-wide ploidy analysis of the tumor cells showed variable median ploidies. The regional copy number was estimated by processing our data in 20 Kb cases of reads. By applying Fast maximum-likelihood (ML) genetic clustering and Bayesian Information Criterion (BIC) it was detected different sub-clones in each sample; in sample number one, there where four clusters of sub-clones characterized cancer heterogeneity which also showed copy number changes on entire chromosome arms and mutations at variant detection level, other samples resulted with different number of clusters. Among cells from sample one, it was detected *VHL* gene mutations that have been reported in RCC samples as signatures of clinical prognosis. Genomic features were detected since the highest resolution applied using this single cell technology. This highest resolution allow the identification and stratification of sub-clones as a new alternative to understand the extensive heterogeneity that characterize RCC.

## PgmNr 1079: SCOPE: A normalization and copy number estimation method for single-cell DNA sequencing.

### Authors:

R. Wang<sup>1</sup>; D.Y. Lin<sup>1,2</sup>; Y. Jiang<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 2) Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC; 3) Department of Genetics, School of Medicine, University of North Carolina, Chapel Hill, NC

---

**INTRODUCTION:** Whole-genome single-cell DNA sequencing (scDNA-seq) enables characterization of copy number profiles at the cellular level. This increases resolution and decreases ambiguity in assessing tumor heterogeneity and tracking cancer evolutionary history. ScDNA-seq data is, however, sparse, noisy, and highly variable even within a homogeneous cell population due to the biases and artifacts that are introduced during library preparation and sequencing procedure. Existing methods either build an optimized normal/reference set for normalization or adopt a cell-specific normalization procedure for removing biases, followed by a post hoc procedure for ploidy estimation and adjustment. They cannot adequately remove such artifacts and, furthermore, they do not address the challenges and complexities caused by aberrant copy number changes or the complicating factor of tumor ploidy.

**METHODS:** SCOPE integrates both null and non-null genomic regions for unbiased estimation of both GC content bias and latent factors for cell- and region-specific background correction. The distinguishing features of SCOPE include: (i) utilization of cell specific Gini coefficients for quality control and for identification of normal cells, which are then used as negative control samples in a Poisson latent factor model for read depth normalization; (ii) modeling of GC content bias using an expectation-maximization algorithm embedded in the Poisson regression models to account for the different copy number states along the genome; and (iii) a cross-sample iterative segmentation procedure to identify breakpoints that are shared across cells with the same genetic background.

**RESULTS:** SCOPE is shown to more accurately estimate subclonal copy number aberrations and to have higher correlation with array-based copy number profiles of purified bulk samples from the same breast cancer patient. Ploidy estimates by SCOPE are highly concordant with those from previous reports based on single-cell flow sorting. We further show that the copy number profiles returned by SCOPE can also be recapitulated by whole-exome sequencing and single-cell RNA-sequencing. We finally demonstrate SCOPE on scDNA-seq data that was produced using the 10X Genomics single-cell CNV pipeline, showing that it can reliably recover proportions of the cancer cell spike-ins from a background of normal cells and that it can successfully reconstruct cancer subclonal structure across 10,000 breast cancer cells.

## PgmNr 1080: Sparse conditional generate adversarial networks (CGAN) for individualized biomarker selection and treatment effect estimation.

### Authors:

S. Fang <sup>1</sup>; S. Guo <sup>2</sup>; Y. Liu <sup>3</sup>; Q. Ge <sup>3</sup>; S. Schrodi <sup>2,4</sup>; M. Xiong <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Univ Texas MD Anderson Cancer Center, Houston, Texas.; 2) Marshfield Clinic Research Institute, Marshfield, WI; 3) University of Texas School of Public Health, Houston, Texas; 4) University of Wisconsin-Madison, WI

---

Next generation genomic, epigenomic, sensing and image technologies produce ever deeper multiple omics, physiological, imaging and phenotypic data with millions of features. Integrating omics, physiological and imaging data provides invaluable information for identification of individualized biomarkers that will be used for estimation of individualized treatment effects and optimal selection of individualized therapy. The classical methods for biomarker identification use average treatment effect information. However, treatment response is heterogeneous. Only using average treatment effect information presents a problem in selecting the optimal treatment for each individual to ensure that the right therapy is offered to “The right patient at the right time.” Unfortunately, estimating individualized treatment effects and design of individualized therapy is beyond the state-of-the-art of the current biomarker selection paradigm. A key issue for individualized treatment estimation is to estimate counterfactuals of treatment. However, counterfactuals are unobserved and are therefore a missing value problem. The classical treatment effect estimation methods cannot accurately estimate the counterfactuals due to lack of methods for missing value estimation. The recently developed conditional generative adversarial nets (CGAN) are accurate tools for estimating the counterfactuals. Imaging, omics and physiological data involve millions of features. Since neural networks (NN) are complicated nonlinear functions, identifying biomarkers from omics and imaging data in CGAN is a challenging task. To identify individualized biomarkers and estimate individualized treatment effects, we combined a novel instance-wise feature selection method that consists of three neural networks: a selection network, a prediction network and baseline network, and CGAN. The algorithms are applied to the TCGA Uterine Corpus Endometrial Carcinoma (UCEC) dataset with histological imaging, gene expression (mRNA and miRNA), methylation data and 145 chemotherapy response records from 67 unique individuals. Real data analysis results show that the proposed algorithms substantially outperform the state-of-the-art methods. We find our prediction model has strong predictive performance on mislabeled drug response (N=134) compared with overall survival (OS).

## PgmNr 1081: Allele-specific QTL fine-mapping with PLASMA.

### Authors:

A.T. Wang <sup>1,2</sup>; A.H. Shetty <sup>2,3</sup>; E. O'Connor <sup>2</sup>; C. Bell <sup>2</sup>; M.M. Pomerantz <sup>2</sup>; M.L. Freedman <sup>2</sup>; A. Gusev <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Massachusetts Institute of Technology, Cambridge, MA.; 2) Dana Farber Cancer Institute, Boston, MA.; 3) Cambridge University, Cambridge, UK

---

Although quantitative trait locus (QTL) associations have been identified for many molecular traits such as gene expression, it remains challenging to distinguish the causal nucleotide from nearby variants. In addition to traditional QTLs by association, allele-specific (AS) QTLs are a powerful measure of cis-regulation that are largely concordant with traditional QTLs, and can be less susceptible to technical/environmental noise. However, existing asQTL analysis methods do not produce probabilities of causality for each marker, and do not take into account correlations among markers at a locus in linkage disequilibrium (LD). We introduce PLASMA (PopuLation Allele-Specific Mapping), a novel, LD-aware method that integrates QTL and asQTL information to fine-map causal regulatory variants while drawing power from both the number of individuals and the number of allelic reads per individual. We demonstrate through simulations that PLASMA successfully detects causal variants over a wide range of genetic architectures, with upwards of a 10-fold reduction in credible set size over conventional QTL-based fine-mapping on the same data. We apply PLASMA to RNA-Seq data from 524 kidney tumor samples and show that over 13 percent of loci can be fine-mapped to within 5 causal variants, compared less than 2 percent of loci using existing QTL-based fine-mapping.

**PLASMA furthermore achieves a greater power at 50 samples than conventional QTL fine-mapping does at over 500 samples.** Overall, PLASMA achieves a 6.4-fold reduction in median 95% credible set size compared to existing QTL-based fine-mapping. We additionally apply PLASMA to H3K27AC ChIP-Seq from 28 prostate tumor/normal samples and demonstrate that PLASMA is able to prioritize markers even at small samples, with PLASMA achieving a 1.4-fold reduction in median 95% credible set sizes over existing QTL-based fine-mapping. Variants in the PLASMA credible sets for RNA-Seq and ChIP-Seq were enriched for open chromatin and chromatin looping (respectively) at a comparable or greater degree than credible variants from existing methods, while containing far fewer markers. **Our results demonstrate how integrating AS activity can substantially improve the detection of causal variants from existing molecular data and at low sample size.** More generally, PLASMA's approach in combining QTL and AS signals can potentially improve QTL-GWAS colocalization and gene expression prediction in transcriptome-wide association studies (TWAS).



## PgmNr 1082: Measures of transcript isoform diversity and alternative splicing.

### Authors:

S. Mount <sup>1</sup>; D.R. Crawford <sup>1, 2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Cell Biology and Molecular Genetics, Univ Maryland, College Park, MD.; 2) Cancer Data Science Laboratory, National Cancer Institute, National Institutes of Health, Bethesda, MD

---

The advent of high-throughput sequencing of messenger RNA, coupled with the development of software that allows the high-throughput processing of thousands of samples, makes possible a global description of alternative splicing at the level of a sample, a tissue or a species. We review global measures of diversity and define heteroformity. This measure of transcript diversity reflects the fraction of transcript pairs drawn at random from a single gene that differ. This metric is inspired by the utility of mean heterozygosity in population genetics. The heteroformity of a gene thus varies between 0 (if there is only one isoform) and 1 (the limit, when so many isoforms are present that each transcript is different). The abundance-weighted gene heteroformity of a sample can be visualized as a cumulative distribution, allowing immediate inference of some phenomena.

Application to 11,688 human samples from 30 tissues the Genotype-Tissue Expression (GTEx) project reveals some general patterns. In all samples, about 25% of transcripts are in genes with very low heteroformity. In contrast, the top quartile of transcripts are in genes with over 0.5 heteroformity. Tissues differ significantly; much greater levels of alternative splicing are consistently observed in reproductive and nervous tissue. There is also great individual variation, with specific heart samples varying over threefold in total gene-level heteroformity. We note that sample heteroformity need not correlate with differences in alternative splicing between samples, and that partitioning genes with respect to their overall heteroformity vs. their differential alternative splicing can reveal patterns of developmental regulation.

We are currently applying heteroformity to diverse biological samples and exploring its properties as a robust and useful metric, with a focus on the heteroformity of cancer RNA-seq samples, oncogenic isoforms and tumor-specific antigens generated by alternative splicing.

## PgmNr 1083: Clinical, biochemical, and pathological investigation of a patient with musculocontractural Ehlers-Danlos syndrome caused by a novel pathogenic variant in *DSE*.

### Authors:

M. Minatogawa <sup>1,2</sup>; S. Mizumoto <sup>3</sup>; T. Hirose <sup>4</sup>; T. Yamaguchi <sup>1,2</sup>; K. Wakui <sup>1,2</sup>; T. Watanabe <sup>4</sup>; S. Yamada <sup>3</sup>; T. Kohosho <sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Shinshu University, Matsumoto, Nagano, Japan; 2) Center for Medical Genetics, Shinshu University Hospital, Matsumoto, Japan; 3) Department of Pathobiochemistry, Faculty of Pharmacy, Meijo University, Nagoya, Japan; 4) Laboratory of Anatomy, School of Veterinary Medicine, Rakuno Gakuen University, Ebetsu, Japan; 5) Research Center for Supports to Advanced Science, Matsumoto, Japan

---

Musculocontractural Ehlers-Danlos Syndrome (mcEDS) is a recently identified connective tissue disorder, caused by loss-of-function variants in *CHST14* encoding dermatan 4-*O*-sulfotransferase-1 (mcEDS-*CHST14*) or in *DSE* encoding dermatan sulfate epimerase (mcEDS-*DSE*), both resulting in defective biosynthesis of dermatan sulfate (DS) replaced by chondroitin sulfate (CS). Whereas 42 patients with mcEDS-*CHST14* from 28 families have been reported to date, only three patients from two families with mcEDS-*DSE* have been published. Here, we report clinical, biochemical, and pathological features in the fourth case with mcEDS-*DSE* caused by a novel biallelic pathogenic variant of *DSE*. The patient is a 36-year-old Japanese man born to consanguineous parents who were second cousin. He had recurrent large subcutaneous hematomas from childhood, sometimes requiring intravenous administration of DDAVP. He also showed adducted thumbs, delayed anterior fontanelle closure, slender and cylindrical fingers, scoliosis, bilateral cryptorchidism, bilateral high intraocular pressure and thin retina, mitral valve regurgitation, and a massive diverticular hemorrhage. He had no developmental delay and had three healthy daughters. Next-generation-based custom panel sequencing for hereditary connective tissue disorders identified a novel homozygous variant "c.2601C>A; p.Y867\*" in *DSE*. Epimerase assays, measured using COS-7 cells transfected with a *DSE* sequence including the detected variant, showed a significantly lower, an almost negligible level of epimerase activity, compared with the cells with a wild-type *DSE* sequence. A disaccharide composition analysis of DS and CS showed a negligible amount of DS in the patient's urine, and showed a small but detectable amount of DS, significantly lower than healthy controls though somewhat higher than those with mcEDS-*CHST14*. Transmission electron micrographs (TEM) including cupromeronic blue staining to visualize GAG chains on the skin specimens showed unassembled collagen fibrils with GAG chains linearly stretching from the attached collagen fibrils in some areas, similar to the findings in patients with mcEDS-*CHST14*, but in other areas, tightly packed collagen fibrils curved in close contact along the contour of attached collagen fibrils, similar to the findings in healthy controls. Milder clinical phenotypes and partial biochemical and pathological abnormalities suggest an alternate pathway producing DS in the absent activity of DSE.

## PgmNr 1084: “You can't always get what you want”: Pseudoscleroderma unraveling Hutchinson-Gilford progeria.

### Authors:

A. Pessoni <sup>1</sup>; F. Borges <sup>1</sup>; Z. Carneiro <sup>1</sup>; L. Vagnini <sup>2</sup>; C. Lourenco <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centro Universitario Estacio, Ribeirao Preto, Brazil; 2) Centro Paulista de Diagnostico e Pesquisa - CPDP, Ribeirao Preto, Brazil

---

**BACKGROUND:** Hutchinson–Gilford progeria syndrome (HGPS) was first reported in 1886 as a case of “congenital absence of hair and its appendages,” and, later, Gilford introduced the term progeria in 1904. Although signs and symptoms vary in terms of the age of onset and severity, they are remarkably consistent overall, involving skin, bones, heart, and blood vessels.

**CASE REPORT:** Female patient, born to non-consanguineous parents, was referred for failure to thrive and suspicion of infantile scleroderma with 4 months of age. On clinical examination, there was a markedly thickened, tight skin, especially on the trunk, buttocks and proximal thighs. The skin felt indurated, was shiny and had a ridged appearance on the thighs and a ‘cobblestone’ aspect on the buttocks. There were no other dysmorphic features observed. Her skin biopsy was suggestive of an “atypical scleroderma”, and her investigation was guided to evaluate genetic metabolic disorders with cutaneous manifestations like congenital disorders of glycosylation. Comprehensive diagnostic testing was performed (with brain MRI, echocardiogram, abdominal ultrasound, screening for inborn errors of metabolism) and turned out to be normal. Since there were reports of scleroderma-like lesions in HGPS, molecular genetic testing was realized and patient showed a de novo heterozygous c.1824C>T (G608G) mutation in exon 11 of the LMNA gene on chromosome 1q22.

**DISCUSSION:** HGPS is a rare, uniformly fatal premature aging disease with distinct dermatologic features. Some studies show that skin and hair abnormalities are present within the first 24 months of life in such patients. Usually, sclerodermoid changes are the most frequently reported skin feature, which commonly involved the abdomen and bilateral lower extremities. Progressively, patients present with other typical HGPS findings: prominent superficial veins, dyspigmentation, and alopecia.

**CONCLUSION:** HGPS should be considered a differential diagnosis in children with early onset findings of scleroderma/scleroderma-like lesions. Since those skin findings are very distinct cutaneous manifestations appearing during the first 2 years of life in this disease, awareness of these findings allow earlier diagnosis of HGPS and adequate patient clinical management.

## PgmNr 1085: Congenital contractural arachnodactyly and short stature in a boy with a *de novo* **FBN2** missense variant.

### Authors:

N. Urraca <sup>1</sup>; S. Holtrop <sup>2</sup>; C.W. Brown <sup>1</sup>; H. Lydigsen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatric Genetics Division, University of Tennessee Health Science Center; 2) Le Bonheur Children's Hospital, Memphis, TN

---

Congenital Contractural Arachnodactyly (CCA), also known as Beals syndrome, is a rare connective tissue disorder. Patients with CCA have a Marfanoid habitus, as patients are tall, thin, and have arachnodactyly as well as scoliosis, congenital contractures, muscle hypoplasia and abnormalities of the external ears. Here we present a 20-month-old male referred to genetics because of short stature. Parents are healthy, of normal height, and non-consanguineous. He was born at 39 weeks by a planned C-section. His birth weight was 3.7 kg (75%ile) and length, 49 cm (30%ile). During his initial genetics evaluation at age 8-months, parents reported that he has never been able to extend his legs completely. His physical examination revealed a height <1%ile, relative macrocephaly (75%ile), frontal bossing, ears with normal position and appearance, micrognathia, camptodactyly, extra creases on the forearms, and limitation to full extension of elbows and knees bilaterally. A skeletal survey showed slight anterior, inferior beaking of L1-L2 vertebrae and urine oligosaccharides and mucopolysaccharides excluded mucopolipidoses and mucopolysaccharidoses. A chromosomal microarray and *ADAMTSL2* and *FBN1* sequencing were recommended to assess for possible chromosomal anomalies and geleophysic dysplasia, respectively but were denied by insurance. At age 15-months, he had progression of his joint restriction and underwent trio whole genome sequencing. A heterozygous, *de novo* missense variant (c.44863C>G;p.Cys1621Trp) was identified in *FBN2*. This was classified as a variant of uncertain significance, as there is phenotypic overlap between this patient's features and CCA; however, short stature has not been described in CCA and the variant has never been reported. We suggest that our patient's frontal bossing, micrognathia, camptodactyly, and joint contractures are clinically consistent with the diagnosis of CCA, and that the short stature could be a new feature of the phenotype. CCA is a rare disorder and the description of additional patients will be required to determine if short stature is within the spectrum of the disorder.

## PgmNr 1086: Probing a novel familial lipomatosis syndrome with whole genome sequencing.

### Authors:

J. de Vegvar <sup>1</sup>; K. Yeboa <sup>1</sup>; D. Goldstein <sup>1</sup>; H. Depaz <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Columbia University, Genetics; 2) Harlem Hospital, Surgery

---

**Introduction:** Lipomas are slow-growing adipose tumors that occur more frequently in mature adults aged 40 to 60 years old. Most lipomas are small and weigh only several grams. A giant lipoma is defined as at least 10 cm in one dimension and weighs at least 100 grams. Cases of giant lipomas have been reported, but a clinical syndrome of a solitary giant encapsulated lipoma with a strong familial component has not yet been described. **Goals and hypothesis:** We studied a large cohort of 22 individuals from a St. Croix family including a 42-year-old man born in St. Croix with macrocephaly, intellectual disability and a giant lipoma. This same condition has been traced through three generations of the patient's family including his grandmother, two aunts, uncle and cousin. We hypothesized that this condition is caused by an unknown hereditary factor. **Methods:** We employed chromosomal microarray, karyotype, and whole exome sequencing of the proband. We then captured genomes on four of the family members (3 affected and one unaffected) and performed sequencing analysis by a next generation sequencing platform with a mean coverage of 40 and sufficient depth to call variants at 99% of each targeted genome. **Results:** A disease causing factor was identified as being related to the phenotype of this family. **Conclusion:** Our results demonstrate a successful application of whole genome sequencing to identify a gene for an undescribed autosomal dominant disorder.

## **PgmNr 1087: Autonomic symptoms and autoimmune dysautonomia evaluation in patients with hypermobile Ehlers Danlos syndrome (hEDS) and generalized hypermobility on the EDS spectrum (G-HSD).**

### **Authors:**

J. Huynh <sup>1</sup>; R. Jesudas <sup>2</sup>; K. Saboda <sup>3</sup>; K. Diehl Martinez <sup>4</sup>; A. Sanoja <sup>1</sup>; C. Mauss <sup>3,5</sup>; P. Byers <sup>6</sup>; C. Laukaitis <sup>2,5</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) College of Medicine, Department of Medicine, University of Arizona, Tucson, AZ.; 2) Bleeding & Clotting Disorders Institute, Peoria, IL; 3) Arizona Cancer Center; 4) Genetics Graduate Interdisciplinary Program, University of Arizona, Tucson, AZ; 5) University of Arizona, Tucson, AZ; 6) Departments of Pathology and Medicine, University of Washington, Seattle, WA

---

Hypermobile Ehlers Danlos Syndrome (hEDS) is one of 13 EDS types that is characterized by joint hypermobility. However, its physical symptoms can be similar to patients who have generalized hypermobility on the EDS hypermobility spectrum (G-HSD). The hEDS GENE study was developed to better understand the genetic causes of hEDS and to gather data on disease-related symptomatology in order to develop better hEDS diagnostic methods.

To quantify disease-related symptomatology, each participant in the study completed eight questionnaires, including the COMPASS31. This survey is an abbreviated version of the Autonomic Symptom Profile and prompts participants to answer questions on autonomic symptoms that are grouped into the following subgroups: orthostatic intolerance, vasomotor, secretomotor, gastrointestinal, bladder and pupillomotor symptoms. Totalled scores range between 0-100 with higher scores indicating greater autonomic symptoms. Total weighted scores for participants with hEDS and G-HSD (42.5) are significantly different from unaffected participants (9.4) ( $p < 0.001$ ). However, participants with hEDS and G-HSD did not significantly differ in their subgroup scores except for the GI ( $p < 0.001$ ) and Vasomotor ( $p < 0.002$ ) categories.

Additionally, serum from 40 participants was sent to Mayo Clinical labs and tested with the Autoimmune Dysautonomia Evaluation (DYS1) which tests for the presence of nine antibodies. 17.5% (7/40) of participants had measurable auto-antibodies sometimes associated with dysautonomia. Four had GAD65 levels from 0.03 to 0.06 nmol/L (normal  $\leq 0.02$ ), two had Neuronal (V-G) K<sup>+</sup> Channel antibody levels from 0.09 to 0.1 nmol/L (normal  $\leq 0.02$ ), and one had a positive striated muscle antibody titer at 1:960 (normal  $< 1:120$ ). However, none of the participants tested had measurable levels of the ganglionic acetylcholine receptor (alpha3) autoantibody which is the only antibody on the panel known to cause autoimmune dysautonomia.

Our studies suggest that though patients with hEDS or G-HSD experience greater autonomic symptoms compared to unaffected patients, patients with either hEDS or G-HSD experience similarities in their autonomic symptoms. Though 7/40 patients were found to be positive upon autoimmune dysautonomia antibody evaluation, there is not a high prevalence of antibodies or a clear correlation between antibody presence and symptoms.

## PgmNr 1088: A new mutational hotspot in the *SKI* gene in the context of MFS/TAA molecular diagnosis.

### Authors:

P. Arnaud <sup>1,2,3</sup>; C. Racine <sup>4</sup>; N. Hanna <sup>1,2,3</sup>; J. Thevenon <sup>5</sup>; J.-L. Alessandri <sup>6</sup>; J. Clayton-Smith <sup>7</sup>; C. Coubes <sup>8</sup>; S. Dupuis-Girod <sup>9</sup>; S. Odent <sup>10</sup>; V. Carmignac <sup>11</sup>; C. Thauvin-Robinet <sup>4,11</sup>; C. Le Goff <sup>2</sup>; G. Jondeau <sup>2,3</sup>; C. Boileau <sup>1,2,3</sup>; L. Faivre <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Département de Génétique, AP-HP, Hôpital Bichat, Paris, Paris, France; 2) LVTS, INSERM U1148, Hôpital Bichat, Université Paris Diderot, Paris, France; 3) Centre de Référence Maladies Rares, Syndrome de Marfan et pathologies apparentées, Hôpital Bichat, AP-HP, Paris, France; 4) Centre de Génétique et FHU TRANSLAD, Centre Hospitalier Universitaire Dijon, Dijon, France; 5) Unité de Génétique Clinique, CHU de Grenoble site Nord - Hôpital Couple-Enfant, Dijon, France; 6) Pôle enfants, CHU de la Réunion - Hôpital Félix Guyon, Saint-Denis, France; 7) Manchester Center for Genomic Medicine, Manchester Regional Genetics Service, Genetic Medicine, St Mary's Hospital, Manchester, UK; 8) Équipe Maladies Génétiques de l'Enfant et de l'Adulte, Département Génétique Médicale, Maladies rares et médecine personnalisée, CHRU de Montpellier - Hôpital Arnaud de Villeneuve, Montpellier, France; 9) Hospices Civils de Lyon, Hôpital Femme-Mère-Enfant, Service de Génétique Clinique, Centre Hospitalier Universitaire de Lyon, Bron, France; 10) Hôpital Sud, Service de Génétique Clinique, Centre Hospitalier Universitaire de Rennes, Rennes, France; 11) Laboratoire de génétique chromosomique et moléculaire, CHU de Dijon - Dijon, France

### Background

*SKI* pathogenic variations are associated with Shprintzen-Goldberg Syndrome (SGS), which is a rare systemic connective tissue disorder characterized by craniofacial, skeletal and cardiovascular features. Intellectual disability and craniosynostosis are almost consistent in the carriers of pathogenic variants in this gene. So far, the clinical description was quite homogeneous within patients, and the known pathogenic variations were located in two different hotspots of the *SKI* gene.

### Methods and Results

In the course of diagnosis of Marfan syndrome (MFS) and related disorders, we identified 6 probands (2 females and 4 males) carrying 3 different pathogenic variants in the *SKI* gene affecting the same amino acid Thr180 located in the Dachshund-homology domain (DHD). They were all sporadic cases, diagnosed at an age ranging from 2 to 14 y.o. For 5 of the probands, the family screening revealed that the molecular events occurred *de novo*. A specific clinical examination was done for all these patients by a multidisciplinary team. They all displayed a milder morphological phenotype that did not initially lead to the clinical diagnosis of SGS. They presented a marfanoid habitus with a systemic score according to the revised Ghent nosology ranging from 4 to 12. Only 3 of them had learning disorders or mild intellectual disability. Four out of 6 presented a moderate cardiovascular involvement with Thoracic Aortic Aneurysm (TAA) or Mitral Valve Prolapse (MVP).

### Conclusion

This report extends the phenotypic spectrum of variants identified in the *SKI* gene. We describe a new mutational hotspot in the *SKI* gene associated with a marfanoid syndrome with no or a mild form of

intellectual disability. A cardiovascular involvement was seen in a significant number of cases, illustrating the need to adapt the medical treatment and follow-up for these patients.

**Keywords:** *SKI* gene, Shprintzen-Goldberg syndrome, Marfan syndrome, Phenotypic spectrum



## PgmNr 1089: Detection of a novel large deletion in *FBN1* gene by MLPA.

### Authors:

J. Bene<sup>1,2</sup>; G. Büki<sup>1</sup>; L. Pintér<sup>1</sup>; B. Melegh<sup>1,2</sup>; K. Hadzsiev<sup>1,2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Medical School, University of Pécs, Pécs, Hungary; 2) Szentagothai Research Centre, University of Pécs, Pécs, Hungary

---

**Objectives:** Marfan syndrome (MFS) is an autosomal dominant multisystem disorder of the connective tissue. It has varying clinical features, most frequently the skeletal, ocular, and cardiovascular systems are involved. The disorder caused by mutations in the fibrillin-1 gene (*FBN1*), mapped to chromosome 15q21.1. Majority of classic MFS patients bear small alterations (i.e. point mutations, indels), however, a smaller number of patients have larger genomic rearrangements in *FBN1* gene.

**Methods:** The study included 20 patients. All of the patients fulfilled the revised Ghent criteria. After negative results of systematic sequencing of *FBN1*, *TGFBR1* and *TGFBR2* genes MLPA (P065-P066 MRC-Holland) analyses were performed in our patient cohort.

**Results:** In contrast to the generally detected large deletions, we identified a novel two-exon deletion (exon 46-47) in two patients, who are close relatives. This deletion results in the loss of 31-32<sup>nd</sup> calcium binding EGF-like domain which is responsible for the development of classic Marfan phenotype.

**Conclusions:** Less than 10% of the disease causing mutations in *FBN1* gene are copy number alterations, in which single or multiple exon deletions can be detected by MLPA in a cost-effective manner even in the NGS era. Our findings emphasize the importance of screening for large *FBN1* deletions in clinical genetic testing, especially for those with classic Marfan phenotype.

## PgmNr 1090: Otological features in patients with musculocontractural Ehlers-Danlos syndrome caused by mutations in *CHST14* (mcEDS-*CHST14*).

### Authors:

T. Kosho <sup>1,2,3</sup>; M. Kawakita <sup>4</sup>; S. Iwasaki <sup>5,6</sup>; H. Moteki <sup>5</sup>; S. Usami <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 2) Center for Medical Genetics, Shinshu University Hospital, Matsumoto, Japan; 3) Research Center for Supports to Advanced Science, Matsumoto, Japan; 4) Shinshu University School of Medicine; 5) Department of Otolaryngology, Shinshu University School of Medicine, Matsumoto, Japan; 6) Department of Otolaryngology, International University of Health and Welfare, Mita Hospital, Tokyo, Japan

**Background:** Musculocontractural Ehlers-Danlos Syndrome (mcEDS) is a recently identified connective tissue disorder, caused by loss-of-function mutations in *CHST14* encoding dermatan 4-O-sulfotransferase-1 (mcEDS-*CHST14*) or in *DSE* encoding dermatan sulfate epimerase (mcEDS-*DSE*), both resulting in defective biosynthesis of dermatan sulfate (DS) replaced by chondroitin sulfate (CS). To date, 42 patients with mcEDS-*CHST14* from 28 families and three patients with mcEDS-*DSE* from two families have been published. Mild hearing impairment was reported in some patients, but has not been systematically investigated. The aim of this study is to delineate otological features in patients with mcEDS-*CHST14*.

**Materials and methods:** Our cohort includes nine patients (eight ears from four male patients; 10 ears from five female patients) with mcEDS-*CHST14*, who had systemic otological investigation in Shinshu University Hospital. The mean age is 18 years old (range, 10–28 years old). Otological investigation comprised subjective symptoms, audiogram, distortion product otoacoustic emissions (DPOAE), and tympanogram.

**Results:** Only one patient felt hearing impairment. Audiogram showed hearing impairment of high-frequency sound in 12 ears (67%), low-frequency sound in six ears (33%), and specifically middle-frequency sound in two ears (11%). Normal hearing was detected in one ear (6%). DPOAE, available in 13 ears, showed hearing impairment of high-frequency sound in 10 ears (77%) and all-frequency sound in two ears (15%). Normal hearing was detected in one ear (8%). Tympanogram, available in 16 ears, showed Ad types in 12 ears (75%). No abnormalities were detected in four ears (25%).

**Discussion:** Hearing impairment is frequent in patients with mcEDS-*CHST14*, detected in 94% of ears and predominantly for high-frequency sound but with a variability including low-frequency type or middle-frequency type. In eight ears (62%), results in DPOAE were compatible to those in audiogram. Ad type in tympanogram did not seem relevant to specific types of hearing impairment. Interestingly, they tended to have conductive hearing impairment for <1kHz sound and sensorineural hearing impairment for >2kHz sound. Regular surveillance using multiple measures is recommended in patients with mcEDS-*CHST14*, and complex pathology was suggested for hearing impairment in these patients.

## PgmNr 1091: Keratin 13 mutations in two Chinese families and the mechanism of white sponge nevus.

### Authors:

W. Cai<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Tongji University School of Medicine, Shanghai, P. R. China; 2) Center for Translational Neurodegeneration and Regenerative Therapy, Shanghai Tenth People's Hospital, Tongji University, Shanghai, P. R. China

---

White sponge nevus (WSN) is an autosomal dominant hereditary disease. Keratin 4 (*KRT4*) and Keratin 13 (*KRT13*) gene mutations were involved in the WSN. We recruited two WSN Chinese families, and oral lesions biopsy with hematoxylin and eosin staining showed that patients had significantly pathological characteristics. The mutations of *KRT4* and *KRT13* gene were detected by PCR and direct sequencing. The multiple alignments of *KRT13* from 28 diverse species homology analysis were performed by the ClustalW program. The *KRT13* expression was measured by Real-Time RT-PCR and Western blot analysis. We revealed the pathological pathway of the WSN expression profile by RNA sequencing (RNA-seq). Sequencing analysis revealed two mutations of *KRT13* gene: one mutation was 332T>C, and amino acid change was Leu111Pro. Another mutation was 340C>T, and amino acid change was Arg114Cys. The sequence of *KRT13* was highly conserved. Real-Time RT-PCR and Western blot analysis results show that *KRT13* expression level is lower in patient but keep almost no change in mRNA level. When treat cells with MG132, *KRT13* protein level was increased and kept almost the same in normal and patient cell. We identified two heritable mutations in the *KRT13* gene, which were associated with the development of WSN. The abnormal degradation of *KRT13* protein of WSN may probably associate with abnormal ubiquitination process. Further RNA-seq analysis suggests that the abnormal degradation of *KRT13* protein of WSN may probably associate with Keratin 7 (*KRT7*) and abnormal ubiquitination process. All the efforts may contribute to the molecular therapy for WSN. Gene-based diagnosis and gene therapy for WSN may become available in the near future while provide reference and instruction for treating other keratin associated diseases.

**Key Words:** gene mutation; keratin 13(*KRT13*); oral disease; ubiquitination; white sponge nevus (WSN); RNA sequencing (RNA-seq)

**Funding:** This work was supported by the National Natural Science Foundation of China (No. 81600881).

**First Author and Corresponding Author:** Wenping Cai, Tongji University School of Medicine, Shanghai, P. R. China. Phone: 86-21-65987015, Email: wenpingcai@tongji.edu.cn

## PgmNr 1092: Investigating differences between Mendelian and non-Mendelian human orthologs of Zebrafish matrisome genes.

### Authors:

J. Brown; G. Unlu; E. Gamazon; L. Basterache; E. Knapik; N. Cox

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt University, Nashville, Tennessee.

---

Mendelian genes are generally regarded as inherently different from non-Mendelian genes. To interrogate differences between Mendelian genes and non-Mendelian genes, we focused on a small subset of genes which would likely have physical manifestations in zebrafish – zebrafish matrisome genes. 1015 zebrafish matrisome genes were identified, with corresponding 719 unique human orthologs. Of 3924 human genes known to be Mendelian, 269 are human orthologs of zebrafish matrisome genes. Of these genes, 72 are genes associated with described disorders in OMIM. These most frequent phenotypes associated with these disorders include vision and eye problems, movement abnormalities, and neurological disorders. The majority of the phenotypes fall under the neurologic category. Matrisome genes can be broken down into different categories: collagens, ECM glycoproteins, ECM regulators, ECM-affiliated proteins, proteoglycans, and secreted factors. Non-Mendelian matrisome genes are more than twice as likely to be ECM-affiliated proteins compared to Mendelian matrisome genes, and more than three times less likely to be collagens. Compared to Mendelian genes without a described disorder in OMIM, those with a described disorder in OMIM are 1.5 times more likely to be collagens. Using PredixVU developed by the Cox lab, phenotypes associated with predicted altered expression of each of the matrisome genes was identified. Cutoffs of 0.001 for both p-values and  $R^2$ -values were used. For tissue-specific associations, phenotypes associated with non-Mendelian genes were 1.4 times more likely to be rheumatologic and phenotypes associated with Mendelian genes were 1.3 times more likely to be endocrine and metabolic and 1.3 times more likely to be gastrointestinal. As more stringent p-value cutoffs are used, there are more extreme differences between category frequencies between Mendelian and non-Mendelian. When using cross-tissue results, the differences seen between non-Mendelian genes and Mendelian genes generally shrinks, only becoming more pronounced in the injuries category. While these results show some evidence of differences between Mendelian and non-Mendelian matrisome genes, further investigation will explore additional differences between Mendelian and non-Mendelian genes.

## **PgmNr 1093: An expanded cohort of patients indicates thrombospondin type laminin G domain and EAR (*TSPEAR*) is a cause of ectodermal dysplasia but may not directly cause hearing loss.**

### **Authors:**

B. Bowles<sup>1</sup>; A. Ferrer<sup>1</sup>; R. Smith<sup>2</sup>; C. Nishimura<sup>2</sup>; D. Babovic-Vuksanovic<sup>3</sup>; B.C. Lanpher<sup>3</sup>; A. Bloch-Zupan<sup>4,5,6</sup>; B. Gérard<sup>7</sup>; T. Rey<sup>7</sup>; C.M. Li<sup>8</sup>; J. Sullivan<sup>9</sup>; V. Shashi<sup>9</sup>; A. Williams<sup>10</sup>; J.A. Hamm<sup>10</sup>; A. Ankala<sup>11,12</sup>; I.M. Wentzensen<sup>13</sup>; E.W. Klee<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 2) Molecular Otolaryngology and Renal Research Laboratories, University of Iowa Carver College of Medicine, Iowa City, IA; 3) Department of Clinical Genomics, Mayo Clinic, Rochester, MN; 4) Faculté de Chirurgie Dentaire, Institut d'Etudes Avancées, USIAS, FMTS, RARENET Interreg V, Université de Strasbourg, Strasbourg, France; 5) Hôpitaux Universitaires de Strasbourg (HUS), Pôle de Médecine et Chirurgie Bucco-dentaires, Hôpital Civil, Centre de référence des maladies rares orales et dentaires, Filière Santé Maladies rares, European Reference Network, Strasbourg, France; 6) Université de Strasbourg, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258, CNRS-UMR7104, Illkirch, France; 7) Laboratoire de Génétique Médicale des Hôpitaux Universitaires de Strasbourg, Strasbourg, France; 8) Department of Pediatrics, McMaster University, Hamilton, ON; 9) Department of Pediatrics, Duke University, Durham, NC; 10) Department of Medical Genetics, East Tennessee Children's Hospital, Knoxville, TN; 11) Emory University School of Medicine, Atlanta, GA; 12) EGL Genetics LLC, Tucker, GA; 13) GeneDx inc., Gaithersburg, MD

---

Biallelic loss of function variants in the thrombospondin type laminin G domain and EAR (*TSPEAR*) have recently been associated with ectodermal dysplasia and hearing loss phenotypes. While the ectodermal association is supported by multiple studies of affected families, the evidence surrounding hearing loss is not well defined. We assembled a cohort of 10 newly reported, unrelated individuals to investigate the *TSPEAR* hearing loss association in greater depth. Individuals were sequenced using whole exome or targeted hearing loss panels. Skin, hair, and dental features characteristic of ectodermal dysplasia or related syndromes were present in a majority of the cohort patients (6/10 patients). Hearing loss was reported in 4/10 patients in the cohort. These results were confounded by the report of additional variants in other hearing loss genes beyond *TSPEAR* in 3 of these individuals. While variants for this cohort concentrated in exons 3, 7, and 8, there was no genotype/phenotype correlation that separated patients with hearing loss from those with ectodermal dysplasia. When presented alongside previous reports, this evidence supports the association of *TSPEAR* variants with ectodermal dysplasia and tooth agenesis features, but raises question as to whether *TSPEAR* alterations are a direct cause of hearing loss. Further functional evidence is needed to demonstrate this assertion.

## PgmNr 1094: Pulmonary complications and CT findings of vascular Ehlers-Danlos syndrome.

### Authors:

H. Morisaki <sup>1,2</sup>; H. Ogino <sup>2</sup>; K. Minatoya <sup>2</sup>; T. Morisaki <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, Sakakibara Heart Institute, Fuchu, Tokyo, Japan; 2) National Cerebral and Cardiovascular Research Center Hospital, Suita, Osaka, Japan; 3) Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

---

Vascular Ehlers-Danlos syndrome (vEDS) is a rare, autosomal dominantly inherited disorder, caused by heterozygous pathogenic variants in *COL3A1* gene, affecting the connective tissue in multiple organ systems. The patients are predisposed to general tissue fragility, characterized by arterial aneurysm, dissection and rupture, bowel rupture, rupture of the gravid uterus and early mortality. Although not usually related to issues of mortality, pulmonary complications including recurrent pneumothorax, hemoptysis, and intrapulmonary hematoma, are annoying and at times serious complications. We reviewed the medical records of 51 molecularly diagnosed vEDS patients retrospectively to evaluate the lung manifestations characteristic to vEDS in patients in our outpatient clinic for connective tissue and aortic diseases or in inpatient unit for vascular surgery. Data were obtained from study visits, medical records and radiology images.

Thirteen of 51 (25%) patients presented pulmonary complications. Twelve suffered recurrent pneumothorax, 11 had episodes of massive hemoptysis, and 9 had both. Pulmonary CT evaluation revealed multiple cavities or parenchymal hemorrhage in six, mostly in lower lobes and periphery. Mean and median ages of first pulmonary complications are 19 and 18, respectively. In 8 cases, pneumothorax or hemoptysis was the first significant event of tissue fragility. There was no apparent genotype-phenotype correlation. Also, there seemed to be no correlation with the risk of pneumothorax/ pulmonary cysts and tissue fragility in general. One patient with multiple giant bullae presented lung findings as a sole complication without any arterial or bowel problem. But all those with episodes of hemoptysis showed other tissue fragilities, especially vascular complications, meaning hemoptysis is an earlier sign for progressive vascular complications.

It is important to be aware of the possible pulmonary presentations of connective tissue disorders, avoid invasive procedures and consult to geneticists for the possibility of vEDS, when we encounter cases with unique radiologic findings, including atypical cysts with signs of infiltration or multiple large bullae, especially those with recurrent hemoptysis.

## PgmNr 1095: Biallelic *KRT5* mutations in epidermolysis bullosa simplex, including a complete human keratin 5 “knock-out”.

### Authors:

H. Vahidnezhad <sup>1</sup>; L. Youssefian <sup>1</sup>; M. Daneshpazhooh <sup>2</sup>; H. Mahmoudi <sup>2</sup>; A. Karimineajd <sup>3</sup>; J. Fischer <sup>4</sup>; J. Christiansen <sup>5</sup>; H. Schneider <sup>6</sup>; A. Guy <sup>7</sup>; L. Liu <sup>7</sup>; J. McGrath <sup>7</sup>; C. Has <sup>8</sup>; J. Uitto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Dermatology and Cutaneous Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA; 2) Department of Dermatology, Razi Hospital, Tehran University of Medical Sciences, Tehran, Iran; 3) Kariminejad-Najmabadi Pathology & Genetics Center, Tehran, Iran; 4) Institute of Human Genetics, University of Freiburg, Faculty of Medicine, Freiburg Germany; 5) Department of Dermatology and Venereology, Skanes University Hospital, Sweden; 6) Department of Pediatrics, University Hospital Erlangen, Erlangen, Germany; 7) St John's Institute of Dermatology, King's College London, Guy's Campus, London; 8) Department of Dermatology, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Germany

---

Epidermolysis bullosa simplex (EBS) is usually inherited as an autosomal dominant disease due to monoallelic gain-of-function mutations in *KRT5* and *KRT14*. Although autosomal recessive forms of EBS have been associated with mutations in at least 10 genes, recessive EBS due to homozygous biallelic *KRT5* mutations has not been reported previously; it has been hypothesized that it would result in prenatal lethality. We sought the genetic causes of EB in a cohort of 512 distinct EB patients by performing whole exome sequencing (WES) and using a disease-targeting next-generation sequencing (NGS) panel of 21 genes. The pathogenicity and consequences of the mutations were determined by expression profiling and at tissue and ultrastructural levels. Two pathogenic, homozygous missense variants of *KRT5* in two patients with generalized EBS and a homozygous null mutation in a patient who died as a neonate from complications of EB were found. The two missense mutations disrupted keratin 5 expression on immunofluorescence microscopy, and the human “knock-out” of *KRT5* showed no RNA and protein expression. Collectively, these findings identify biallelic *KRT5* mutations with a phenotypic spectrum varying from mild, localized and generalized to lethal, expanding the genotypic profile of autosomal recessive EBS.



## PgmNr 1096: Cardiac findings in a cohort of 47 patients with early onset cobalamin C (*cb/C*) deficiency.

### Authors:

C. Van Ryzin<sup>1</sup>; J.L. Sloan<sup>1</sup>; V. Sachdev<sup>2</sup>; D.R. Rosing<sup>2</sup>; K. Mitchell<sup>2</sup>; A. Brofferio<sup>2</sup>; M.D. Levin<sup>2</sup>; S. Ferry<sup>1</sup>; A. Pass<sup>1</sup>; O.A. Shchelochkov<sup>1</sup>; I. Manoli<sup>1</sup>; C.P. Venditti<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MGMGB, NHGRI, NIH, Bethesda, MD.; 2) NHLBI, NIH, Bethesda, MD

---

Combined methylmalonic acidemia and hyperhomocysteinemia, *cb/C* type, is the most common disorder of intracellular cobalamin metabolism. One variant in *MMACHC*, c.271dup A, accounts for 30-50% of the alleles, and is associated with early onset disease in the homozygous state. While the primary manifestations of *cb/C* are ophthalmologic, neurologic, and hematologic, cardiac disease, including left ventricular non-compaction (LVNC) and other structural heart defects, has been variably noted, and not fully described. Therefore, we reviewed the structural cardiac manifestations in a cohort of 47 ethnically diverse, early onset *cb/C* patients (ages 2-36, 14 females) who received standardized evaluations under an NIH natural history protocol (ClinicalTrials.gov Identifier: NCT00078078) over the span of 15 years. 26/47 (55%) were homozygous for the *MMACHC* c.271dupA variant. We obtained 2D echocardiograms in 44/47 individuals, and cardiac MRI in five patients that could cooperate for imaging. On echocardiogram, 13 (30%) were noted to have hypertrabeculations in the left ventricle suggestive of LVNC. Nine of the 13 (69%) were homozygous for c.271dupA. Four of 13 (31%) ages 2, 12, 19 and 24 years had decreased left ventricular ejection fraction. Four of the five patients who had cardiac MRI had confirmation of LVNC and the other subject had global hypokinesis, reduced ejection fraction at age 9 and notable trabeculations, but did not meet criteria for LVNC. Cardiac findings that were comorbid with LVNC included: fetal hydrops, Ebstein's anomaly, tricuspid regurgitation, pulmonary artery dilation, hypoplastic right heart, persistent left SVC, interatrial septal aneurysm, dilated atria and global hypokinesis. Other findings variably included mildly dilated aorta, right ventricular dilatation, PFO, VSD and biventricular hypertrophy. Cardiac involvement, especially LVNC, was relatively common in early onset *cb/C* patients in our cohort. While the long-term natural history of cardiac function remains unknown in this disorder, patients with reduced LV systolic function may be at risk for progression. The chronic hyperhomocysteinemia and decreased activity associated with progressive vision loss may pose additional risks to cardiac health. Longitudinal assessment, including echocardiographic monitoring, and aggressive management of co-morbidities, such as hypertension and hyperlipidemia, are indicated and should be routine for these patients.



## PgmNr 1097: A novel mutation in *KCNJ11* gene in a late diagnosed hyperinsulinemic hyperglycemia patient.

### Authors:

E. Arslan Ates; A.I. Guney

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department OF Medical Genetics, Marmara University Pendik Training and Research Hospital, Istanbul, MD

---

Hyperinsulinemic hypoglycemia is the most common metabolic disorder of neonates. It is important to diagnose and treat to prevent neurological damage such as intellectual disability, recurrent seizures, neuromotor delay if not noticed. Most common genetic causes of hyperinsulinemic hypoglycemia are *ABCC8* and *KCNJ11* gene mutations in infants.

A one year old boy was referred to us because of afebrile seizures starting at the age of 4-months. He was born to consanguineous parents at term with an uneventful prenatal and perinatal history. Physical examination,, biochemical tests, electroencephalogram and cranial imaging studies were normal. His blood sugar level was found to be 29 mg / dl in one of his biochemical tests and diagnosed as hyperinsulinemic hypoglycemia at the age of 12-months. *ABCC8* and *KCNJ11* genes were evaluated by next generation sequencing method. Homozygous c.392T> C (p.Ile141Thr) mutation was detected in the *KCNJ11* gene. Parents were carrying this mutation in heterozygous state, and referred to an endocrinology department with a diagnosis of MODY. His asymptomatic sister was also homozygous for the mutation and her fasting blood sugar level was detected 42mg / dl.

*KCNJ11* is an ATP-sensitive potassium channel gene that is expressed in endocrine cells, neurons and muscle cells, especially in pancreatic beta cells. Homozygous mutations cause neonatal hypoglycemia and persistent diabetes, and heterozygous mutations have been associated with MODY. *KCNJ11* c.392T> C mutation was not reported previously in databases or population studies. In silico analyses predicted this mutation to be pathogenic and mutation cosegregated with the clinical findings.

Here we present this study to emphasize the importance of molecular genetic diagnosis to prevent the complications of hypoglycemia and also hyperglycemia. This study also expands the mutation spectrum of *KCNJ11* gene and contribute to genotype-phenotype correlation of *KCNJ11* mutations.

## PgmNr 1098: A genetic retrospective study of maturity-onset diabetes of the young (MODY) in two population health studies.

### Authors:

A. Bolze <sup>1</sup>; S. White <sup>1</sup>; E.T. Cirulli <sup>1</sup>; L. Sharma <sup>1</sup>; K.A. Schlauch <sup>2,3</sup>; G. Elhanan <sup>2,3</sup>; J. Grzymalski <sup>2,3</sup>; N.L. Washington <sup>1</sup>; J. Lu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Helix, San Francisco, California; 2) Desert Research Institute, Reno, Nevada; 3) Renown Institute of Health Innovation, Reno, Nevada

---

### Background:

Maturity-Onset Diabetes of the Young (MODY) has been reported to account for 1-2% of all diabetes cases. Like most rare diseases, MODY is usually underdiagnosed and it is reported that up to 90% of MODY cases are misdiagnosed as Type 1 or Type 2 Diabetes. Properly diagnosing MODY has medical implications as the first line of treatment is usually different for MODY patients who may not require insulin medication. MODY is generally inherited in an autosomal dominant manner, with variants in *GCK*, *HNF1A* and *HNF4A* explaining more than 80% of reported cases.

### Aims:

The aims of this retrospective study were: (i) to provide the frequency of known pathogenic or likely pathogenic variants, as well as predicted loss-of-function variants in the population, (ii) to estimate the penetrance of these variants, and (iii) to assess the potential medical implications of screening the general population for pathogenic variants in these three genes.

### Method:

We analyzed the exome sequence data of ~70,000 individuals, with 50,000 individuals from the UK Biobank and 20,000 from the Healthy Nevada project, irrespective of their ancestry. Participation in either project was not based on a personal or family history of diabetes. We used ClinVar to select known MODY pathogenic and likely pathogenic variants and Variant Effect Predictor version e!95 annotations to identify predicted loss-of-function variants in *GCK*, *HNF1A*, and *HNF4A*. Electronic health records were available for all individuals analyzed. We looked in particular at (i) ICD10 codes related to diabetes, (ii) medication code(s) for insulin and metformin, as well as (iii) the age of diabetes diagnosis.

### Results:

The frequency of known pathogenic or likely pathogenic variants (n=14 in Healthy Nevada + 42 in UK Biobank), or predicted loss-of-function variants (n=4 in Healthy Nevada + 12 in UK Biobank) in these 3 genes was about 1 in 1,000 individuals (0.1%). Less than 25% of the individuals carrying a pathogenic or likely pathogenic variant had a diagnosis of MODY (2 in Healthy Nevada + 10 in UK Biobank). Moreover, none of the other individuals carrying a pathogenic or likely pathogenic variant had a diagnosis of diabetes. We further characterized the phenotypes of these individuals with regard to glucose and HbA1c levels. These results will also allow to refine the classification of pathogenic and likely pathogenic variants for MODY.

## PgmNr 1099: Next generation sequencing in the diagnosis of rare metabolic disorders.

### Authors:

E. Nikaina <sup>1</sup>; E. Tsoutsou <sup>2</sup>; S. Amenta <sup>2</sup>; N. Selenti <sup>2</sup>; A.B. Skiathitou <sup>1</sup>; R. Husain <sup>3</sup>; T. Siahaidou <sup>1</sup>; E. Fryssira <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neonatal Unit, First Department of Pediatrics, National and Kapodistrian University of Athens, Aghia Sophia Children's Hospital, Athens, Greece; 2) Medical Genetics, National and Kapodistrian University of Athens, Choremio Research Laboratory, Aghia Sophia Children's Hospital, Athens, Greece; 3) Centre for Inborn Metabolic Disorders, Department of Neuropediatrics, Jena University Hospital, Am Klinikum 1, Germany

---

**Introduction:** Molybdenum cofactor deficiency, is a very rare autosomal recessive metabolic disorder, first described in 1977. It is characterized by neonatal seizures resistant to treatment, opisthotonus and dysmorphic features. Low uric acid and increased levels of urine sulfites are also present. The disorder is attributed to mutations in *MOCS1*, *MOCS2*, *MOCS3* and *Gephyrin* genes.

**Case report:** We report the clinical presentation of a female newborn, the second child of consanguineous phenotypically healthy parents. The mother was a G5P2 woman, with 3 spontaneous abortions. From the first day of life, the newborn presented with trunk hypotonia and hypertonia of the limbs, opisthotonus, intractable seizures and dysmorphic features such as facial asymmetry, overlapping sutures, prominent occipital bone and narrow forehead. She had a low nasal bridge with a short nose and a high arched palate, retrognathia, low set ears, low posterior hairline and 5<sup>th</sup> finger clinodactyly. Cardiological investigation revealed atrial septal defect. Biochemical and metabolic profile showed low plasma uric acid (0.2 mg/dl), positive urine sulfite test and elevated urine sulfocysteine levels (145.6 mmol/mol creatinine). Brain MRI revealed thin corpus callosum and cerebellar hemispheres atrophy.

**Results:** Chromosomes from peripheral blood by G-banding technique was normal. Whole exome sequencing revealed a homozygous missense mutation in *MOCS2* (c.226G>A, leading to p.Gly76Arg, in exon 4 of the *MOCS2* gene). Both parents were found to be carriers of the above mutation. *MOCS2* is involved in the biosynthesis of the molybdenum cofactor of the complementary group B.

**Conclusion:** Whole exome sequencing technique, is the gold standard method of investigation in cases of an inconclusive phenotype in critically ill newborns. An accurate and prompt diagnosis leads to precise therapy and appropriate genetic counselling of the family to avoid the risk of recurrence.

## PgmNr 1100: Coexistence of two infrequent genetic pathologies: Noonan and renal Fanconi syndromes.

### Authors:

M. Garcia-Acero <sup>1</sup>; L. Sotaquira <sup>2</sup>; A. Lopez <sup>1,2</sup>; N. Mejia <sup>1,2</sup>; M.C. Beltran <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics SIGEN, Fundación SantaFe de Bogotá, Bogota, Colombia; 2) Facultad de Medicina. Universidad de los Andes, Bogota, Colombia

**Introduction:** Genetic diseases are associated with variations in DNA, which it includes single-nucleotide variation or large variations of DNA fragments. The clinical variability of genetic conditions can occur by variable expressivity of monogenic diseases and less frequent coexistence of two conditions. The genetic advances have allowed the application of very useful molecular studies in the diagnosis of patients with uncertain pathology. We present the unusual coexistence of two diseases with different pathogenic mechanisms and different clinical manifestations.

**Case report:** 4-year-old female patient, born of non-consanguineous parents, prenatal history of hydronephrosis, neonatal finding of dilated cardiopathy with interatrial communication and pulmonary hypertension and optic nerve hypoplasia. Since the year of age with persistent hydronephrosis, renal tubular acidosis, height and weight low. Patient with chromosomal complement 46, XX without alteration. array Genomic Comparative Hybridization without evidence of deletions or duplication [arr (1 -22, X) x2]. Sequencing of the CTNS gene negative. Clinical exome sequencing (trio), which evidenced a heterozygous variant pathogenic in the PTPN1: c.236A>G(p.Gln79Arg) associated with Noonan syndrome, and another heterozygous variant c.1861G>T (p. Glu621\*) in the gene EHHADH, probably pathogenic, associated with renotubular Fanconi syndrome 3.

**Discussion.** The power detection of molecular alterations depends on the resolution of the test used and the type of technique. In this patient the staggered study with different diagnostic techniques allowed to identify the coexistence two infrequent conditions explaining the patient's phenotype. Noonan syndrome is a rasopathy characterized by low stature, facial dysmorphism and congenital cardiac anomalies, among others. In addition, Renotubular Fanconi syndrome, type 3, is an alteration of the renal tubular absorption characterized urinary hyperexcretion of electrolytes with secondary symptoms and evidence of hypophosphatemic rickets and acidosis. Both phenotypes could be mistaken in individuals with pathologic short stature

**Conclusion.** The coexistence of these genetic conditions has not been previously reported with different etiology and high comorbidity associated with a late diagnosis; these results emphasize the importance of clinical exome in patients with clinical phenotypes that do not configure for a specific entity.

## PgmNr 1101: Kallman syndrome: A case report.

**Authors:**

J. Prieto; J. Villalta; R. Quero; M. Manotas; M. Guerra; A. Paredes

[View Session](#) [Add to Schedule](#)

**Affiliation:** Instituto de Genetica Humana, Universidad Javeriana, Bogota Cundinamarca, Colombia

---

Congenital hypogonadotrophic hypogonadism is a rare disease caused by deficiency in the production, secretion or activity of gonadotropin-releasing hormone (GnRH), which is the main regulatory hormone of the reproductive axis. The clinical characterization of this disorder and in the setting of an isolated hypogonadotrophic hypogonadism is due to infertility and partial or incomplete puberty.

Congenital hypogonadotrophic hypogonadism may present as an isolated deficiency of GnRH or be associated with other developmental anomalies such as: cleft lip and palate, dental agenesis, ear abnormalities, congenital deafness, renal agenesis, skeletal alterations or bimanual synkinesias.

Kallman syndrome is a congenital hypogonadotrophic hypogonadism associated with anosmia or hyposmia. This alteration results from an abnormal embryogenic migration of the GnRH neurons in the olfactory plaque to the anterior cerebral region. The association of congenital hypogonadotrophic hypogonadism with an alteration in olfactory perception (anosmia or hyposmia) is found in around 50% of cases of Kallmann syndrome. Thanks to advances in mass sequencing techniques, more than 25 genes related to Kallmann syndrome have been identified, including CHD7. We present the case of a 34-year-old female patient with a history of delayed puberty, learning disorder, hypogonadotrophic hypogonadism and hyposmia, with a G-banded karyotype with normal chromosomal complement 46, XX and a clinical-led exome, which reports a heterozygous variant c.1159C> T (p.Q387 \*) in the CHD7 gene. This mutation has been described not only in the Kallmann syndrome together with the association of olfactory alterations, but also in congenital hypogonadotrophic hypogonadism and CHARGE syndrome.

Research on the molecular basis of the disease and the various signaling pathways involved will help improve the diagnosis and management of patients with hypogonadotrophic hypogonadism, as well as develop more precise genetic counseling and screening.

## PgmNr 1102: Rare *BMP4* sequence variants in patients with hypogonadotropic hypogonadism.

### Authors:

A.K. Topaloglu <sup>1</sup>; R. Yildirim <sup>2</sup>; L.D. Kotan <sup>3</sup>; G. Akkus <sup>4</sup>; E. Unal <sup>5</sup>; I. Turan <sup>6</sup>; S.O. Dilek <sup>3</sup>; M. Tastan <sup>3</sup>; F. Gurbuz <sup>3</sup>; B. Yuksel <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Pediatric Endocrinology, University of Mississippi Med Ctr Jackson, Mississippi; 2) Diyarbakir Pediatric Hospital, Clinic of Pediatric Endocrinology, Diyarbakir, Turkey; 3) Cukurova University, Faculty of Medicine, Division of Pediatric Endocrinology, Adana, Turkey; 4) Antakya State Hospital, Clinic of Endocrinology and Metabolism, Antakya, Hatay, Turkey; 5) Dicle University, Faculty of Medicine, Division of Pediatric Endocrinology, Diyarbakir, Turkey; 6) Sanliurfa Training and Research Hospital, Clinic of Pediatric Endocrinology, Sanliurfa, Turkey

*BMP4* is a member of the bone morphogenetic protein family which is part of the transforming growth factor-beta superfamily. *BMP4* is involved in the embryonic development of various organ and tissues including the craniofacial and olfactory structures, pituitary, eyes, heart, and kidneys. Mutations in this gene are associated with orofacial cleft and microphthalmia in human patients. *BMP4* plays an important role in the embryonic development of the GnRH neurons (Forni et al 2013, Layman et al 2011) and anterior pituitary by regulating diverse cellular responses, such as cell differentiation, migration, adhesion, and proliferation (Massague et al 2000). *BMP4* also has been described as inhibiting FSH production particularly (Nicol et al 2008). Recently, a heterozygous truncating mutation was described in a 6 years old prepubertal child with combined pituitary hormone deficiency (Rodriguez-Contreras et al 2019). Mutations in the BMP genetic network including *BMP4* were reported to be found in patients with hypogonadotropic hypogonadism (HH) in a meeting abstract (Cassatella et al ASHG 2013). However, no detailed description of *BMP4* mutations in the etiology of HH has been found in the literature. Here we present three independent patients with isolated HH who harbor rare seemingly deleterious sequence variants in *BMP4*.

Methods: We screened the whole exome sequencing data from 215 HH patients from Turkey.

Patient number	DNA change	Zygosity	Protein change	Minor allele frequency in gnomAD	CADD score
1	NM_001202.3:c.450C>G	Hom	NP_001193.2:p.(Asn150Lys)	0.00001	22.5
2	NM_001202.3:c.41G>A	Het	NP_001193.2:p.(Cys14Tyr)	0.00001	24.9
3	NM_001202.3:c.751C>T	Het	NP_001193.2:p.(His251Tyr)	0.0001	29.6

Results: We identified apparently deleterious rare *BMP4* sequence variants in three independent patients with HH. All patients were normosmic. Interestingly, variant p.N150K has been reported to cause renal hypodysplasia in heterozygosity as well as homozygosity in two independent Turkish patients (Weber et al 2008). Co-occurrence of HH with kidney anomalies are well known, especially with certain anosmic HH genes. In all of the patients both FSH and LH were similarly deficient, i.e. there was not a particularly more pronounced deficiency of FSH over LH.

Conclusions: These results suggest that inactivating variants in *BMP4* may compromise the central

part of the hypothalamo-pituitary-gonadal axis and result in normosmic HH.



## PgmNr 1103: Genomic characterization of *STX16* deletion in pseudohypoparathyroidism.

### Authors:

M. Kim <sup>1</sup>; J. Ko <sup>2</sup>; Y. Yeon <sup>1</sup>; S. Cho <sup>1</sup>; M. Seong <sup>1</sup>; S. Park <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea; 2) Department of pediatrics, Division of Clinical Genetics, Seoul National University College of Medicine, Seoul, Korea

---

Pseudohypoparathyroidism type 1b (PHP-1b) is a rare imprinting disorder, characterized by renal parathyroid hormone resistance, but the absence of physical features of Albright hereditary osteodystrophy. A common heterozygous 3-kb deletion of *STX16* gene was previously reported in multiple unrelated probands. This deletion cause isolated loss of methylation at *GNAS* exon A/B. We aimed to identify *STX16* deletion and exact breakpoint and to gain insights into the mechanism of *STX16* deletion. We investigated 10 patients who were diagnosed as PHP-1b but didn't have any pathogenic sequence variants in *GNAS*. Methylation-specific multiplex ligation-dependent probe amplification assay of *STX16* were performed to assess the allelic dosage. Junction PCR and Sanger sequencing around presumptive breakpoint area was carried out to reveal the exact breakpoint. A *STX16* gene deletion was detected in 2 of 9 probands (22.2%). Deletion range was the same for both probands which was from g.57,243,566 to g.57,246,545 (2,979 bp), according to the February 2009 human reference sequence (hg19, build37), consistent with previous reports. The 5' and 3' break points were located in the mammalian-wide interspersed repeat (MIR) sequences suggesting an MIR-mediated non-allelic homologous recombination (NAHR) as the putative mechanism for the common *STX16* 3-kb deletion. To the best of our knowledge, this is the first study to reveal that the MIR was involved in the genomic rearrangement.

## PgmNr 1104: Novel *KMT2D* mutation in Kabuki syndrome with midgut malrotation and hyperinsulinism.

### Authors:

T. Phetthong <sup>1</sup>; T. Tim-Aroon <sup>1</sup>; P. Poomthavorn <sup>2</sup>; D. Wattanasirichaigoon <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 2) Division of Endocrinology and Metabolism, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

---

Kabuki syndrome (KS) is a heterogeneous phenotypic genetic syndrome, mainly caused by loss of function in *KMT2D* and *KDM6A*. Gastrointestinal anomalies in KS patients have been rarely reported. The present patient was a female infant, born at 37-week gestation. Her growth parameters included weight of 2,714 g (10<sup>th</sup> centiles), length of 49 cm (50<sup>th</sup> centiles), and head circumference of 32 cm (10<sup>th</sup> centiles). At one hour after birth, she developed respiratory distress and found to have hypoglycemia (blood sugar 34 mg/dL). She was treated with bolus dose of 10% dextrose. Critical blood samples showed elevated serum insulin level at 10.4 µU/ml, prompting diagnosis of hyperinsulinism. Due to refractory response to high dose of glucose infusion, diazoxide treatment was started on day of life (DOL) 5 which resulted in gradual improvement of hypoglycemia. On DOL15, the patient developed feeding difficulty, abdominal distension and frequent vomiting. Barium swallowing revealed abnormal position of duodenal and jejunal loops, indicating midgut malrotation, prompting surgical operation with Ladd procedure performed. She was discharged home on DOL 45. Due to her minor dysmorphic features including smooth philtrum, thin vermilion of lips and microretrognathia, genetic evaluation was carried out. There was no specific recognized syndrome was suspected. Cytogenomic and SNP microarray (CytoSNP850, Illumina) analysis was performed which revealed normal findings. Whole exome sequencing revealed a *de novo* novel pathogenic heterozygous *KMT2D* mutation, c.13564A>T (pLys5122\*). *KMT2D* is a transcriptional activator that induces the transcription of target genes by covalent histone modification and involved in adhesion-related cytoskeletal regulation, which might affect cell growth and survival effecting to coupling of apoptosis and Left-Right patterning controls organ looping. Therefore, this might explain the pathogenic mechanism of *KMT2D* mutation leading gut malrotation in the present patient. Retrospectively, the patient did not show typical facial phenotypes of KS such as long palpebral fissure and eversion of the lateral one-third of lower eyelid, probably due the characteristic facial phenotypes of KS was not evident at that young age. In conclusion, we address the benefit of WES in establishing diagnosis of rare disorder in patient without characteristic features of a known syndrome and the rare phenotype of KS namely gut malrotation.

## PgmNr 1105: Molecular diagnosis of osteogenesis imperfecta: Targeted gene panel sequencing.

### Authors:

S.G. Temel <sup>1</sup>; L. Aliyeva <sup>1</sup>; N. Kaya <sup>1</sup>; O. Demirbağ <sup>1</sup>; E. Eren <sup>2</sup>; O. Tarım <sup>2</sup>; S. Özemri Sağdıç <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Medical Genetics, Faculty of Medicine, Univ Bursa Uludag, Gorukle, Bursa, Turkey; 2) Dept of Pediatric Endocrinology, Faculty of Medicine, Univ Bursa Uludag, Gorukle, Bursa, Turkey

---

Osteogenesis imperfecta (OI) is a hereditary disease characterized by increased bone fragility and low bone mass. The incidence is one in 10,000 births. Mutations in the *COL1A1* and *COL1A2* genes that encode type I collagen, are responsible for more than 90 percent of all cases of OI.

To understand the etiology of the disease, to identify the carriers and the underlying genetic cause is extremely important for appropriate genetic counseling. In this study, 29 patients who admitted to our clinic between July 2018 and June 2019 with OI were screened for related genes. Exon and the exon-intron boundaries were amplified using OI Paragon kit (14 genes related with OI) and sequence analysis was performed via Illumina MiSeq™ System. Sophia DDM platform and Sophia Genetics' MOKA software was used for variant analysis and annotations. Pathogenicity of variants were analyzed using online tools (HGMD, ClinVAR, MutationTaster, SIFT, Polyphen) and ACMG criteria. 37 variants in *BMP1*, *COL1A1*, *COL1A2*, *FKBP10*, *LEPRE1*, *PLS3*, *SP7*, *SERPINH1* and *PLOD2* genes were identified. According to ACMG classification, 28 of these variants were variant of unknown significance (VUS), while nine of them were pathogenic and possibly pathogenic. 14 of these variants were novel. Our study affirms that genetic screening of patients with OI symptoms by using multigene panel not only eases the diagnosis but also help to provide accurate and effective genetic counseling for prenatal diagnosis and preimplantation genetic diagnosis options for patients and their family members for the risk of developing the diseases in the future.

## PgmNr 1106: Strong genotype-phenotype correlations in *ACTG2* are the primary determinants of disease severity in visceral myopathy.

### Authors:

N. Assia Batzir <sup>1,2</sup>; P.K. Bhagwat <sup>1,3</sup>; A. Larson <sup>4</sup>; Z. Coban Akdemir <sup>1</sup>; S. Jhangiani <sup>5</sup>; B. Isidor <sup>6</sup>; R. Littlejohn Okashah <sup>7</sup>; P. Magoulas <sup>1,2</sup>; R. Marom <sup>1,2</sup>; S. McLean <sup>7</sup>; K.M. Nugent <sup>7</sup>; E. Roeder <sup>7</sup>; J.E. Posey <sup>1</sup>; J.R. Lupski <sup>1,2,5,8</sup>; A.L. Beaudet <sup>1,2,8</sup>; M.F. Wangler <sup>1,2,3,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College Medicine, Houston, Texas; 2) Texas Children's Hospital, Houston, Texas; 3) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas 77030, USA; 4) Section of Clinical Genetics and Metabolism, Department of Pediatrics, University of Colorado School of Medicine, Aurora, CO, USA; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 6) CHU de Nantes, Service de Génétique Médicale, Nantes 44093 Cedex 1, France; INSERM, UMR-S 957, 1 Rue Gaston Veil, Nantes 44035, France; 7) Department of Pediatrics, Children's Hospital of San Antonio, Baylor College of Medicine, San Antonio, TX, 78207, USA; 8) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA

Visceral myopathy (MIM #155310) is a disorder manifesting primarily as abnormal intestinal and bladder peristalsis, with a clinical spectrum that includes Megacystis Microcolon Intestinal Hypoperistalsis Syndrome (MMIHS) and Chronic Intestinal Pseudo-Obstruction (CIPO). Symptoms range from intrauterine bladder distension with severe postnatal feeding intolerance and dependence on parenteral nutrition to intermittent constipation with or without bladder dysfunction. A number of genes have been identified underlying these phenotypes with the vast majority of molecularly-diagnosed cases caused by dominant variants in *ACTG2*; however, the overall genetic architecture of visceral myopathy has not been well-characterized.

We report in the clinical features, molecular diagnostic rates and genotype-phenotype correlations in 51 families with visceral myopathy ascertained based on clinical features of megacystis, functional bladder or GI obstruction or microcolon. Targeted *ACTG2* sequencing or exome sequencing yielded a molecular diagnosis in 64% (34/53) of cases, of which 97% (33/34) is attributed to *ACTG2* and 3% (1/34) to *MYLK*. As a group, the *ACTG2*-negative cases had a higher likelihood of having a more favorable clinical outcome and more restricted disease. Strikingly, missense variants in five conserved Arginine residues in *ACTG2* accounted for 49% (26/53) of the disease burden in this cohort; all alleles were due to CpG dinucleotide changes. Poor outcomes, defined as total parenteral nutrition (TPN) dependence, death or transplantation, were invariably associated with one of the Arginine missense alleles within the *ACTG2*-positive cases in the cohort. Further characterization of specific Arginine missense mutations suggests an allelic series with a severity spectrum of p.R178 > p.R257 > p.R40 along with other less frequently reported recurrent sites p.R63 and p.R211.

These results provide a seemingly robust genotype-phenotype correlation for *ACTG2*-related disease and emphasize the important role of Arginine missense variants in visceral myopathy. Our findings potentially have important implications on the diagnostic strategy and patient counseling.

## PgmNr 1107: Expanding the phenotype of inflammatory bowel disease in Hermansky-Pudlak syndrome.

### Authors:

B.R. Gochuico <sup>1</sup>; X. Parisi <sup>1,2</sup>; N.R. Shelman <sup>3</sup>; M.A. Merideth <sup>1</sup>; W.J. Introne <sup>1</sup>; T. Heller <sup>4</sup>; W.A. Gahl <sup>1,5</sup>; M.C.V. Malicdan <sup>1,5</sup>; K.J. O'Brien <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland; 2) Medical Research Scholars Program, Clinical Center, National Institutes of Health, Bethesda, Maryland; 3) Department of Pathology, Boston Children's Hospital; 4) National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland; 5) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, Maryland

**Background:** A subpopulation of patients with Hermansky-Pudlak syndrome (HPS) develop inflammatory bowel disease. A paucity of information is available about disease prevalence, clinical characteristics, genotyping analysis, and histologic features. The objective of this study is to expand the understanding of HPS inflammatory bowel disease.

**Methods:** Medical records of all children and adults with HPS evaluated at the National Institutes of Health Clinical Center from 1995 to 2018 were retrospectively reviewed. Their clinical characteristics, genotyping results, and histologic findings were analyzed.

**Results:** Inflammatory bowel disease affected 35 (13.5%; 11 male, 24 female) of 259 patients with HPS. The mean age of onset was 20 years with a range of 1 to 52 years. Six children had very early-onset inflammatory bowel disease. The most common symptoms were hematochezia, abdominal pain, and loose stools. Fistulae or extra-intestinal manifestations developed in 31% or 23%, respectively. Genotyping showed that patients with HPS-1, HPS-3, HPS-4, or HPS-6 were diagnosed with inflammatory bowel disease; HPS-3 was associated with mild symptoms. Seventy-four percent of patients received medical therapy. The most common treatments were systemic corticosteroids, anti-tumor necrosis factor alpha drugs, and systemic non-steroidal immunosuppressants. Seven of 13 patients who received anti-tumor necrosis factor alpha therapy had prolonged clinical responses. Bowel resection was performed in 37%. Active cryptitis, chronic inflammatory changes, granulomas, and ceroid lipofuscinosis were histopathologic findings in three colectomy specimens.

**Conclusion:** Inflammatory bowel disease resembling Crohn's disease affects a subpopulation of patients with HPS-1, HPS-3, HPS-4, or HPS-6. Very early-onset inflammatory bowel disease can develop in children with HPS. Severity of disease is highly variable and is generally mild in patients with HPS-3. Treatment of patients with HPS inflammatory bowel disease was similar to that of patients with Crohn's disease.

## PgmNr 1108: *TRPC6* pathogenic variant in a Japanese boy with infantile nephrotic syndrome.

### Authors:

H. Hanafusa<sup>1,2</sup>; Y. Hidaka<sup>3,4</sup>; T. Murase<sup>3</sup>; T. Yamaguchi<sup>1,5</sup>; M. Minatogawa<sup>1,5</sup>; K. Nakamura<sup>1,5</sup>; K. Wakui<sup>1,5</sup>; K. Takano<sup>1,5</sup>; T. Kosho<sup>1,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Medical Genetics, Shinshu university Hospital, Matsumoto, Japan; 2) Problem-Solving Oriented Training Program for Advanced Medical Personnel: NGSD (Next Generation Super Doctor) Project, Matsumoto, Japan; 3) Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan; 4) Department of Molecular Genetics, Wakayama Medical University, Kimiidera, Japan; 5) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 6) Research Center for Supports to Advanced Science, Matsumoto, Japan

---

Transient receptor potential channel C6 (TRPC6) coded by *TRPC6* is involved in slit membrane formation in podocytes. Abnormalities in proteins that form slit membranes, including TRPC6, are known to cause glomerular disease. Originally, *TRPC6* variants were reported as the cause of adult-onset focal segmental glomerulosclerosis (FSGS), the main cause of steroid resistant nephrotic syndrome (SRNS). However, pediatric cases of FSGS due to *TRPC6* variants have also been reported. Here, we report a patient with rapidly progressing infantile nephrotic syndrome, in whom a pathogenic *TRPC6* variant was identified.

The patient is a 2-years-old Japanese boy who is the second child from unrelated healthy parents. The pregnancy and delivery were uneventful. At age 8 months, edema appeared and gradually increased, and he was diagnosed with NS. He received prednisolone therapy for four weeks. But the symptom did not improve, and he was admitted to our hospital for further investigation and treatment. Renal biopsy demonstrated diffuse mesangial sclerosis (DMS). Clinical exome analysis, using a MiSeq and TruSight One Sequencing Panel (Illumina, San Diego, CA, USA) including 4813 genes associated with human genetic disorders, revealed a heterozygous missense variant of *TRPC6* (MN\_004621:c.523C>T causing p.Arg175Trp). At age 1 year and 5 months, peritoneal dialysis was introduced because of respiratory failure resulting from overhydration triggered by infection.

The same variant was not found in his parents, nor registered in databases of healthy individuals. The variant was reported to cause childhood-onset SRNS in a patient with a renal pathology as FSGS (Wang et al, 2017). Therefore, it would be a cause of infantile nephrotic syndrome with a renal pathology as DMS in the current patient. Considering that patients with SRNS caused by *TRPC6* variants, though the number is limited, were reported to tolerate renal transplantation without relapse (Santín et al, 2009; Sun et al, 2017), renal transplantation would be a reasonable choice for this patient. Genetic investigation could be useful in evaluating and treating patients with infantile or childhood-onset severe nephrotic syndrome.

## PgmNr 1109: Heterozygous *CTNNB1* and *TBX4* variants in a patient with abnormal lung growth, pulmonary hypertension, microcephaly, and spasticity.

### Authors:

J.A. Karolak <sup>1,2</sup>; P. Szafranski <sup>1</sup>; D. Kilner <sup>3</sup>; C. Patel <sup>4</sup>; B. Scurry <sup>5</sup>; E. Kinning <sup>6</sup>; K. Chandler <sup>7</sup>; S.N. Jhangiani <sup>8</sup>; Z.H. Coban Akdemir <sup>1</sup>; J.R. Lupski <sup>1,8,9,10</sup>; E. Popek <sup>11</sup>; P. Stankiewicz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poznan, Poland; 3) Department of Respiratory and Sleep Medicine, Queensland Children's Hospital, South Brisbane, Queensland, Australia; 4) Genetic Health Queensland, Royal Brisbane and Women's Hospital, Brisbane, Queensland, Australia; 5) Pathology Queensland, Royal Brisbane and Women's Hospital and Lady Cilento Children's Hospital, Brisbane, Queensland Australia; 6) West of Scotland Regional Genetics Service, Queen Elizabeth Hospital, Glasgow, UK; 7) Manchester Centre for Genomic Medicine, Saint Mary's Hospital, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 9) Texas Children's Hospital, Houston, TX; 10) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 11) Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX

The canonical wingless (Wnt) and fibroblast growth factor (FGF) signaling pathways are essential for regulation of human organ development. Perturbations of these pathways and disruptions from biological homeostasis have been associated with abnormal morphogenesis of multiple organs, including the lung. Components of Wnt and FGF signaling, *CTNNB1* and *TBX4*, respectively, modulate epithelial-mesenchymal interactions required for proper morphogenesis and lung development. *CTNNB1* variants have been identified mainly in patients with neurodevelopmental disorders, visual defects, and different types of cancers whereas *TBX4* variants have been associated with ischiocoxopodopatellar syndrome, pulmonary hypertension (PAH), developmental delay with heart and limb defects (PMID: 20206336), and recently with lethal lung developmental diseases (PMID: 30639323). Using exome sequencing in a term infant with lung growth abnormality, severe microcephaly, and spasticity, who died at four months of age due to progressively worsening of PAH and respiratory failure, we identified a *de novo* heterozygous nonsense variant c.1603C>T (p.Arg535\*) in *CTNNB1* and a heterozygous missense variant c.1198G>A (p.Glu400Lys) in *TBX4* inherited from the unaffected father. Both variants are rare and predicted to be likely deleterious. Interestingly, the *CTNNB1* variant c.1603C>T was reported in two unrelated infants with low oxygen saturation or persistent PAH during the neonatal period coexisting with microcephaly, spasticity, intellectual disability, and behavioral anomalies (PMID: 27915094). While severe microcephaly and spasticity detected in our patient overlap the clinical features in these two subjects, the observed lung phenotype is distinct and includes abnormal lung growth with enlarged simplified alveoli, prominent lymphangiectasia, and PAH with mild venous hypertensive changes. We suggest that perturbations in *CTNNB1* and *TBX4* act synergistically to produce a severe lung phenotype. Our data expand the known phenotypic spectrum of *CTNNB1* and *TBX4* variants, support the contention that

dual molecular diagnoses can parsimoniously explain blended phenotypes, and further implicate a recently proposed complex multi-locus mutational burden underlying the lung developmental diseases (NIH-R01HL137203, NIH-UM1HG006542).



## PgmNr 1110: Exome sequencing in biliary atresia family trios: Investigating ciliopathy genes.

### Authors:

A. Sabo<sup>1</sup>; J. Schraw<sup>2</sup>; J. Jayaseelan<sup>1</sup>; Y. Lei<sup>3</sup>; R. Finnell<sup>3</sup>; W. Nembhard<sup>4</sup>; G. Shaw<sup>5</sup>; S. Carmichael<sup>5</sup>; P. Romitti<sup>6</sup>; M. Werler<sup>7</sup>; M. Browne<sup>8</sup>; A. Olshan<sup>9</sup>; M. Canfield<sup>10</sup>; A. Agopian<sup>11</sup>; M. Feldkamp<sup>12</sup>; M. Bamshad<sup>13</sup>; D. Nickerson<sup>13</sup>; J. Chong<sup>13</sup>; E. Blue<sup>13</sup>; J. White<sup>13</sup>; J. Mullikin<sup>14</sup>; F. Pangilinan<sup>14</sup>; L. Brody<sup>14</sup>; D. Muzny<sup>1</sup>; R. Gibbs<sup>1</sup>; S. Harpavat<sup>2,15</sup>; P. Lupo<sup>2</sup>; The National Birth Defects Prevention Study

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Pediatrics, Baylor College of Medicine, Houston, TX; 3) Center for Precision Environmental Health, Baylor College of Medicine, Houston, TX; 4) Fay W. Boozman College of Public Health, University of Arkansas for Medical Sciences, Little Rock, AR; 5) Stanford University School of Medicine, Stanford, CA; 6) University of Iowa, Iowa City, IA; 7) Boston University, Boston, MA; 8) State University of New York, Albany, NY; 9) University of North Carolina, Chapel Hill, NC; 10) Texas Department of State Health Services, Austin, TX; 11) Department of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas School of Public Health, Houston, TX; 12) University of Utah School of Medicine, Salt Lake City, UT; 13) Center for Mendelian Genomics, University of Washington, Seattle, WA; 14) NISC Comparative Sequencing Program, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 15) Gastroenterology, Hepatology and Nutrition, Texas Children's Hospital, Houston, TX

---

Biliary atresia (BA) is a rare disease of liver and bile ducts that occurs in infancy. Children with BA undergo a surgical procedure to reestablish bile flow, but 70-80% of patients ultimately require liver transplant, making BA the leading cause of liver transplant in children. To investigate the genetic causes of BA, we performed an exome sequencing variant prioritization analysis on 54 case-parents trios recruited as part of the National Birth Defects Prevention Study, with 91% of cases having isolated BA and no other congenital anomaly. Whole-exome sequencing (WES) was performed on Illumina platform using SeqCap EZ Exome capture reagent. After performing standard quality control exclusions, single nucleotide variants and small insertion/deletions were identified using Platypus and annotated using ANNOVAR. In each trio, we prioritized *de novo*, rare homozygous, and rare compound heterozygous (CH) protein-altering variants. On average each trio was found to have one *de novo*, two homozygous and three CH candidate variants. We Sanger sequence validated a *de novo* missense variant in the syntaxin-binding protein 4 gene (*STXBP4*); the same variant was identified in clinical exome data from a child with liver failure. We identified two probands with different rare compound heterozygous variants in the polycystic kidney disease 1 like 1 gene (*PKD1L1*), a recently identified BA candidate gene associated with ciliary calcium signaling and establishing left-right asymmetry. A focused review of ciliopathy genes identified one patient that carried CH variants in inversin (*INVS*), a gene involved in primary cilia function and left-right axis determination. Utilizing StringDB, we have established a significant interaction enrichment ( $p < 0.004$ ) among the 152 genes harboring rare CH variants in cases, with cytoskeleton organization and calcium ion binding as most significantly enriched biological process and molecular function. Our study results suggest that BA is heterogeneous in its underlying genetic etiology. In the future, larger studies utilizing more

comprehensive genome sequencing methods better suited to assay structural and non-coding variants will be needed to better understand the genetic component of BA.

## PgmNr 1111: Identification of two novel pathogenic mutations of *SLC22A12* in two renal hypouricemia families with semidominant phenotypes.

### Authors:

J.H. Wu <sup>1</sup>; K. Fujita <sup>2</sup>; M. Hosoyamada <sup>3</sup>; K. Ichida <sup>2</sup>; H. Sakurai <sup>4</sup>; M.y. Chung <sup>5,6</sup>; Y.T. Tsao <sup>7,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; 2) Department of Pathophysiology, Tokyo University of Pharmacy and Life Sciences, Tokyo, Japan; 3) Department of Human Physiology and Pathology, Teikyo University, Tokyo, Japan; 4) Department of Pharmacology and Toxicology, Kyorin University School of Medicine, Tokyo, Japan; 5) Department of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, Taiwan, ROC; 6) Department of Medical Research, Taipei Veterans General Hospital, Taipei, Taiwan, ROC; 7) Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan, ROC; 8) Division of Nephrology, Department of Medicine, Taoyuan General Hospital, Ministry of Health and Welfare, Taoyuan, Taiwan, ROC

---

Renal hypouricemia (RHUC, OMIM 220150 and 612076) with clinical presentations of low serum uric acid (UA) concentrations (less than 2.0 mg/dL) tends to complicate with exercise-induced acute renal failure (EIAHF). In patients with RHUC, many loss-of-function mutations have been detected in *SLC22A12* (OMIM: 607096, RHUC1) and *SLC2A9* (OMIM: 606142, RHUC2), which encode two major urate transporters in renal tubules. Here we report two patients from unrelated Taiwanese families suffered from hypouricemia with the initial presentation of EIAHF, and two novel missense variants (NM\_144585.3:c.1376T>G, NP\_653186.2:p.L459R in family I and c.365A>G, p.Y122C in family II) in *SLC22A12* identified via direct sequencing of *SLC22A12* and *SLC2A9*. Both mutations had allele frequencies below 0.01 validated in 200 Taiwanese cohorts and impaired uric acid uptake was revealed in oocyte injection assays. Biochemistry examinations showed hypouricemia in family members heterozygous for each mutation (mean serum UA level: 1.76 mg/dL, 95% CI: 1.50-2.02 mg/dL) and even lower serum UA levels in probands (mean: 0.95 mg/dL, 95% CI: 0.85-1.05 mg/dL, p<0.05). These probands are compound heterozygotes with different possible loss-of-function variants of *SLC22A12*. In conclusion, our results suggest that these two novel mutations lead to RHUC by impairing function of urate transporters and account for biochemically semidominant phenotypes in these two families.

## PgmNr 1112: Very low-level *FOXF1* parental somatic mosaicism in families with alveolar capillary dysplasia with misalignment of pulmonary veins.

### Authors:

P. Stankiewicz<sup>1</sup>; Q. Liu<sup>1</sup>; N.G. Xie<sup>2</sup>; L.R. Wu<sup>2</sup>; G. Rocha<sup>3</sup>; S. Fernandes<sup>4,5</sup>; I.F. Lo<sup>6</sup>; S. Lam<sup>6</sup>; D. Mowat<sup>7,8</sup>; E.K. Fiorino<sup>9</sup>; M. Edelman<sup>10</sup>; J. Fox<sup>11</sup>; D.A. Hayes<sup>12</sup>; D. Witte<sup>13</sup>; A. Parrot<sup>14</sup>; E. Popek<sup>15</sup>; P. Szafranski<sup>1</sup>; D.Y. Zhang<sup>2</sup>; J.A. Karolak<sup>1,16</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Department of Bioengineering, Rice University, Houston, TX; 3) Department of Pediatrics, Hospital de Sao Joao, Porto, Portugal; 4) Department of Genetics, Faculty of Medicine, Porto University, Porto, Portugal; 5) Instituto de Investigação e Inovação em Saúde, Porto Portugal; 6) Clinical Genetic Service, Department of Health, Hong Kong SAR, China; 7) Centre for Clinical Genetics, Sydney Children's Hospital, Randwick Sydney, NSW, Australia; 8) School of Women's and Children's Health, The University of New South Wales, Sydney, NSW, Australia; 9) Division of Pediatric Pulmonary Medicine, Cohen Children's Medical Center – Northwell Health, New Hyde Park, NY; 10) Division of Pediatric Pathology, Cohen Children's Medical Center – Northwell Health, New Hyde Park, NY; 11) Division of Medical Genetics, Cohen Children's Medical Center – Northwell Health, New Hyde Park, NY; 12) Division of Pediatric Cardiology, Cohen Children's Medical Center – Northwell Health, New Hyde Park, NY; 13) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 14) The Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 15) Department of Pathology and Immunology, Baylor College of Medicine, Houston, TX; 16) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poznan, Poland

---

Alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV) is a rare lethal neonatal developmental lung disease. The vast majority of ACDMPV cases are sporadic and causative *de novo* point mutations or copy-number variant (CNV) deletions involving *FOXF1* or its distant lung-specific enhancer have been described. Importantly, in a few familial ACDMPV, the pathogenic *FOXF1* variants were transmitted from the somatic mosaic or heterozygous unaffected parent. To assess the scale and ratio of parental somatic mosaicism, we screened 18 unrelated ACDMPV families with the *FOXF1* variant determined to be apparent *de novo* (n=14), of unknown parental origin (n=1), or inherited from a parent suspected to be somatic and/or germline mosaic (n=3), using highly-sensitivity blocker displacement amplification (BDA), droplet digital PCR (ddPCR), quantitative PCR, long-range PCR, and array comparative genomic hybridization. BDA is a PCR-based technology that preferably amplifies allele with SNVs or CNVs, allowing for rare event quantitation with precision similar to ddPCR (PMID: 29805844). We have identified and characterized four (22%) families with parental mosaicism levels ranging between <0.03% and 19%, confirming that the rate of parental somatic mosaicism in ACDMPV is higher than previously thought. In one family, mosaic allele ratio in different tissues originating from all three germ layers ranged between <0.03% and 0.65%, suggesting that *FOXF1* variant might have occurred during early embryonic development. Comparison between results obtained using BDA and ddPCR for *FOXF1* variants indicates that both methods can accurately and efficiently quantitate low-level mosaicism. Our study further implies the importance of a systematic

screening of parental samples for low-level somatic mosaicism and shows that parents of children with ACDMPV who are found negative for *FOXF1* variants by the routine detection methods may benefit from re-analyses using more sensitive and quantitative methods.

## PgmNr 1113: Overcoming challenges in identification, annotation, and interpretation of variants in DSD genes.

### Authors:

E.C. Delot <sup>1,2</sup>; H. Barseghyan <sup>1,2</sup>; A. Parivesh <sup>1</sup>; H. Lee <sup>3</sup>; E. Vilain <sup>1,2</sup>; the DSD Translational Research

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Medicine Research, Children's National Medical Center, Washington, D.C., USA; 2) Genomics and Precision Medicine, School of Medicine and Health Sciences, George Washington University Washington, D.C., USA; 3) UCLA Depts of Human Genetics & Pathology and Laboratory Medicine, Los Angeles, CA, USA

Whole-genome sequencing (WGS) of 94 individuals with a wide spectrum of Disorders/Differences of Sex Development (DSD) revealed challenges of variant calling and classification for several well-known DSD genes. Using two different analysis pipelines, pathogenic variants were identified in *AR*, *SRY*, *WT1*, *NR5A1/SF1* and *SRD5A2*, with a 31% diagnosis rate in the non-syndromic 46,XY DSD subcohort.

Sequencing of *AR* proved challenging with low gene coverage and read depth, in particular next to the polyQ or polyG repeats. We were able to identify a complex frameshift variant missed by commercial clinical testing. WGS established a firm diagnosis of 5 $\alpha$ -reductase deficiency in 4 patients with homozygous or compound heterozygous variants in *SRD5A2*. Two of these had undergone clinical exome sequencing, which had missed the diagnosis. We found that *SRD5A2* is incorrectly annotated in ENSEMBL (but not RefSeq) as a non-coding transcript, resulting in ENSEMBL-based annotators filtering out *SRD5A2* variants. We recommend reanalysis of negative exomes for patients with the appropriate phenotypes. Identification of the V89L polymorphism in *SRD5A2* was also difficult because of an error of alignment to Hg19 by some platforms. Availability of larger, more ethnically diverse reference databases (such as ExAC/gnomAD) also prompted reclassification as likely benign of previously published pathogenic variants (e.g. in *MAP3K1*).

To assess evidence available in the databases used by laboratories to determine variant pathogenicity, we examined ClinVar variants in 69 DSD genes. For several of the long-known DSD genes, available proof was obvious, with the highest numbers of variants reported for *AR*, *NROB1/DAX1*, *CYP21A2*, *SRD5A2*, *SRY*, *NR5A1/SF1*, or *WT1*. However, historical variants, with pathogenicity strongly supported by published evidence from *in vitro* or animal models, were often missing, resulting in gross under-estimation of available evidence for several genes (e.g. *AMHR2*, *SRY* or *CYP21A2*). The difficulty of interpreting whether a ClinVar-reported variant is causative of the DSD phenotype is increased for genes causing syndromic disorders where the genital phenotype is incompletely penetrant (e.g. *DHCR7* in CHARGE) or where XY and XX individuals are differentially affected (e.g. *ATRX*).

These challenges highlight the need for standardization of variant annotation tools and expert human interpretation of sequencing data in patients with rare disorders of complex etiology such as DSD.

## PgmNr 1114: Generation of a Zebrafish model to study the role of *GPR101* in embryonic development and somatic growth.

### Authors:

G. Trivellin <sup>1</sup>; C.H. Tsai-Morris <sup>2</sup>; L.C. Hernández-Ramírez <sup>1</sup>; T. Gupta <sup>2</sup>; H.A. Burgess <sup>2</sup>; B. Feldman <sup>2</sup>; C.A. Stratakis <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Section on Endocrinology and Genetics, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, 20892, Maryland, United States; 2) Division of Developmental Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, 20892, Maryland, United States

---

**Background:** X-linked acrogigantism is a novel syndrome of early childhood-onset pituitary gigantism caused by duplications of *GPR101*, a gene that is predominantly expressed in the brain.

**Aim:** To elucidate the role played by *GPR101* during embryonic development and its effects on body growth.

**Methods:** *gpr101* knockout (KO) zebrafish were generated by CRISPR-Cas9. Growth was monitored until 9 weeks post-fertilization (wpf) in fish housed both in groups and individually. The mRNA expression of *gpr101*, pituitary and hypothalamic hormones (*gh1*, *prl*, *pomca*, *pomcb*, *trh*), and growth regulators (*igf1a*, *igf2a*, *igf2b*) was measured by qPCR at 7 dpf and 6 wpf (heads). Development of maternal-zygotic *gpr101* KO (MZ*gpr101*) embryos from fertilization to 24 hpf was monitored by time-lapse microscopy. Morphometric analysis of MZ*gpr101* brains was done by staining 6 dpf larvae with lysotracker and imaging with a confocal microscope.

**Results:** Two frame-shifting alleles - a deletion of 5 bp and a deletion of 13 bp - were selectively propagated from F founders. Both frameshifts were present at the mRNA level, and *gpr101* expression between mutants and wild type (WT) controls was similar, excluding genetic compensation triggered by mutant mRNA degradation. A significant decrease in standard length, body weight, and BMI in KOs was seen at 3, 6, and 9 wpf ( $p < 0.05$  for all). Genotype frequencies followed Mendelian distribution. Sex did not influence the three variables at 9 wpf, when fish reach sexual maturity. The smaller size was not accompanied by significant differences in the expression of hypothalamic or pituitary hormones (at 6 wpf). MZ*gpr101* were characterized to investigate the contribution of maternal factors. The majority of embryos died during the blastula-gastrula transition. Pericardial edema was present in about half the surviving embryos. The remaining WT-looking larvae showed decreased expression of *gh1* ( $p = 0.07$ ) and *igf2a* ( $p < 0.05$ ), enlarged hypothalamus and sensory ganglia, and a smaller dorsal diencephalon ( $p < 0.01$  for all). Individually-housed MZ*gpr101* were already smaller before feeding started, excluding its effect on growth.

**Conclusions:** This animal model represents the first *in vivo* evidence that *gpr101* regulates body growth. We also unraveled an essential role for *gpr101* during blastulation/gastrulation. Finally, the morphometric findings suggest that *gpr101* might be a negative regulator of the canonical Wnt pathway. Future studies will validate this hypothesis.

## PgmNr 1115: Investigating *SFRP4* causing Pyle disease in an African child.

### Authors:

C. Oosterwyk <sup>1</sup>; N. Manyisa <sup>1</sup>; G. Agenbag <sup>1</sup>; E. Chimusa <sup>1,2</sup>; A. Wonkam <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pathology, University of Cape Town, Cape Town, Western Province, South Africa; 2) African Institute for Mathematical Sciences (AIMS), Cape Town, South Africa; 3) Department of Medicine, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

---

Pyle disease (PD) is a rare autosomal recessive disorder leading to widened metaphyses of long bones, cortical bone thinning and genu valgum. Four mutations causing PD in the *SFRP4* gene have been reported in four unrelated families of different ethnic backgrounds worldwide. The aim of this project was to investigate the involvement of mutations in *SFRP4* in a South African (SA) patient. Sanger sequencing was performed for the *SFRP4* gene in the SA patient of mixed ancestry. This was followed-up by targeted sequencing of variants suspected to be pathogenic in the unaffected parents and sibling to confirm heterozygosity. In silico bioinformatics analysis of the putative candidate variants to assess their pathogenicity was performed. The study revealed a novel putative pathogenic variant in intron 5 of *SFRP4* (c.855+4delAGTA) in a homozygous state, in the South African patient with PD. The parents were both heterozygous for the same mutation. The novel variant was present in the splice junction, and in silico prediction provided evidence that may affect the donor site causing aberrant splicing, and therefore pathogenicity. This study has reported for the first time the implication of a mutation in the *SFRP4* gene in an African patient presenting with PD and adds to the evidence that PD is seemingly due to a mutation in the single genes in many world populations. Understanding PD molecular pathophysiology may provide insight and a better understanding of bone remodelling and possible target for new therapeutic approaches, that will impact the treatment of other skeletal dysplasias including osteoporosis.



## PgmNr 1116: Biallelic hypomorphic variants p.Arg1133Cys and p.Arg1379Cys in *COL2A1*: Further delineation of phenotypic spectrum of recessive type 2 collagenopathies.

### Authors:

G.S. Bhavani <sup>1</sup>; H. Shah <sup>2</sup>; A. Moirangthem <sup>1</sup>; A. Shukla <sup>1</sup>; O.H. Kim <sup>3</sup>; G. Nishimura <sup>4</sup>; G.R. Mortier <sup>5</sup>; K.M. Girisha <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India; 2) Department of Orthopedics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India; 3) Department of Pediatric Radiology, Woorisoa Children's Hospital, Seoul, Republic of Korea; 4) Center for Intractable Diseases, Saitama Medical University Hospital, Saitama, Japan; 5) Center for Medical Genetics, University of Antwerp/Antwerp University Hospital, Antwerp, Belgium

The phenotypic spectrum of type 2 collagenopathies ranges from lethal achondrogenesis type 2 to milder osteoarthritis with mild chondrodysplasia and all of them are monoallelic except recent two reports on association of biallelic variants in *COL2A1* with spondyloepiphyseal dysplasia congenita in two children.

Two affected brothers aged 26 and 24 years were evaluated for short stature (heights -11 and -10 SD), micromelia, lumbar hyperlordosis and joint hyperlaxity. Platyspondyly, hip dysplasia, short femoral necks, metaphyseal irregularities, short and broad metacarpals, metatarsals and phalanges of both hands and feet were noted on radiographs. Exome sequencing of both siblings revealed a biallelic pathogenic variant, c.4135C>T(p.Arg1379Cys) in exon 53 of *COL2A1*. Healthy, consanguineous parents were heterozygous carriers.

In the second family, proband had difficulty in walking and mild scoliosis by age 7 years. He had pectus carinatum, kyphosis, leg length discrepancy and mild joint laxity. His 10 years-old sister had restricted hip movements and hypermetropia. Platyspondyly, coxa valga, epi-metaphyseal dysplasia, delayed carpal ossification and generalized osteopenia were noted in both. They had a novel homozygous pathogenic variant, c.3190C>T(p.Arg1133Cys) in exon 47 of *COL2A1*. The variant segregated in Mendelian fashion and carriers of the variant were unaffected.

MutationTaster, PolyPhen2 and SIFT predict these variants to be disease causing. p.Arg1379Cys is observed in heterozygous state in 6 individuals in gnomAD with a frequency of 0.00002387 and p.Arg1133Cys is not reported in gnomAD. Both variants were absent in in-house database of 623 exomes.

Though all six patients from four families exhibit a spondylo-epimetaphyseal dysplasia, they demonstrate a wide variation in severity of short stature and involvement of epiphyses, metaphyses and vertebrae. The variant, p.Arg1133Cys (family 2) in the triple helical domain is associated with a mild, p.Arg437Trp (Tham *et al.*, 2015) and p.Pro458Leu (Barat-Houari *et al.*, 2015) in the same domain with moderate and p.Arg1379Cys (family 1) in C-propeptide domain with severe phenotype. We hypothesize the variants are likely to be hypomorphic, given the underlying mechanisms of disease causation for known heterozygous variants in *COL2A1*. With this report, we describe additional four individuals with homozygous hypomorphic variants in *COL2A1* and expand the

phenotypic spectrum of recessive type 2 collagenopathy.

## PgmNr 1117: A case report of recessive aggrecan mutations in a girl with spondyloepiphyseal dysplasia.

### Authors:

E. Carter; J. Davis; C. Raggio

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center Skeletal Dysplasias, Hospital for Special Surgery, New York, New York.

---

Skeletal dysplasias are a heterogeneous group of >400 disorders impacting the size and shape of the skeleton. The aggrecanopathies comprise a single group within the current skeletal dysplasia nosology. We report a child with compound heterozygous *ACAN* mutations and a skeletal dysplasia phenotype less severe than what is reported in the one family with SEMD Aggrecan type (OMIM 612813) but more severe than the changes described in the one family with SED Kimberley type (OMIM 608361). An 11 year-old girl with a history of short stature (height -7 SD), scoliosis, joint abnormalities, and brachydactyly was referred to our service for work-up and genetic analysis. Prior consultations had included a differential diagnosis of acrodysostosis versus a FilaminA disorder. Family history was significant in that both the mother (4'9" tall) and the father (5'3" tall) had a history of chondromalacia patellae. Radiographic review detected short stature, platyspondyly, mild scalloping of the posterior aspects of the lumbar vertebral bodies, thoracolumbar scoliosis, epiphyseal changes, and genu valgum with an osteochondral defect in the left knee. Exome sequencing detected two compound heterozygous mutations in the *ACAN* gene: p.S1396LfsX5 and p.A171V. Our proband's mother was found to carry the p.A171V mutation; paternal DNA was unavailable for confirmatory testing. *ACAN* encodes the protein aggrecan, which plays an important role in development and function of articular cartilage. Aggrecan-related bone disorders include a variety of phenotypes, ranging from dominant idiopathic short stature to recessive spondyloepimetaphyseal dysplasia (SEMD), aggrecan type, which is characterized by marked short stature (66-71 cm final height). We present clinical and X-ray characteristics of a girl with an intermediate skeletal dysplasia phenotype that falls between the recessive, more severe SEMD Aggrecan type and the dominant, milder SED Kimberley. Although the specific mutations detected in our family have not been reported before, we believe they cause our proband's skeletal dysplasia findings, and speculate that carrier status impacts final height in other family members.

## PgmNr 1118: Loss of *Fkbp10* in murine tendons and ligaments recapitulates key features of Bruck syndrome.

### Authors:

J. Lim <sup>1</sup>; C. Lietman <sup>1</sup>; J. Rai <sup>2</sup>; M. Weis <sup>2</sup>; D. Park <sup>1</sup>; D. Eyre <sup>2</sup>; D. Krakow <sup>3</sup>; B. Lee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor College of Medicine, Houston, TX; 2) University of Washington, Seattle, WA; 3) David Geffen School of Medicine at UCLA, Los Angeles, CA

---

Osteogenesis Imperfecta (OI) is a heterogeneous genetic disorder that is characterized by low bone mass, increased susceptibility to fractures and progressive bone deformities. Recently, we and others have reported that loss-of-function mutations in FK506 binding protein 10, 65 kDa (*FKBP10*) can cause progressively deforming OI, with a subset of patients exhibiting joint contractures, also known as Bruck Syndrome. To determine if *Fkbp10* in tendons and ligaments is required for proper joint formation, we genetically ablated *Fkbp10* using *Scx-Cre* transgenic mice. Interestingly, conditional removal of *Fkbp10* in *Scx-Cre*-expressing cells induced hindlimb contractures associated with significant gait defects. In contrast, *Fkbp10* deletion in *Col2-Cre*-expressing cells did not cause hindlimb contractures, indicating that *Fkbp10* may be dispensable in chondrocytes. Furthermore, *Scx-Cre*-mediated deletion of *Fkbp10* markedly reduced telopeptide lysyl hydroxylation of type I collagen and collagen cross-linking. Histological analyses of the knee and ankle joints at multiple postnatal stages revealed that *Fkbp10* mutant mice develop ectopic chondroid lesions in tendons and ligaments, which subsequently undergo heterotopic ossification. Furthermore, mRNA-seq of the developing chondroid lesions in *Fkbp10* mutant mice showed significant enrichment of extracellular matrix organization genes, which overall resulted in increased expression of fibrillar collagens, matrix metalloproteases and small leucine rich proteoglycans. Taken together, our mouse model recapitulates several key features of Bruck Syndrome that may provide a valuable tool for identifying novel therapies.

## **PgmNr 1119: X-linked hypophosphatemia: All eight individuals representing separate American families carrying the *PHEX* 3'UTR mutation c.\*231A>G tested positive for an exon 13-15 duplication.**

### **Authors:**

S. Mumm <sup>1,2</sup>; M. Huskey <sup>1</sup>; S. Duan <sup>1</sup>; V. Wollberg <sup>2</sup>; P. Smith <sup>2</sup>; V. Bijanki <sup>2</sup>; G.S. Gottesman <sup>2</sup>; M.P. Whyte <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Division of Bone and Mineral Diseases, Washington University School of Medicine, St Louis, Missouri.; 2) Center for Metabolic Bone Disease and Molecular Research, Shriners Hospital for Children, St. Louis, Missouri

---

In 2008, a single nucleotide change (c.\*231A>G) located 3-bp upstream of the putative polyadenylation signal in the 3'UTR of *PHEX* was associated with X-linked hypophosphatemia (XLH) in seven patients from six unrelated Midwest USA families (Bone 43: 663-6, 2008). In 2015, we reported (JBMR 30: 137-43, 2015) this mutation in six apparently sporadic instances of XLH in Midwest children, but then found mothers carrying the 3'UTR mutation who remarkably appeared unaffected or mildly affected by XLH. Subsequent investigation among > 300 patients with XLH identified several multigenerational American families harboring this unique *PHEX* mutation. Their XLH was distinctly less severe than for those patients carrying a mutation that disrupts the *PHEX* coding region. Furthermore, females compared with males with the 3'UTR defect seemed still more mildly affected, and were often untreated or even undiagnosed with XLH (JBMR 33 Suppl 1: 425-6, 2018). In 2014, a boy with autism and hypophosphatemic rickets (a second cousin to a 3'UTR mutation patient) was referred with copy number microarray performed commercially, showing a copy number gain of *PHEX* exons 13-15. Therefore, we performed copy number real-time qPCR for exons 12-16, (primers failed for exon 13) in four females and four males from apparently “unrelated” families with the 3'UTR mutation. Controls were three sex-matched XLH patients with coding region *PHEX* mutations and a healthy male and female. The six XLH patients with conventional mutations or unaffected controls showed the appropriate copy number for all *PHEX* exons (one copy for males and two for females). In contrast, all eight subjects with the 3'UTR mutation exhibited one extra copy of *PHEX* exons 14 and 15 in their genomic DNA. Thus, the duplication likely resides within or near *PHEX* but its precise location and orientation remain unknown. It may be that all individuals carrying the 3'UTR defect harbor this exon 13-15 duplication. If so, their mild form of XLH would reflect a shared “founder”. However, it also remains to be determined whether this duplication disrupts normal translation or function of *PHEX*. Perhaps the mild XLH phenotype is not from the 3'UTR mutation, but instead due to the exon 13-15 duplication rendering the 3'UTR change a benign variant that co-segregates. Alternatively, the duplication may be a benign variant. Functional assays for *PHEX* mRNA stability and protein function will clarify the etiology of this mild form of XLH.

## PgmNr 1120: Osteogenesis Imperfecta fractures show delayed healing and increased possibility of re-fracture in murine models.

### Authors:

J. Zieba <sup>1</sup>; E. Munivez <sup>1</sup>; C. Ambrose <sup>2</sup>; A. Castellon <sup>1</sup>; M.M. Jiang <sup>1</sup>; B. Lee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, Texas.; 2) Orthopaedic Surgery, UTHealth, Houston, Texas

---

Osteogenesis Imperfecta (OI) is the most common genetic bone dysplasia that is phenotypically and genetically complex. It is characterized by bone deformities and fractures caused by low bone mass and impaired bone quality. Roughly 85-90% of cases are dominantly inherited and result from mutations in genes encoding type I collagen (*COL1A1* and *COL1A2*), the major protein of the bone matrix. 10-15% of OI cases are recessively inherited and the majority result from mutations in members of the prolyl-3-hydroxylation complex including Cartilage Associated Protein (CRTAP). OI patients are at an increased risk of fracture throughout their lifetimes and anecdotal evidence suggests successful fracture recovery. However, non-union has been reported in 24% of fractures and 52% of osteotomies and many stabilization techniques result in additional surgery due to re-fracture. Additionally, re-fractures typically go unreported making the frequency of re-fractures in OI patients unknown. Thus, there is an unmet need to better understand the mechanisms by which OI affects fracture healing. It is our hypothesis that OI fractures undergo suboptimal healing and that this process results in ultimately weaker bone leading to the increased possibility of re-fracture. Using an open tibial fracture model, we show a decrease in callus size in both *Col1a2*<sup>G610C/+</sup> and *Crtap*<sup>-/-</sup> OI mouse models post-fracture suggesting impaired healing. Callus cartilage distribution is altered in the OI models suggesting alterations in cartilage formation or differentiation and may be associated with the observed decrease in callus size. Fracture calluses in both models also exhibited a significant decrease in polar moment of inertia (pMOI) indicating a decrease in resistance to torsional stress. Most importantly, via three-point-bending biomechanical testing, we determined that in wild type bone, the healed fracture site resulted in stronger bone compared to the unfractured tibia. However, *Crtap*<sup>-/-</sup> healed fractured tibia show a trend towards being mechanically weaker and exhibiting decreased ductility when compared to the contralateral unfractured bone. This implies the possibility that OI fractures do not heal properly and may be a prime location for re-fracture. These data provide valuable insight into the effect of the ECM on fracture healing, a poorly understood mechanism, as well as support the need for treatment of primary fractures to decrease incidence of refracture and deformity in OI patients.

## PgmNr 1121: Novel pathogenic variants support constitutive dysregulation of the noncanonical WNT pathway gives rise to Robinow syndrome.

### Authors:

C. Zhang<sup>1</sup>; J.F. Mazzeu<sup>2,3</sup>; B.J. Shayota<sup>1,4</sup>; J. White<sup>5</sup>; Z.C. Akdemir<sup>1</sup>; K. Domic<sup>6</sup>; S. Banka<sup>7</sup>; S.N. Jhangiani<sup>8</sup>; L. Potoki<sup>1,4</sup>; D.M. Muzny<sup>8</sup>; J.E. Posey<sup>1</sup>; R.A. Gibbs<sup>8</sup>; J.R. Lupski<sup>1,4,8,9</sup>; V.R. Sutton<sup>1,4</sup>; C.M.B. Carvalho<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, BCM, Houston, TX; 2) University of Brasilia, Brasilia, Brazil; 3) Robinow Syndrome Foundation, Anoka, MN; 4) Texas Children's Hospital, Houston, TX; 5) University of Washington, Seattle, WA; 6) Department of Pediatric Endocrinology and Diabetes, University Clinical Center Zagreb, Croatia; 7) The University of Manchester, UK; 8) Human Genome Sequencing Center, BCM, Houston, TX; 9) Department of Pediatrics, BCM, Houston, TX

---

Robinow syndrome (RS) is a genetically heterogeneous disorder characterized by skeletal dysplasia and a distinctive facial appearance. RS can segregate as an autosomal-recessive (AR) trait (RRS [MIM 268310]) or autosomal-dominant (AD) trait (DRS [MIM 180700]). Previous studies have revealed locus heterogeneity and identified several genes that can cause RS; including *DVL1*, *DVL3*, *FZD2*, *WNT5A* and *NXN*. The aforementioned Robinow genes identified thusfar all play a role in the WNT/planar cell polarity (PCP) signaling pathway. We performed Sanger sequencing and exome sequencing (ES) on a cohort of 12 subjects from 10 families with RS or a RS-like phenotype. Two unrelated individuals were identified to have pathogenic variants in *DVL1*: a recurrent single base pair deletion (c.1519delT;Trp507Glyfs\*142), and a single nucleotide deletion (c.1556del;p.Gly519Aspfs\*130). Two subjects had previously-reported pathogenic variants *FZD2*: a missense variant (c.1300G>A;p.Gly434Ser) and a nonsense variant (c.1644G>A;p.Trp548\*). In addition, three likely pathogenic variants in *WNT5A* were found: a private missense variant (c.461G>T; p.Cys154Phe) in both a patient and his affected father, and a heterozygous missense variant in exon 3 (c.248G>A; p.Cys83Tyr) in an unrelated individual. One subject had variants in *NXN*, including a homozygous nonsense variant (c.817G>A; p.Gln273\*) *in trans* with a ~1Mb telomeric deletion on chromosome 17p containing *NXN*. Such clinical findings can elicit a range of potential differential diagnoses, including other skeletal disorders. In our cohort, pathogenic variants in Aarskog-Scott syndrome (MIM: 305400) related genes were found in three individuals: two hemizygous frameshift variants (c.367del; p.Leu123\*) and a hemizygous 1bp deletion in *FGD1* (c.1422del; p.Phe474Leufs\*34). Plus, we found one subject with a heterozygous missense variant in *PPP1CB* that causes Noonan syndrome-like disorder with loose anagen hair 2 (MIM: 617506) (c.146C>G; p.Pro49Arg). Note that PP1C can stabilize Dvl and sustain Wnt signaling in mammalian cells and zebrafish embryos. In conclusion, individuals in our cohort were found to carry pathogenic or likely pathogenic variants in genes associated with either RS or disorders with phenotypic overlap with RS. Our findings confirm that constitutive dysregulation of the noncanonical WNT pathway in humans gives rise to RS, which has deepened our insight into underlying mechanisms of the skeletal development and WNT signaling.



## PgmNr 1122: *GNPNAT1* variant causes a novel form of rhizomelic skeletal dysplasia due to impaired chondrocyte differentiation.

### Authors:

S. Naz <sup>1</sup>; N. Ain <sup>1,2</sup>; M. Baroncelli <sup>3</sup>; T. Ishaq <sup>1</sup>; O. Nilsson <sup>3,4</sup>; O. Makitie <sup>2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Biological Sciences, University of the Punjab, Lahore, Pakistan; 2) Department of Molecular Medicine and Surgery and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 3) Pediatric Endocrinology Unit and Center for Molecular Medicine, Department of Women's and Children's Health, Karolinska Institutet and University Hospital, Stockholm, Sweden; 4) School of Medical Sciences, Örebro University and Örebro University Hospital, Örebro, Sweden; 5) Children's Hospital, University and Helsinki University Hospital, Helsinki, Finland

Glucosamine 6-phosphate N-acetyltransferase GNPAT1 catalyzes transfer of an acetyl residue to D-Glucosamine 6-phosphate in the second step of a four-step pathway yielding UDP-N-acetyl-alpha-D-glucosamine (UDP-GlcNAc). UDP-GlcNAc participates in both N- & O- linked glycosylation of proteins in concert with glycosyltransferases. We identified a consanguineous family with four individuals affected with severe skeletal dysplasia characterized by marked rhizomelia, platyspondyly and short stature with adult heights ranging from 99 to 114 cm (SD -7 to -9.6). Whole genome sequencing was completed for two affected and two unaffected individuals. Computational modeling to predict the variant's effect on GNPAT1 was completed by I-Mutant Suite on the protein crystal structure. To explore its biological role, *Gnpnat1* was knocked down by small interfering RNA (siRNA) in rat epiphyseal chondrocytes and the effects on cellular proliferation, differentiation and apoptosis were studied. Analysis of the whole genome sequencing data revealed a homozygous missense variant in *GNPNAT1* (OMIM 616510) in all affected individuals. The GNPAT1 variant affects an amino acid conserved in all multicellular species examined except mosquitoes. The predicted pathogenic variant was absent from public databases and in DNA of 190 ethnically matched controls. *GNPNAT1* is ubiquitously expressed and the encoded protein is localized to cytoplasmic organelles including endoplasmic reticulum and Golgi bodies. The variant affects the acetyltransferase domain of the protein. Computational modeling of the crystal structure predicted a large decrease in stability (DDG Value Prediction, -0.70 Kcal/mol) of the enzyme due to the effect of variant on GNPAT1. Knockdown of *Gnpnat1* in rat epiphyseal chondrocytes did not have an effect on apoptosis. Cell proliferation was somewhat impaired. However, significantly decreased expression of chondrocyte differentiation markers (*Acan*, *Alpl*, *Col2a1*) was observed, suggesting impaired epiphyseal chondrocyte differentiation in the absence of GNPAT1. Variants in *GFPT1* (OMIM 138292) and *PGM3* (OMIM 172100), encoding enzymes involved in the same pathway as GNPAT1, have previously been associated with different human genetic disorders. Our findings implicate a *GNPNAT1* variant in a severe form of skeletal dysplasia and suggest that the disease represents a novel disorder of glycosylation. Funded by Swedish Research Council, Sweden & Higher Education Commission, Pakistan.



## PgmNr 1123: *IFT172* mutation as the nineteenth cause of oral-facial-digital syndrome.

### Authors:

M. Yamada <sup>1</sup>; T. Uehara <sup>1</sup>; H. Suzuki <sup>1</sup>; T. Takenouchi <sup>2</sup>; N. Morisada <sup>3</sup>; K. Tominaga <sup>4</sup>; M. Onoda <sup>5</sup>; K. Kosaki <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.; 2) Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan; 3) Department of Clinical Genetics, Hyogo Prefectural Kobe Children's Hospital, Kobe, Hyogo, Japan.; 4) Department of Cardiology, Hyogo Prefectural Kobe Children's Hospital, Kobe, Hyogo, Japan; 5) Department of Plastic Surgery, Hyogo Prefectural Kobe Children's Hospital, Kobe, Hyogo, Japan

---

The defects in cilia structures or function cause disease expanding multiple organs associated with cilia, that is called ciliopathy. Ciliopathy is classified into subgroups by its clinical features including Joubert syndrome, Bardet-Biedl syndrome, short-rib thoracic dysplasia, and Oral-facial-digital syndrome (OFD). *IFT172* plays a critical role in intraflagellar transport and its mutations cause various ciliopathy-related syndrome phenotype but not OFD. Herein, we document a male patient with OFD syndrome phenotype who had biallelic mutations in *IFT172* gene. Clinical features compatible with the diagnosis included post-axial polydactyly, laterality defects of the heart tube (i.e., single atrium, an atrioventricular septal defect, and a persistent left superior vena cava), and multiple oral frenula. The patient had compound heterozygous variants c.39+5G>A and c.1478T>G (NM\_01566.2): The former variant changes the guanine nucleotide that is critical for normal recognition of the splicing donor site, and the latter variant changes Leu residue that is highly conserved down to *Chlamydomonas reinhardtii*. This is the first report of biallelic *IFT172* mutations who had an OFD syndrome phenotype. Hence, *IFT172* represents the 19th OFD gene. The documentation of a mutation in *IFT172*, in addition to *IFT57*, confirms the IFT pathway is one of the pathogenetic mechanism for OFD.

## PgmNr 1124: The phenotypical spectrum of BRD4 defects.

### Authors:

C. Thiel <sup>1</sup>; N.N. Hauer <sup>1</sup>; C. Vogl <sup>1</sup>; B. Popp <sup>1</sup>; C. Buettner <sup>1</sup>; S. Uebe <sup>1</sup>; H. Sticht <sup>2</sup>; A.B. Ekici <sup>1</sup>; P. Klinger <sup>3</sup>; C. Kraus <sup>1</sup>; U. Trautmann <sup>1</sup>; A. Wiesener <sup>1</sup>; H.G. Doerr <sup>4</sup>; A. Reis <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Human Genetics, Friedrich-Alexander-University of Erlangen-Nuremberg FAU, Erlangen, Germany; 2) Institute of Biochemistry, Friedrich-Alexander-University Erlangen-Nuremberg FAU, Erlangen, Germany; 3) Department of Orthopaedic Rheumatology, Friedrich-Alexander-University Erlangen-Nuremberg FAU, Erlangen, Germany; 4) Department of Pediatrics and Adolescent Medicine, Friedrich-Alexander-University Erlangen-Nuremberg FAU, Erlangen, Germany

The growth of an individual is a variable and highly heterogeneous trait. We performed WES in individuals with short stature. In 2 independent non-consanguineous families we identified 2 heterozygous missense *de novo* variants in *BRD4* (c.1856G>T; c.2513A>T). Screening for further individuals with *BRD4* variants, we identified an individual with a *de novo* frameshift (c.2728delC) variant.

In the 4 previously reported patients with frameshift and missense variants in *BRD4* a Cornelia de Lange-like phenotype, growth deficit, microcephaly and intellectual disability was described. In our study, the patient with the frameshift variant and the patient with the c.1856G>T variant both presented with intellectual disability and Cornelia de Lange-like facial gestalt. Microcephaly was not present in the patient with the frameshift variant. Short stature seems to be a more consistent feature in individuals with missense variants.

*Bromodomain Protein 4* (*BRD4*) is a member of the bromodomain (BET) protein family involved in binding to hyperacetylated genomic regions of promoters and enhancers. *BRD4* mediates CDK9 activity to influence transcription elongation by RNA polymerase II and thus regulates gene expression, cell differentiation and cell cycle. Expression analysis showed that the missense variants lead to reduced expression levels. As the frameshift variant does not alter *BRD4* expression levels we propose a potential shorter, functionally impaired protein. These results indicate a loss-of-function effect of all three identified variants. Immunofluorescence analysis of cells from both patients with missense variants confirmed a significantly reduced cellular growth and smaller cell size when compared to controls. scRNASeq Analysis in HEK293 cells with CRISPR/Cas9 mediated *BRD4* defects highlighted a complex dysregulation of effector proteins (CDK9, CCND1/2, PCNA, DNTTIP1/2, TOP2A and SPA-1).

In conclusion, our findings expand the clinical spectrum of mutations in *BRD4* ranging from idiopathic short stature without distinct facial gestalt to a Cornelia de Lange-like phenotype. Furthermore, we confirm the proposed loss-of-function effects of *BRD4* mutations. We hypothesize that the variability of the phenotype might in part be explained by different effects of the variants on target effector genes.

## **PgmNr 1125: MPS-specific physical symptom score (PSS) and adaptive functions in MPS IVA: A cross sectional study.**

### **Authors:**

A. Ahmed <sup>1</sup>; B. Diethelm-Okita <sup>2</sup>; A. Schneider <sup>3</sup>; K. Rudser <sup>4</sup>; J.B. Eisengart <sup>5</sup>; K. King <sup>6</sup>; S. Lewis <sup>7</sup>; C.B. Whitley <sup>8</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Division of Clinical Behavioral Neuroscience Department of Pediatrics, University of Minnesota, 353-25 717 Delaware St. S.E. Minneapolis, MN 55414; 2) Pediatric Gene Therapy, 13-146 PWB, 516 Delaware St SE Minneapolis, MN 55455; 3) 13-145 PWB, Department of Pediatrics, 420 Delaware St SE Minneapolis, MN 55455; 4) Room 134, Department of Biostatistics 717 Delaware Street SE Minneapolis, MN 55414; 5) Room 381, Department of Pediatrics 717 Delaware St SE Minneapolis MN 55414; 6) Department of Pediatrics, University of Minnesota 420 Delaware St SE Minneapolis, MN 55455; 7) Department of Pediatric Neuropsychology, 3rd floor 353-36 717 Delaware St SE 1932K MPLS, MN 55414; 8) Pediatric Gene Therapy, Department of Pediatrics, University of Minnesota, 13-146 PWB 516 Delaware St SE Minneapolis, MN 55455

---

**Introduction:** MPS IV or Morquio syndrome is an inherited autosomal recessive lysosomal storage disease caused by deficiency in N-acetylgalactosamine-6-sulphate phosphatase, encoded by GLANS gene mapped to 16q24. Previously an MPS-specific physical symptom scale was developed and introduced a Physical Symptom Score (PSS) (PMC 4864710) by quantifying the somatic disease burden in MPS I, II and VI. Now we investigate the association of PSS with adaptive functions in nine MPS IV patients from a cross-sectional study (NIH-U54NS065768).

**Hypothesis:** The goal of our study was to investigate the physical symptoms in MPS IV patients who have been treated with enzyme replacement therapy (ERT) and to examine the relationship of physical symptoms with adaptive behavior and daily living skills.

**Methods:** Nine patients with MPS IV aged 11.6 to 53.3 years, among them 6 female and 3 male who were enrolled in "Longitudinal Studies of Brain Structure and Functions in MPS Disorders," were recruited from University of Minnesota. Somatic disease burden was measured by the PSS, and adaptive behavior was measured with the Adaptive Behavior Composite standard score and Daily Living Skills (DLS) standard score from the Vineland Adaptive Behavior Scales, Second Edition (VABS-II). Within this sample, 8 MPS IV patients were administered the NIH-developed PROMIS questionnaires to assess pain, anger, anxiety, relationships, support, and fatigue.

**Results and Discussion:** While all nine patients had average adaptive skills, we have found higher/worse PSS score is associated with a lower/worse Vineland composite standard score ( $P=0.021$ ) and DLS score ( $P=0.033$ ). The association between the PSS and age was not statistically significant in MPS IV. The pilot PROMIS data found no significant elevations for MPS IV in areas described above. Limitations include small sample size and lack of control data suggesting the need for further analysis on a larger cohort.

**Conclusion:** Adaptive functions are negatively associated with PSS at a statistically significant level

(See Table 1). Future longitudinal analysis of more patients will provide more detailed information on emotional, social and adaptive outcomes with somatic burden of disease.

Table 1 Linear trend between predictor (predominantly PSS) and adaptive functions

Outcome	Predictor	N	Slope (95% CI)	P-value
PSS	Age (yr)	9	0.04 (-0.03, 0.10)	0.267
Vineland Composite	PSS	9	-7.88 (-14.58, -1.17)	0.021
Vineland DLS	PSS	9	-5.81 (-11.15, -0.46)	0.033

## PgmNr 1126: Mutation spectrum in 147 Chinese patients with craniosynostosis: Findings of a custom sequencing panel study.

### Authors:

J. Chen <sup>1,2</sup>; M. Peng <sup>1,2</sup>; Y. Wu <sup>3</sup>; X. Mu <sup>3</sup>; S. Wang <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Chinese Academy of Sciences Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; 2) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200438, China; 3) Department of Plastic Surgery, Huashan Hospital, Fudan University, Shanghai 200040, China

---

Craniosynostosis (CS), one or more of the cranial sutures in an infant skull prematurely fuses, is a skeletal disease caused by genetic or environmental factors, or interplay of them. Most genetic studies on CS were carried out in Caucasian populations, while only a few case reports were performed in East Asian populations. In our study, we established a Chinese craniosynostosis patient cohort without prior molecular diagnosis. We aimed to assess the impact of a targeted gene sequencing for the genetic diagnosis, and to describe the characteristics of mutation spectrum in Chinese craniosynostosis patients. We firstly designed a panel to sequence exonic regions of 17 genes based on a comprehensive literature review. We then performed panel sequencing and bioinformatic analysis in 147 probands. The confirmation and inheritance of suspected variants were validated in probands and their available parents by Sanger sequencing. Variants were identified according to the ACMG standards, based on silico prediction and Sanger validation. A total of 80 cases were identified with pathogenic or likely pathogenic variants. In total, we identified 43 different variants in 9 genes (*EFNB1*, *ERF*, *FGFR1*, *FGFR2*, *FGFR3*, *POR*, *TCF12*, *TGFBR2*, *TWIST1*). We further showed that the distribution of variants has no significant differences between Chinese and Caucasian patients of Apert, Crouzon and Pfeiffer syndromes. However, in Crouzon patients, the mutation frequency at the C342 and Y340 residues were significantly different between the two populations. In addition, eleven novel variants were identified, which had never been reported in Caucasian patients. In summary, our custom sequencing panel can provide reasonably high-yield diagnosis in Chinese craniosynostosis patients. The differences and novel findings of our study suggested great potential to carry out systematic genetic studies in Chinese craniosynostosis patients.

## PgmNr 1127: Parental gonosomal *COL2A1* mosaicism contributes to intrafamilial recurrence in a family with type 2 collagenopathy.

### Authors:

K. Yamamoto <sup>1,2</sup>; T. Kubota <sup>1</sup>; S. Takeyari <sup>1</sup>; K. Miyata <sup>1</sup>; Y. Nakano <sup>1</sup>; H. Nakayama <sup>1</sup>; M. Fujiwara <sup>1,3</sup>; Y. Ohata <sup>1</sup>; T. Kitaoka <sup>1</sup>; K. Yanagi <sup>4</sup>; T. Kaname <sup>4</sup>; Y. Okada <sup>2</sup>; K. Ozono <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Pediatrics, Osaka university Graduate School of Medicine, Osaka, Japan; 2) Dept. of Statistical Genetics, Osaka university Graduate School of Medicine, Osaka, Japan; 3) The 1st Dept. of Oral and Maxillofacial Surgery, Osaka University Graduate School of Dentistry, Osaka, Japan; 4) Dept. of Genome Medicine, National Center for Child Health and Development, Tokyo, Japan

---

The *COL2A1* gene encodes the alpha-1 chain of procollagen type 2. Mutations in the *COL2A1* gene are associated with several different types of skeletal dysplasia collectively known as type 2 collagenopathies. The main inheritance pattern of type 2 collagenopathies is autosomal dominant, and in some cases, germline mosaicism and gonosomal mosaicism have been reported. Our patients were 6-years-old boy and 4-years-old girl. They had severe short stature (-8.5SD and -10.0SD, respectively), flat nose, and short limbs. Their radiograph showed abnormality of vertebra and epiphyseal metaphyses. Their parents had no phenotypes. We conducted whole exome sequencing (WES) to the family because autosomal recessive or germline mosaicism of skeletal dysplasia such as spondyloepimetaphyseal dysplasia or kniest dysplasia was suspected. The heterozygous *COL2A1* mutation (c.3121 A>G) was found in the two patients, and then we diagnosed type 2 collagenopathy. Next, to investigate whether gonosomal mosaicism was raised within the family, we conducted target deep sequencing to their parents without mutations by WES. About 1% *COL2A1* mosaicism was identified in the mother's tissues. We assumed that the mother's gonosomal mosaicism would arise in the epiblast, the mutation rate in egg cells would be high, and the intrafamilial recurrence rate of the disease would also be higher than by the germline mosaicism. This study suggests the importance of detecting the low levels of somatic mosaicism in parents whose children have a suspicious germline mosaicism, and the usefulness of target deep sequencing to detect them.

## PgmNr 1128: Anabolic effects of nitric oxide on bone revealed by deficiency of argininosuccinate lyase.

### Authors:

Z. Jin <sup>1</sup>; J. Kho <sup>1</sup>; B. Dawson <sup>1</sup>; M. Grover <sup>2</sup>; M.M. Jiang <sup>1</sup>; Y. Chen <sup>1</sup>; D. Palmer <sup>1</sup>; P. Ng <sup>1</sup>; S. Nagamani <sup>1</sup>; B. Lee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030.; 2) Stanford school of medicine, Department of Pediatric Endocrinology, Stanford, CA, 94305.

Argininosuccinate lyase deficiency (ASLD) is the second most common urea cycle disorder. Some of the manifestations in ASLD like neurocognitive deficiencies and hypertension are observed even in individuals without documented hyperammonemia and thus are likely caused by mechanisms other than elevated ammonia. ASL is the only mammalian enzyme capable of synthesizing arginine, the sole precursor for nitric oxide synthase (NOS)-dependent nitric oxide (NO) synthesis. Moreover, we have previously shown that ASL is also required for channeling extracellular arginine to NOS for NO production. Hence, loss of ASL leads to a cell-autonomous deficiency of NO production. ASL deficiency is thus a Mendelian model for cell-autonomous, NOS-dependent NO deficiency. NO is an important signaling molecule that influences a wide range of biological processes including bone metabolism. Previous studies have shown that NO donors may prevent bone loss and fractures in postmenopausal women. However, the function of NO in bone formation remains controversial.

Here, we reported that loss of ASL led to decreased NO production and impaired osteoblast differentiation in osteoblastic cell lines and primary osteoblasts derived from a hypomorphic mouse model of ASLD (*Asl*<sup>Neo/Neo</sup>). Osteoblast-lineage specific *Asl* knockout mice (*Osteocalcin Cre; Asl*<sup>f/f</sup>) resulted in decreased bone mass, which was due to reduced bone formation. Mechanistically, the anabolic effect of NO on osteoblast differentiation may be mediated through glycolysis, as indicated by decreased abundance of glycolytic enzymes and reduced glycolytic rate and capacity in *Asl* deficient osteoblastic lineage cells. Caveolin 1 (*Cav-1*) has been reported to be an endogenous negative regulator of NO synthesis. Therefore, we hypothesized that ablation of *Cav-1* may rescue the low bone mass in *Asl*<sup>Neo/Neo</sup> mice. Indeed, we observed that bone mass, NO production, as well as glycolysis were restored in *Asl*<sup>Neo/Neo</sup>; *Cav-1*<sup>+/-</sup> mice as compared to *Asl*<sup>Neo/Neo</sup> mice. Finally, we generated induced pluripotent stem cells (iPSCs) from an individual with ASLD and low bone mass. Consistently, when iPSCs were differentiated into osteoblasts, the mineralization capacity was impaired from the ASL-deficient osteoblasts as compared to three unaffected controls. Taken together, our findings reveal that ASLD is a unique genetic model for studying NO-dependent osteoblast function, and suggest that NO-glycolysis pathway may be a new target for bone anabolism.

## PgmNr 1129: An emerging ribosomopathy affecting the skeleton due to bi-allelic variations in *NEPRO*.

### Authors:

K.M. Girisha <sup>1</sup>; D.L. Narayanan <sup>1</sup>; A. Shukla <sup>1</sup>; K. Neethukrishna <sup>1</sup>; G.S. Bhavani <sup>1</sup>; H. Shah <sup>2</sup>; G.R. Mortier <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Kasturba Medical College, Manipal Academy of Higher Education, Manipal, India; 2) Department of Orthopedics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India; 3) Center of Medical Genetics, University of Antwerp and Antwerp University Hospital, Antwerp, Belgium

---

Cartilage hair hypoplasia (CHH), anauxetic dysplasia 1 and anauxetic dysplasia 2 are rare metaphyseal dysplasias caused by biallelic pathogenic variants in *RMRP* and *POP1*, which encode the components of RNase-MRP endoribonuclease complex (RMRP) in ribosomal biogenesis pathway. Nucleolus and neural progenitor protein (NEPRO), encoded by *NEPRO* (*C3orf17*), is known to interact with multiple protein subunits of RMRP. We ascertained a six-years-old girl with skeletal dysplasia and some features of CHH. *RMRP* and *POP1* did not harbor any causative variant in the proband. Parents-child trio exomes revealed a candidate bi-allelic variant, c.435G>C, p.(Leu145Phe) in *NEPRO*. Two families with four affected individuals with skeletal dysplasia and a homozygous missense variant, c.280C>T, p.(Arg94Cys) in *NEPRO*, were identified from literature and their published phenotype was compared in detail to the phenotype of the child we described. All the five affected individuals have severe short stature, brachydactyly, skin laxity, joint hypermobility and joint dislocations. They also have short metacarpals, broad middle phalanges and metaphyseal irregularities. Protein modeling and stability prediction showed that the mutant protein has decreased stability. Both the reported variants are in the same domain of the protein. Our report delineates the clinical and radiological characteristics of an emerging ribosomopathy caused by biallelic variants in *NEPRO*.



## PgmNr 1130: The role of COL27A1 in skeletal development: Additional pathogenic variants from Turkish kindreds and mouse modeling of the Steel syndrome Puerto Rican founder allele.

### Authors:

C. Gonzaga-Jauregui<sup>1</sup>; G. Yesil<sup>2</sup>; H. Nistala<sup>1</sup>; A. Gezdirici<sup>2</sup>; Y. Bayram<sup>3,4</sup>; K. Nannuru<sup>5</sup>; D. Pehlivan<sup>3,6</sup>; J. Jimenez<sup>5</sup>; Y. Sahin<sup>7</sup>; I.S. Paine<sup>3</sup>; Z. Coban-Akdemir<sup>3</sup>; A.E. Schlesinger<sup>8,9</sup>; N. Ramirez<sup>10</sup>; A. Santiago<sup>11</sup>; N. Wu<sup>3,12</sup>; K. Nierman<sup>5</sup>; J. Walls<sup>5</sup>; W. Poueymirou<sup>5</sup>; C.J. Siao<sup>5</sup>; V.R. Sutton<sup>3,8</sup>; J.E. Posey<sup>3</sup>; A. Economides<sup>1,5</sup>; S. Carlo<sup>10,13</sup>; D.H. Tegyay<sup>14</sup>; R.A. Gibbs<sup>3,15</sup>; J.R. Lupski<sup>3,8</sup>; Baylor-Hopkins Center for Mendelian Genetics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY, USA.; 2) Department of Medical Genetics, Bezmi Alem Vakif University Faculty of Medicine, Istanbul, Turkey.; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA.; 4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.; 5) Regeneron Pharmaceuticals Inc., Tarrytown, NY, USA.; 6) Department of Pediatrics, Division of Pediatric Neurology, Baylor College of Medicine, Houston, TX, USA.; 7) Department of Medical Genetics, Necip Fazlî City Hospital, Kahramanmaraş, Turkey.; 8) Texas Children's Hospital, Houston, TX, USA.; 9) Department of Radiology, Baylor College of Medicine, Houston, TX, USA.; 10) Mayagüez Medical Center, Mayagüez, Puerto Rico.; 11) San Jorge Children's Hospital, Santurce, Puerto Rico.; 12) Department of Orthopedic Surgery, Beijing Key Laboratory for Genetic Research of Skeletal Deformity, and Medical Research Center of Orthopedics, Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences; 13) Ponce Health Sciences University, Ponce, Puerto Rico.; 14) New York Institute of Technology, College of Osteopathic Medicine, New York, NY, USA.; 15) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA.

---

Previously we reported the identification of a homozygous p.Gly697Arg variant in *COL27A1* as a founder mutation in Puerto Rico segregating with Steel syndrome (STLS, MIM #615155); a rare osteochondrodysplasia characterized by short stature, congenital bilateral hip dysplasia, carpal coalitions, and scoliosis. Here we report homozygosity and segregation of this variant in five probands from the initial clinical report and an additional family of Puerto Rican descent with multiple affected individuals. These data confirm that the variant allele represents a founder mutation and that it is the molecular cause of the clinically described Steel syndrome. Additionally, we report three novel and distinct homozygous pathogenic variants in *COL27A1* identified in unrelated consanguineous Turkish kindreds where the probands were ascertained due to unusual osteochondrodysplastic phenotypes. We developed an animal model for the p.Gly697Arg orthologous mutation in the murine *Col27a1* that recapitulates some of the major skeletal features observed in Steel Syndrome individuals including reduced length, scoliosis, and a rounded skull shape. Two pathogenic missense variants were functionally characterized *in vitro* and support a cell non-autonomous mechanism of disease. Characterization of the *in vivo* model shows that the phenotype is likely the result of abnormal collagen deposition in the extracellular matrix and disorganization of the proliferative zone of the

growth plate. The hypothesis that carrier states for this autosomal recessive osteochondrodysplasia may contribute to common complex traits is further explored. Our findings augment our understanding of *COL27A1* biology and expand on the allelic architecture that can potentially underlie both rare and common disease phenotypes.

## PgmNr 1131: Respiratory function in the *Crtp*KO mouse model of osteogenesis imperfecta.

### Authors:

R. Morello <sup>1,4,5</sup>; M. Dimori <sup>1</sup>; M.E. Heard-Lipsmeyer <sup>1</sup>; S.D. Byrum <sup>2</sup>; S.G. Mackintosh <sup>2</sup>; J.L. Carroll <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Physiology & Biophysics, Univ Arkansas Med Sciences, Little Rock, Arkansas.; 2) Department of Biochemistry and Molecular Biology, Univ Arkansas Med Sciences, Little Rock, Arkansas.; 3) Department of Pediatrics, Univ Arkansas Med Sciences, Little Rock, Arkansas.; 4) Department of Orthopaedic Surgery, Univ Arkansas Med Sciences, Little Rock, Arkansas.; 5) Division of Genetics, Univ Arkansas Med Sciences, Little Rock, Arkansas.

---

Respiratory disease is a leading cause of mortality in patients with Osteogenesis imperfecta (OI), a congenital disease that causes severely reduced bone mass and is most commonly caused by dominant mutations in type I collagen genes, *COL1A1* and *COL1A2*. Because type I collagen is present in most tissues, the disease is systemic and patients with moderate or severe OI can suffer from respiratory distress at birth, and then declining respiratory function with increased vulnerability to lung infections in adulthood which can lead to fatal pneumonia. Previous studies proposed that impaired respiratory function in OI patients was secondary to skeletal deformities however recent evidence suggested the existence of a primary lung defect. How collagen defects causing OI negatively affect lung function is not understood and poorly studied. Here, we analyzed the lung phenotype of *Crtp*KO mice, a mouse model of recessive OI with defective collagen synthesis, to corroborate our hypothesis that collagen defects are responsible for their compromised lung parenchyma architecture and consequently alter the respiratory pattern and mechanical properties of the respiratory system. Our results show that lung fibroblasts lacking *Crtp* expression synthesize type I collagen with altered post-translation modifications, consistent with those observed in bone and skin of *Crtp*KO mice. These collagen alterations likely weaken the lung tissue and cause the changes in parenchymal lung architecture and alveolar airspace observed in *Crtp*KO mice. Unrestrained whole body plethysmography showed a significant increase in the ratio of inspiratory time over expiratory time (TI/TE) concomitant with an increase of the diaphragm duty cycle (TI/TCT), the primary inspiratory muscle, in *Crtp*KO compared to WT mice. Closed chest measurements using the force oscillation technique (FOT) showed an increased dynamic resistance but no significant changes in compliance or elastance. Further assessments of lung tissue properties indicated no changes in central airways resistance but significant increases in both lung tissue damping and tissue hysteresivity in *Crtp*KO compared to WT mice suggesting an alteration of the alveolar structure and peripheral lung ventilation heterogeneity. This is the first evidence that collagen defects in OI cause primary changes in several respiratory parameters and thus negatively impact lung function with features of both restrictive and obstructive lung disease.

# PgmNr 1132: Characterization of the Robinow syndrome skeletal phenotype with dual-energy x-ray absorptiometry and genotype-phenotype correlations with osteosclerosis.

## Authors:

B.J. Shayota <sup>1</sup>; C. Zhang <sup>1</sup>; R.J. Shypailo <sup>2</sup>; R. Maricevich <sup>3</sup>; P.F. Austin <sup>4</sup>; J.F. Mazzeu <sup>5</sup>; C. Carvalho <sup>1</sup>; V.R. Sutton <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA.; 2) USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 3) Division of Plastic Surgery, Department of Surgery, Baylor College of Medicine, Houston, TX, USA; 4) Department of Urology, Baylor College of Medicine, Houston, TX, USA; 5) Division of Clinical Medicine, Faculty of Medicine, University of Brasília, Brasília, Brazil

## Background

Robinow syndrome (RS) is a genetically heterogeneous skeletal dysplasia with a growing list of associated genes. Recent reports suggest much more phenotypic heterogeneity including a few reports of an osteosclerotic form of RS. We endeavored to investigate the full spectrum of skeletal anomalies in a genetically diverse cohort of RS patients with a focus on the bone structure.

## Subjects & Methods

Seven individuals with molecularly confirmed RS, including four with *DVL1* variants and single individuals with variants in *WNT5A*, *ROR2*, and *GPC4* underwent a musculoskeletal physical exam and dual-energy x-ray absorptiometry (DEXA) scan. Measures in bone mineral content (BMC) and bone mineral density (BMD) of the total body and non-cranial skeleton were taken and compared to height and age matched controls for pediatric and adult subjects.

## Results

Skeletal examination revealed variability in limb shortening anomalies including mesomelia (3/7), rhizomelia (2/7), and micromelia (2/7) as well as limited supination (3/7), genu valgus (1/7), and scoliosis (1/7). DEXA scan measures compared to height-matched controls revealed increased total body BMC (4/7), non-cranial BMC (2/7), total body BMD (6/7), and non-cranial BMD (3/7). DEXA scan measures compared to age-matched controls revealed increased total body BMC (3/7), non-cranial BMC (0/7), total body BMD (3/7), and non-cranial BMD (1/7). Cranial osteosclerosis based on BMD for age was present in 3/7 subjects, one of which also had non-cranial osteosclerosis. One subject had osteopenia.

When comparing the skeletal phenotype based on genotype, cranial osteosclerosis was only observed in *DVL1*-RS (2/4) and *GPC4*-RS (1/1) subjects, while individuals with a *ROR2*-RS and *WNT5A*-RS had normal or low BMD. Notably, the *GPC4*-RS subject's cranial osteosclerosis was complicated by choanal atresia, bilateral conductive hearing loss, and cranial nerve III, VI and VII palsy secondary to stenosis of nerve foramina.

## Conclusion

The spectrum of skeletal anomalies including the micro-architecture of the bones observed in RS has considerable variability with some genotype-phenotype correlations. In particular, osteosclerosis appears to be a feature observed only in those with variants in *DVL1* and *GPC4* based on this study

and prior publications. Additionally, the osteosclerosis has a predilection for the cranium compared to the non-cranial skeleton, which can result in comorbidities like cranial nerve palsy and hearing loss.

## PgmNr 1133: A mouse model of mandibulofacial dysostosis with microcephaly reveals a role for EFTUD2 in the neural tube.

### Authors:

L. Jerome-Majewska <sup>1</sup>; M. Beauchamp <sup>2</sup>; A. Djedid <sup>3</sup>; J. Majewski <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Human Genetics and Anatomy and Cell Biology, McGill Univ, Montreal, Quebec, Canada; 2) McGill University Health Centre, Montreal, Quebec, Canada; 3) Human Genetics, McGill Univ, Montreal, Quebec, Canada

---

Splicing, the removal of introns from pre-messenger RNA is an essential step for expression of most genes in multicellular organisms and for expanding the number of proteins coded by genomes. Exome sequencing has shown that mutations in splicing factors and small nuclear ribonucleoprotein particles (snRNPs) which form core components of the spliceosome, are responsible for craniofacial malformations. Our group is using a mutant mouse line with conditional mutation in *Eftud2*, the gene mutated in mandibulofacial dysostosis with microcephaly (MFDM) to uncover the etiology of this syndrome. *In situ* hybridization revealed strong expression of *Eftud2* in ectodermal and mesodermal components of the future craniofacial region at embryonic days (E) 7.5 and 8.5, and by E9.5 expression was also found in the body wall and developing heart. To examine the requirement for *Eftud2* in neural crest cells, the Wnt1-Cre2 transgenic line was used to delete exon 2 of *Eftud2*. *Eftud2* flox homozygous mutant embryos carrying the Wnt1-Cre2 transgene displayed hypoplasia of the midbrain and pharyngeal arches starting at E9.5. By E11.5, most embryos also had an open neural tube and all embryos showed exencephaly at E14.5. Cartilage preparations revealed an absence of cartilage in the head, reduction/or absence of Meckel's cartilage, and abnormal inner ear development. Since deletion of exon 2 is predicted to generate a truncated protein with partial function, our data suggest that normal levels of *Eftud2* is crucial in neural crest cells for normal craniofacial development. Future studies are focused on elucidating the molecular and transcriptional basis of MFDM using this mouse model.

## PgmNr 1134: Variants in *ANK2* brain-specific transcript as novel cause of holoprosencephaly.

### Authors:

C. Ngongang <sup>1</sup>; P. Kruszka <sup>1</sup>; J.L. Everson <sup>2</sup>; A.F. Martinez <sup>1</sup>; R.J. Lipinski <sup>2</sup>; M. Muenke <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Department of Comparative Biosciences, School of Veterinary Medicine, University of Wisconsin-Madison, Madison, WI

---

Loss-of-function mutations in Ankyrin 2 gene (*ANK2*; OMIM 106410) are known as a cause of human arrhythmias and Long QT syndrome (OMIM 600919). Different Ankyrin isoforms encoded by alternatively spliced transcript variants have been described. This includes a brain-specific 440-kD Ankyrin which contains a 200-kD segment encoded by exon 38, not typically found in cardiac *ANK2* transcripts. Whereas *Ank2* null mice show brain structural defects such as hypoplasia of the corpus callosum, dilated cerebral ventricles, extensive degeneration of the optic nerve with death by postnatal day 21, no association between *ANK2* variants and defects of the human forebrain development, including holoprosencephaly (HPE) has been reported to date.

Here we present two unrelated female patients aged 2 years each, diagnosed with lobar holoprosencephaly by Magnetic Resonance Imaging. Thorough phenotyping of each patient revealed no associated feature in the first patient. The second patient had a single central maxillary incisor and cleft of the lip and palate. No extracephalic features were identified in both patients. Using trio-based exome sequencing, novel *de novo* heterozygous variants in exon 38 of *ANK2* were identified. The first patient had a stop-gain variant c.7804G>T (p.E2602\*) predicted deleterious and loss-of-function intolerant (CADD score: 51; pLI score: 1), and the second patient had a deleterious missense variant c.9848C>T (p.S3283L)(CADD score: 29) in addition to a *de novo* frameshift variant in *SIX3* c.613del (p.Asp205Thrfs\*46). While mutations in *SIX3* alone can cause HPE, the low rate of *de novo* variants and the incomplete penetrance frequently observed for this gene are suggestive of a possible digenic etiology for the second patient. It is anticipated that genes regulating forebrain patterning and involved in the pathogenesis of HPE are expressed in the prosencephalic neural folds that give rise to the forebrain vesicle during primary neurulation. Gene expression analysis in mouse embryos by *in situ* hybridization at GD8.25 demonstrated specific *ANK2* expression in the prosencephalic neural folds.

This study presents for the first time an association between variants in *ANK2* brain-specific transcripts and HPE, revealing *ANK2* gene as a potential regulator of human forebrain development and modulator of HPE pathogenesis.

## PgmNr 1135: The refinement of 16p13.3 microdeletion syndrome from a case presentation of a girl with epilepsy, intellectual disability, and distinctive dysmorphic features.

### Authors:

Y. Kuroda <sup>1</sup>; H. Murakami <sup>1</sup>; Y. Kimura <sup>2</sup>; Y. Enomoto <sup>2</sup>; Y. Tsurusaki <sup>2</sup>; T. Uehara <sup>3</sup>; K. Kosaki <sup>3</sup>; K. Kurosawa <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan.; 3) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

---

Eight patients have been reported with 16p13.3 microdeletion syndrome demonstrating intellectual disability and epilepsy. The responsible region for this phenotype encompassed five genes (*TBC1D24*, *ATP6V0C*, *AMDHD2*, *CEMP1*, and *PDPK1*). Here, we describe an additional patient demonstrating moderate intellectual disability, epilepsy, skeletal abnormalities, and distinctive dysmorphic features with a minimal overlapping region harboring *TBC1D24* and *ATP6V0C*. The patient was a 10-year-old girl. She developed tonic-clonic seizures at 8 months and was started with an antiepileptic drug. Myoclonic seizures began at 7 years, which were controlled with multiple antiepileptic drugs. She started to walk alone at 30 months and speak recognizable words at 7 years. Dysmorphic features were noted including a wide forehead, arched eyebrows, long palpebral fissures, long eyelashes, bulbous nasal tip, short and prominent philtrum, and pointed chin. She had skeletal abnormalities including chest and spine deformities and metacarpophalangeal joint contracture. Copy number variation analysis from trio exome sequencing data detected a *de novo* microdeletion at 16p13.3. Chromosomal microarray analysis revealed a microdeletion of 284 kb at 16p13.3 (from position 2,280,210 to 2,565,055 bp). The present patient delineated the critical region for intellectual disability and epilepsy for 16p13.3 microdeletion syndrome.



## PgmNr 1136: Familial cleidocranial dysplasia: Case, clinical, and radiological correlation.

### Authors:

A. Rivera-Cameras <sup>1,2</sup>; D. García-Cruz <sup>1</sup>; I.M. Salazar-Davalos <sup>3</sup>; A.V. Ruíz - Ramírez <sup>1,2</sup>; M.A. Ramírez-Hernández <sup>1,2</sup>; M.A. Aceves-Aceves <sup>4</sup>; I.P. Dávalos-Rodríguez <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Doctorado en Genética Humana, CUCS, Universidad de Guadalajara, Mexico; 2) División de Genética, CIBO, IMSS; 3) UNIVA; 4) UMAE,HE,CMNO,IMSS

**INTRODUCTION:** Cleidocranial dysplasia, OMIM #119600, is an autosomal dominant disorder characterized by hypoplasia or aplasia of the clavicles with narrow, sloping shoulders that can approximate each other, delayed fusion of cranial sutures, wide-open fontanels at birth that may persist throughout life. It is caused by mutations in *RUNX2* (6p21) involved in differentiation of osteoblasts and bone formation. The prevalence of CDD is 1/1,000,000.

**OBJECTIVE:** Present a family affected with case of CDD and its clinical-radiological correlation.

**CASE REPORT:** Patient 1 (*proposita*): 4 years-old female, product of the 3rd pregnancy. **Physical examination:** weight 15,100kg (pc11), height 104cm (pc21), HC 49cm (pc13), midfacial hypoplasia, hypertelorism, delayed fusion of cranial sutures, wide-open fontanels, hypoplasia of the clavicles with narrow, sloping shoulders that can be approximated one to another; scoliosis and difficulty speaking. **X-ray:** Presence of facio-cranial disproportion, hypoplasia of clavicles, dorsal scoliosis and enlargement of lumbar vertebrae.

**Patient 2** (brother): 9 years-old male, product of 1st pregnancy. **Physical examination:** weight 22,400Kg (pc2), height 125cm (pc4), HC 51cm (pc3), hypertelorism, delayed fusion of cranial sutures, wide-open fontanels, hypoplasia of the clavicles with narrow, sloping shoulders that can be approximated one to another, scoliosis, cognitive and intellectual functions normal. **X-ray:** presence of facio-cranial disproportion, hypoplasia of clavicles, dorsal scoliosis with thickened vertebrae and apparent bifurcation, enlargement of lumbar vertebrae with apparent platyspondyly.

**Patient 3** (mother): 30 year-old female, with familial history of affected mother and half-sister 17 year-old. Physical examination: weight 60kg, height 1.56cm, HC 50cm, broad flat forehead, midface hypoplasia, hypertelorism, hypoplasia of clavicles with narrow, sloping shoulders that can be approximated each other, cognitive and intellectual functions normal. **X-ray:** vertebral anomalies (thickened lumbar vertebrae) and hypoplasia of clavicles.

**CONCLUSIONS:** A family with clinical and radiological manifestations compatible with CDD was studied. The diagnosis was done on basis of clinical and radiological findings. Molecular genetics testing can be used to confirm the diagnosis in patients with atypical clinical and radiological features. Genetic counseling should be provided to affected families, management and treatment is multidisciplinary.

## PgmNr 1137: Variant in *OTUD6B* gene cause intellectual developmental disorder with dysmorphic facies, seizures, and distal limb anomalies: A case report.

### Authors:

J. Rojas Martinez; R. Quero; M. Manotas; M. Guerra; J. Villalta

[View Session](#) [Add to Schedule](#)

**Affiliation:** Instituto de Genética Humana, Facultad de Medicina, Pontificia Universidad Javeriana, Bogotá, Colombia.

---

The syndrome of intellectual developmental disorder with dysmorphic facies, seizures, and distal limb anomalies is a multisystemic disorder of autosomal recessive inheritance with phenotypic variability characterized by microcephaly, global developmental delay, absence of language, hypotonia, growth retardation with prenatal onset, feeding difficulties, structural abnormalities in the brain, and congenital anomalies including heart defects and musculoskeletal abnormalities. It is caused by biallelic mutations in the *OTUD6B* gene that encodes a member of the deubiquitinating enzyme subfamily that contains the ovarian tumor domain (OTU).

We present the case of a 16-year-old male adolescent, product of the second pregnancy of consanguineous parents (first cousins), with severe intellectual disability, focal epilepsy, absence of language, notorious delay in neurodevelopment and bilateral profound sensorineural hearing loss associated with craniofacial dysmorphism due to microcephaly, bilateral palpebral ptosis, right iris coloboma, bifid tongue, winged neck, pectum excavatum, breast hypertelorism, clinodactyly, shortening of the fifth finger in the hands and generalized xeroderma.

Trio whole exome sequencing revealed a pathogenic variant c.433C> T (p.Arg145 \*), in homozygosis in the *OTUD6B* gene, which confirmed the diagnosis in the patient and the status of heterozygous carriers in both parents.

It has been reported fourteen individuals from eight independent families worldwide with homozygous mutations in the *OTUD6B* gene as a cause of the developmental disorder with dysmorphic facies, seizures, and distal limb anomalies. At our knowledge, our patient is the first case reported in America.

## PgmNr 1138: A novel 3 MB deletion on 6p24 excluding *TFAP2A* results in a mild phenotype in a multiplex family with orofacial clefting.

### Authors:

C.G. Willett <sup>1</sup>; A.L. Petrin <sup>2</sup>; E. Massiello <sup>3</sup>; C. Restrepo <sup>4</sup>; L.C. Valencia-Ramirez <sup>4</sup>; T.C. Cox <sup>5</sup>; T. Roscioli <sup>6</sup>; L.M. Moreno-Urbe <sup>2</sup>; M.L. Marazita <sup>3</sup>; E.J. Leslie <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Emory University, Atlanta, GA; 2) University of Iowa, Iowa City, IA; 3) University of Pittsburgh, Pittsburgh, PA; 4) Fundacion Clinica Noel, Medellin, Colombia; 5) University of Missouri Kansas City, Kansas City, MO; 6) Sydney Children' Hospital, Sydney, Australia

---

Nonsyndromic orofacial clefting (OFC) has a complex etiology but multiplex families are likely to be enriched for high effect variants that reflect novel genetic factors or causal mechanisms. In a previous effort to characterize large copy number variants as genetic risk factors for OFC following a genome-wide association study (GWAS), we identified a novel 3 Mb deletion on 6p24 in three affected members of the same family from Colombia. The reported pedigree had a dominant inheritance pattern and included 5 individuals with cleft lip (CL) or cleft lip and palate (CLP): a female proband (CLP), her brother (CLP), her mother (CL), and a maternal aunt (CLP) and uncle (CL). All affected family members with DNA (4 of 5 individuals, including one not typed in the GWAS) carried the deletion, which was inherited from the proband's unaffected grandmother. Exome sequencing of this family found no segregating single nucleotide variants in OFC candidate genes, supporting the pathogenicity of this deletion. The 3Mb deleted region included 12 genes and a gene desert spanning approximately 840 kb. Two of the deleted genes are notable for their associated phenotypes: *DSP* (woolly hair and keratoderma) and *BMP6* (decreased growth), but neither are strong OFC candidates. The 3' breakpoint of the deletion is located in the middle of a conserved 2.5Mb topologically associated domain (TAD) identified from published Hi-C data in human embryonic stem cells and containing *TFAP2A*. In addition to removing the 5' boundary of the TAD, the deletion includes one enhancer active in human cranial neural crest cells and human fetal craniofacial tissue and a second with activity in neural crest in mouse embryos. *TFAP2A* mutations and deletions cause Branchiooculofacial Syndrome (BOFS), a dominant syndrome that includes OFC as a primary phenotypic feature in addition to branchial and ocular anomalies. Except for OFC and characteristic facial features (broad nasal root and slight hypertelorism), the affected members of this family lack additional features of BOFS. Cumulatively, these data suggest that deletion of distal *TFAP2A* regulatory elements result in a craniofacial phenotype consistent with mild features of BOFS and that these elements are excellent candidates for being involved in nonsyndromic OFCs.

## PgmNr 1139: Phenotypic expansion of *KMT2D*-associated disorder: Beyond Kabuki syndrome.

### Authors:

D. Baldrige<sup>1</sup>; R.C. Spillmann<sup>2</sup>; D.J. Wegner<sup>1</sup>; J.A. Wambach<sup>1</sup>; F.V. White<sup>3</sup>; K. Sisco<sup>1</sup>; T.L. Toler<sup>1</sup>; P.I. Dickson<sup>1</sup>; F.S. Cole<sup>1</sup>; V. Shashi<sup>2</sup>; D.K. Grange<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Washington University in St. Louis, St. Louis, MO.; 2) Pediatrics, Duke University Medical Center, Durham, NC; 3) Pathology and Immunology, Washington University School of Medicine, St. Louis, MO

---

Pathogenic variants in *KMT2D*, which encodes lysine specific methyltransferase 2D, cause autosomal dominant Kabuki Syndrome 1 (MIM: 147920), associated with distinctive dysmorphic features that include arched eyebrows, long palpebral fissures with eversion of the lower lid, large protuberant ears, and persistent fetal finger pads. Most disease-causing variants described to date are putative loss-of-function alleles, although 15-20% of cases are attributed to missense variants. We describe here four cases (including one previously published case, Sakata et al. 2017) with de novo *KMT2D* missense variants, and who have shared but unusual clinical findings not typically seen in Kabuki syndrome, including athelia (absent nipples), choanal atresia, hypoparathyroidism, delayed or absent pubertal development, and extreme short stature. Two of the four had severe interstitial lung disease. These individuals also lack the typical dysmorphic features found in Kabuki syndrome. All of these missense variants cluster within a 40-amino-acid region of the protein that is located just N-terminal of an annotated coiled coil domain. These findings significantly expand the phenotypic spectrum of features associated with variants in *KMT2D* beyond those seen in Kabuki syndrome and suggest a possible new underlying disease mechanism in these cases.

## PgmNr 1140: Expanding the spectrum of *KLHL7*, novel mutation, and evidence for multilocus pathogenic variation.

### Authors:

A. Gezirici<sup>1</sup>; D. Pehlivan<sup>2</sup>; J. Punetha<sup>2</sup>; E. Yilmaz Gulec<sup>1</sup>; M. Elmas<sup>3</sup>; S. Isikay<sup>4</sup>; Y. Bayram<sup>5</sup>; Z. Coban Akdemir<sup>2</sup>; J. M. Fatih<sup>2</sup>; S. N Jhangiani<sup>2</sup>; E. Karaca<sup>2</sup>; R. A Gibbs<sup>2,6</sup>; N. H. Elcioglu<sup>7</sup>; J. R Lupski<sup>2,6,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, ISTANBUL, Turkey; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 3) Department of Medical Genetics, Afyon Kocatepe University, Afyon, Turkey; 4) Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, School of Health Sciences, Gaziantep, Turkey; 5) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA; 7) Department of Pediatric Genetics, Marmara University Medical School, Istanbul, Turkey; 8) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA; 9) Texas Children's Hospital, Houston, Texas, USA

**Background:** Mutations in Kelch-like 7 (*KLHL7*) have recently been described to cause a constellation of clinical findings with descriptions of both autosomal recessive Crisponi syndrome (CS)/cold-induced sweating syndrome type 1 (CISS1)-like (CISS3), Bohring-Opitz syndrome (BOS)-like presentation and autosomal dominant retinitis pigmentosa 42 (RP42). CISS3/BOS is mainly characterized by clinical features of severe developmental delay/intellectual disability (DD/ID), hyperthermia, recurrent infections, feeding difficulties, joint contractures, and craniofacial dysmorphism. Heterozygous pathogenic variants in the BACK domain mainly result in RP42 phenotype, whereas biallelic pathogenic variants in the Kelch-repeat domain result in CISS3/BOS.

**Material and Methods:** In this study, we performed exome sequencing and family-based rare variant analyses on 117 families with a shared clinical feature of arthrogryposis.

**Results:** A total of eight affected individuals from four unrelated families were found to have biallelic pathogenic variants in *KLHL7*. The stopgain variant c.565C>T:p.Arg189\* resides in the BACK domain of *KLHL7*, and is novel. The remaining three families had variants that were reported previously. One family showed evidence for multilocus pathogenic variation including rare variants in *HOXA11* and *TNRC6C* in addition to *KLHL7*, each potentially contributing to the proband's phenotype. Subjects from three of the 4 families shared recognizable facial features whereas multilocus pathogenic variation subject had dysmorphic craniofacial features that were distinct from the other 3 families. The syndromic features observed in all individuals were severe DD/ID with brain anomalies (4/4) and arthrogryposis (4/4). Additional features not present in all cases included prematurity (3/4), genitourinary system anomalies (2/4), congenital heart disease (2/4); and nephrolithiasis (1/4), intestinal obstruction (1/4) and cleft palate (1/4), representing a potential expansion of the known clinical features associated with biallelic *KLHL7* pathogenic variants.

**Conclusions:** Our findings confirm that CISS3 is a clinically recognizable disorder with distinctive facial dysmorphic features and we further delineate the syndromic features of this syndrome by describing biallelic variants including one novel potentially pathogenic variant. We further validated that biallelic variants in the *KLHL7* BACK domain can cause CISS3 in addition to RP42.



## PgmNr 1141: Biallelic frameshift variants in *CNKSR1* and a novel deletion in *ZEB2* identified in a patient dysmorphologically diagnosed with Mowat-Wilson syndrome.

### Authors:

T. Kaname <sup>1</sup>; K. Yanagi <sup>1</sup>; M. Takeshita <sup>1</sup>; K. Naritomi <sup>2</sup>; Y. Matsubara <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Gen Med, Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Japan; 2) Okinawa Nanbu Rehabilitation and Medical Center, Naha, Japan; 3) Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Japan

---

Mowat-Wilson syndrome (MOWS) is an autosomal dominant disorder characterized by distinctive facial appearance, intellectual disability, delayed motor development, microcephaly and short stature. It is known that most patients with MOWS have de novo heterozygous mutations in the *ZEB2* gene. We encountered a 3-year-old female patient with distinctive facial appearance (narrow bifrontal diameter, frontal bossing, telecanthus, epicanthal folds, prominent antihelices, prominent ears, uplifted earlobes, prominent nasal bridge), intellectual disability, delayed motor development, hypotonia and short stature. She was clinically diagnosed with MOWS. Since direct sequencing for *ZEB2* coding exons displayed no pathogenic variants in the patient, we performed whole exome analysis for trio in the family.

After routine SNV calling and variant filtering procedure, no pathogenic SNVs in *ZEB2* but biallelic frameshift variants (p.Thr282fs from mother and p.Ala503fs from father) in *CNKSR1* were found in the patient. Recently, it has been reported that a recognizable autosomal recessive intellectual disability was caused by defect of the *CNKSR1* gene. The p.Thr282fs was recurrent and the p.Ala503fs was novel. However, the patient's findings did not match the phenotypes of *CNKSR1* mutations, which were previously reported. In addition, a homozygous variant of p.Pro284HisfsTer74 in *CNKSR1*, which was near the p.Thr282, was registered 224/273,674 in the gnomAD. Thus, copy number analysis was performed using whole exome data, which revealed an approximately 360kb deletion around intron 7 of *ZEB2* in the patient. The deletion was confirmed by the breakpoint analysis.

Based on these findings and data, we concluded that the patient's symptoms were caused by the novel deletion in *ZEB2* rather than biallelic variants in *CNKSR1*.

## PgmNr 1142: Phenotype assessment of a Brazilian Cornelia de Lange syndrome (CDLS) cohort.

### Authors:

V. Kim <sup>1</sup>; H. Aoi <sup>2</sup>; I. Furquim <sup>1</sup>; J.R.M. Ceroni <sup>1</sup>; J.L.P. Verissimo <sup>1</sup>; C. Berlim <sup>1</sup>; D.M.B. Lopes <sup>1</sup>; R.S. Honjo <sup>1</sup>; D.R. Bertola <sup>1</sup>; N. Matsumoto <sup>2</sup>; C.A. Kim <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics Unit, Instituto da Crianca - Hospital das Clinicas - HC-FMUSP, Sao Paulo, Brazil; 2) Department of Human Genetics, Yokohama City University, Yokohama, Japan

---

Introduction: Cornelia de Lange syndrome (CDLS) is a rare genetic disorder caused by mutations in five different genes (*NIPBL*, *SMC1A*, *HDAC8*, *SMC3*, *RAD21*), characterized by dysmorphic features (synophrys, long eyelashes, short nose, hirsutism, microcephaly), limb defects, growth retardation and neurodevelopmental delay. The disorder has a wide clinical variability, thus our objective is to assess the clinical characteristics of 24 Brazilian patients. Materials and Methods: Whole exome sequencing (WES) of the patients and their parents were performed in clinical diagnosis of CDLS. Results: From the 24 cases, pathogenic variants in *NPBL* were found in 22 cases and the other 2 cases were confirmed in *SMC1A*. The main clinical findings in confirmed CDLS patients were: highly arched eyebrows (95.8%), synophrys (91.6%), long curly eyelashes (87.5%), hypertrichosis (79.1%), microcephaly (70.8%), anteverted nostrils (75%), thin upper lip (66.6%), long philtrum (66.6%), downturned corners of the mouth (66.6%), syndactyly (54.1%), abnormal palmar crease (50%), depressed nasal bridge (45.8%), short neck (45.8%), cryptorchidism (41.6%), hypoplastic genitalia (37.5%) and micrognathia (29.1%). Prenatal growth retardation is reported in 79.1%, height or length below 5th percentile for age in 75% and weight below 5th percentile for age in 62.5%. Other secondary features are gastroesophageal reflux (70.8%), epilepsy (33.3%), heart anomalies (25%), hearing loss (25%) and myopia (20.8%). Neurodevelopmental delay were present in 87.5%, intellectual disability in 58.3% and language delay in 70.8%. Behavior problems were reported in all patients, except three, such as attention deficit disorder (75%), anxiety (66.6%) and self-injurious behavior (54.1%), aggression (45.8%). Other findings were abnormal palmar crease (52%), syndactyly (44%) and oligodactyly (12%). The current age varies from 8mo to 40 y (mean 11y5mo, median 8y). Conclusion: The clinical and molecular evaluation were a pioneer in Brazilian CDLS patients and WES is an important exam for these patients and their family for the correct diagnosis, follow-up and genetic counseling.



## PgmNr 1143: Tracheal cartilaginous sleeve in patients with Beare-Stevenson syndrome.

### Authors:

K. Kurosawa <sup>1</sup>; E. Seki <sup>2</sup>; K. Enomoto <sup>1</sup>; K. Tanoue <sup>2</sup>; M. Tanaka <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, Kanagawa Children's Med Ctr, Yokohama, Japan; 2) Department of General Medicine, Kanagawa Children's Med Ctr, Yokohama, Japan; 3) Department of Pathology, Kanagawa Children's Medical Ctr, Yokohama, Japan

---

Tracheal cartilaginous sleeve (TCS), a life-threatening complication in patients with craniosynostosis syndromes, is a solid cartilaginous tube lacking the pars membranacea, caused by defective C-shaped tracheal rings. Without tracheostomy, it has a 90% mortality by 2 years. To increase knowledge regarding TCS and Chiari malformation for respiratory distress in Beare-Stevenson syndrome (BSS), we report 2 cases with BSS and TCS. Patient 1 was the second child of healthy Japanese parents. After birth, she was intubated for severe respiratory failure caused by airway obstruction and choanal stenosis. Chest computed tomography (CT) and bronchoscopy showed dorsal recession of the tracheal internal wall, suggesting TCS. At 5 years, she developed severe sleep apnea. Although bronchoscopy showed improvements in the tracheal wall deformity, spinal magnetic resonance imaging revealed a hyper-intense signal at the C2 level. Genetic analysis revealed a pathogenic FGFR2 variant c.1124A>G (p.Tyr375Cys). Patient 2 was the third child of healthy parents. At age 10, he developed obstructive respiratory distress. Chest CT showed tracheostomy tube obstruction by a TCS-related dented deformity. At 12 years of age, he died suddenly after respiratory distress caused by tracheostomy tube obstruction. The tip of the tracheostomy tube had caused tracheal wall injury resulting in the formation of a dented deformity. Stacking of the tracheostomy tip into the dented deformity was considered the likely cause for the respiratory distress. Genetic analysis identified a pathogenic FGFR2 variant c.1115C>G (p.Ser372Cys). Early interventions with tracheostomy and cervical cord decompression improved severe respiratory distress. This shows that the mechanism and severity of TCS vary, and early airway evaluation is critical in craniosynostosis syndromes. To our knowledge, this is the first report demonstrating an abnormal tracheal structure due to long-term injury by the distal tip of the tracheostomy tube, leading to airway obstruction in BSS. Our report delineates the potential risks of tracheostomy tube obstruction by TCS and emphasizes the importance of investigations with bronchoscopy for BSS patients.

## PgmNr 1144: Clinical features of individual with Birk Barel syndrome.

### Authors:

E. Nishi <sup>1</sup>; Y. Mishima <sup>1</sup>; Y. Kishimoto <sup>2,3</sup>; T. Kaname <sup>2</sup>; N. Okamoto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Osaka Women's and Children's Hospital, Osaka, Japan.; 2) Department of Genome Medicine, National Research Institute for Child Health and Development, Tokyo, Japan; 3) Medical Genome Center, National Research Institute for Child Health and Development, Tokyo, Japan

---

Birk-Barel syndrome also known as *KCNK9* imprinting syndrome is characterized by congenital central hypotonia, severe feeding difficulties, delayed development, intellectual disability, and dysmorphic features including dolichocephaly with bitemporal narrowing, short philtrum, tented upper lip, palatal abnormalities, and small mandible. It is caused by heterozygous mutation in *KCNK9* on chromosome 8q24. The diagnosis of the *KCNK9* imprinting syndrome is established in a proband with suggestive clinical findings and detection of the heterozygous *KCNK9* pathogenic variant p.Gly 236 Arg (G236R) on the maternal allele. To date 19 individuals with a molecularly confirmed diagnosis have been reported: 15 from an Arab-Israeli family and four cases with different pedigrees [Barel et al., 2008, Graham et al., 2016].

Here we describe a Japanese individual with *KCNK9* imprinting syndrome caused by recurrent mutation in *KCNK9*. The individual was 2 years and 7 months old girl, the third child of healthy 28-year-old mother and 38-year-old father. She was born at the normal delivery at 40 2/7 weeks of gestation. Apgar scores were 7 and 9 at 1 minute and 5 minutes, respectively. At birth, profound hypotonia, high palate, and talipes equinovagum were noted. After birth, she required oxygen inhalation and phenobarbital due to apnea, and needed tube feeding for a year because of poor feeding. Her development has been significantly delayed; she started rolling over and sitting independently at 1 year 2 months old, crawling at 2 years and 2 months old, and speaking a few single words at 2 year 4 months. She has not started standing with support and walking independently at 2 years and 7 months old. Brain MRI at 4 months old did not reveal any structural brain abnormalities. We identified *de novo* c.706G>A variant, (G236R) by whole exome sequence.

The *KCNK9* variant (G236R) reduces the outward current of the TASK3 channel encoded by *KCNK9* gene, and nonsteroidal anti-inflammatory fenaminic acid such as mefenamic acid (MFA) has been reported to stimulate the channel [Veale et al., 2014, Takahira et al., 2005]. To date, only two affected individuals have been treated with MFA, with noted increased energy while on the medication [Graham et al 2016]. To establish the efficacy of the MFA treatment, the further studies of additional individuals of Birk-Barel syndrome with MFA treatment and the long-term observation of them are expected.

## PgmNr 1145: Impact of few rare genetic disorders on oral health.

**Authors:**

E. Severin; S. Albu; C. Albu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dept Human Genetics, Carol Davila Univ Med Pharm, Bucharest, Romania

---

**Background.** According to the EU definition, accepted by Romanian authorities, a rare disease (RD) is defined as a life-threatening or chronically debilitating condition that afflicts fewer than 5 in 10,000 persons within the general public. It is estimated that 80% of RDs are a result of genetic causes. RDs are characterised by a broad diversity of disorders and symptoms that vary not only from disease to disease but also from patient to patient suffering from the same disease. Craniofacial features, including dental anomalies and diseases, vary considerably from individual to individual as well. Oro-dental findings do not increase the risk of death among people with RDs, but the quality of life becomes worse. **Objectives.** The study based on our experience and a systematic review aimed to study the impact of genetic disorders on oral health in rare diseases and to provide better knowledge of craniofacial problems in some RDs. **Material and Methods.** To collect evidence that genetic disorders play a key role in RDs, an electronic literature search was conducted using different medical databases and additional information resources. The variables studied included craniofacial features, teeth development anomalies, dental occlusion, oral mucosa and oral pathologies. **Results.** The results included few RDs (Cherubism, Cleidocranial dysplasia, Hypohidrotic ectodermal dysplasia and Osteogenesis imperfecta) associated with craniofacial abnormalities and/or oro-dental problems. Genetic causes, signs and symptoms are described. **Conclusions.** Given the high rate of craniofacial, oral and dental abnormalities, people with RDs need a thorough oral evaluation and the pediatric dentist should be familiar with common, rare, and individual findings in people with RDs and anticipate complications which allow better-individualized treatment.

## PgmNr 1146: Whole-exome sequencing analysis of multiplex families with nonsyndromic orofacial clefts identifies variants in genes for Mendelian OFC syndromes.

### Authors:

K.K. Diaz Perez <sup>1</sup>; J.T. Hecht <sup>2</sup>; G.L. Wehby <sup>3</sup>; J.C. Murray <sup>3</sup>; M.L. Marazita <sup>4</sup>; E.J. Leslie <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Emory University, Atlanta, Georgia.; 2) McGovern Medical School at UT Houston, Houston. Texas; 3) University of Iowa, Iowa City, Iowa; 4) University of Pittsburgh, Pittsburgh, Pennsylvania

---

Among the most common congenital facial abnormalities worldwide are orofacial clefts (OFCs), which occur when an infants' lip and/or mouth do not form properly during pregnancy leading to an opening in the lip or roof of the mouth. Most OFCs occur in the absence of other significant structural or developmental abnormalities and are thus considered nonsyndromic, with both genetic and environmental influences on etiology. The last decade of OFC research has focused on common genetic variants, but increasing evidence points to a role for rare variants in a subset of cases. We performed whole-exome sequencing (WES) to identify novel, protein-coding variants in a set of 72 multiplex nonsyndromic OFC families from diverse populations with quasi-Mendelian inheritance patterns. Here we summarize three families with mutations in genes previously associated with OFC syndromes. The first was a Guatemalan pedigree with four male siblings affected with cleft lip and palate (CLP); both parents and a female sibling were unaffected. The most likely causal variant was a frameshift mutation of *TP63* (p. His536Thrfs\*18), shared by the affected individuals, the female sibling, and their mother. The second family was a three-generation Hungarian family where two siblings had cleft palate (CP), their father had CLP, and their paternal grandfather had bifid uvula and syndactyly of the hands and feet. These individuals all shared a missense mutation in *IRF6* (p. Leu22Pro). Neither family meets the diagnostic criteria for the known syndromes caused by *TP63* and *IRF6*. The third family was from Pittsburgh with two siblings and their father, all affected with CP. We identified a 32bp deletion that results in a frameshift mutation in *SMC3* (p. Leu676Argfs\*5) shared among all three affected individuals and not found in two unaffected siblings; DNA from a paternal aunt with CLP was unavailable for study. Mutations in *SMC3* cause Cornelia de Lange syndrome but are primarily missense or in-frame indels, so we hypothesize this variant may cause nonsyndromic OFC through a different mechanism. Overall, our study provides evidence that WES can successfully identify likely pathogenic variants in multiplex OFC families and uncover some of the "missing heritability" of OFCs. More importantly, this work highlights the blurring of the distinction between syndromic and nonsyndromic OFCs and may suggest a role for clinical sequencing in patients with a family history of OFCs.

## PgmNr 1147: Complex rearrangements of *CREBBP* detected in patients with Rubinstein-Taybi syndrome (RSTS).

### Authors:

Y. Enomoto <sup>1</sup>; Y. Tsurusaki <sup>1</sup>; Y. Kuroda <sup>2</sup>; H. Murakami <sup>2</sup>; Y. Kimura <sup>1</sup>; S. Shiiya <sup>2</sup>; T. Naruto <sup>1</sup>; Y. Kido <sup>3</sup>; K. Kurosawa <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Dokkyo Medical University Koshigaya Hospital, Saitama, Japan

---

Rubinstein-Taybi syndrome (RSTS) is characterized by distinct facial features, short stature, broad thumbs and first toes and moderate to severe intellectual disability. Their distinct facial features include downward slanting palpebral fissures, arcuate eyebrow, beaked nose with nasal bridge, high-arched palate and grimacing smile. RSTS is rare autosomal dominant disease arising in rate of about 1/100,000-125,000. Haploinsufficiency of causative genes including *CREBBP* and *EP300* is the underlying mechanism of the disease. Both genes have similar functions as histone acetyltransferase that regulates transcription via chromatin remodeling. 50-70% of patients have mutation in *CREBBP* and 3% have mutation in *EP300*. Microdeletions in *CREBBP* are found in 10% of patients. Here we report complex rearrangements of *CREBBP* in two patients with RSTS. We performed copy number variation (CNV) analysis using panel-based exome data. Patient 1 showed the dispersed deletions in the region encompassing *CREBBP* and *ADCY9*, suggesting small genomic rearrangements. Patient 2 showed mosaic deletion (deletion allele : about 25%, normal allele : about 75%) in exon 7-13 of *CREBBP*. He showed mild phenotype of RSTS. These results indicate that the underlying mechanism of RSTS are highly variable for genomic rearrangement and suggest that the detection system for complex rearrangements are required in the context of clinical setting for RSTS.

# PgmNr 1148: A biallelic variant in the *DACH1* gene (Dachshund Family Transcription Factor 1) causes postaxial polydactyly type A.

## Authors:

M. Umair <sup>1</sup>; Z. Lin <sup>3</sup>; F. Ahmad <sup>2</sup>; M. Bilal <sup>2</sup>; M. Younus <sup>3</sup>; A. Khan <sup>1</sup>; M. Alfadhel <sup>1, 4</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Medical Genomics Research Department, KAIMRC, Riyadh, Riyadh, Saudi Arabia; 2) Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad, Pakistan; 3) State Key Laboratory of Membrane Biology and Beijing Key Laboratory of Cardiometabolic Molecular Medicine, Peking University, Beijing 100871, China.; 4) Division of Genetics, Department of Pediatrics, King Abdulaziz Medical City, Ministry of National Guard-Health Affairs (NGHA), King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia.

---

**Background:** Polydactyly or hexadactyly is a very common limbs deformity mostly characterized by having an extra digit/toe with or without a bone. Currently, mutations in nine genes have been implicated in causing a non-syndromic form of polydactyly including *GLI3*, *GLI1*, *ZNF141*, *IQCE*, *MIPOL1*, *STKLD1*, *FAM92A1*, *KIAA0825*, and *PITX1*.

**Methods:** DNA from single affected individual [IV-1] was subjected to whole exome sequencing (WES), followed by Sanger sequencing to verify segregation of the variants in all the available family members. Protein homology modeling and molecular docking were performed using standard methods.

**Results:** A pathogenic disease-causing biallelic missense variant (**c.569G>A; Cys190Tyr**) was identified in the *DACH1* gene that segregated perfectly with the disease phenotype within the family. The variant was absent in 155 exomes from unrelated Pakistani individuals, different online publically available databases, and in 245 ethnically matched controls. Structural analysis of the DACH1 protein revealed secondary structure change that might result in loss of downstream interaction.

**Conclusion:** Taken together, the present study report on the first direct evidence of the involvement of *DACH1* gene in causing nonsyndromic postaxial polydactyly in humans. This signifies impotence and yet unexplored role of *DACH1* in limbs development.

## PgmNr 1149: Severe Noonan syndrome phenotype associated with a germline Q71R MRAS variant.

### Authors:

H. Suzuki <sup>1</sup>; T. Takenouchi <sup>2</sup>; T. Uehara <sup>1</sup>; S. Takasago <sup>3</sup>; S. Ihara <sup>4</sup>; H. Yoshihashi <sup>5</sup>; K. Kosaki <sup>1</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 2) Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan; 3) Department of Cardiology, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 4) Department of Neurosurgery, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan; 5) Department of Medical Genetics, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan

---

Activation of the RAS pathway through either the activation of genes that accelerate the pathway or the suppression of genes that inhibit the pathway leads to a group of disorders collectively referred to as RASopathies. The key molecules of the RAS pathway are *KRAS*, *HRAS*, and *NRAS*. Mutations in these three RAS homolog genes have been shown to be associated with RASopathies. Recently, two patients with a Noonan syndrome phenotype were shown to carry mutations in the yet another RASopathy gene, *MRAS* (muscle RAS oncogene homolog). Here, we report a patient with a severe Noonan syndrome phenotype associated with a germline Q71R MRAS variant, which represents a recurrent substitution in RAS homologs in various cancers. The patient's dysmorphic features included relative macrocephaly, a down-slanted palpebral fissure, hypertelorism, a depressed nasal bridge, and low-set ears with thick lobes; these facial features are strongly associated with RASopathy. We confirmed that the *MRAS* gene represents a causative gene for RASopathy.

## PgmNr 1150: Primary immunodeficiency with chronic enteropathy and developmental delay in a boy arising from a novel homozygous *RIPK1* variant.

### Authors:

Y. Uchiyama <sup>1,2</sup>; C. Kim <sup>3</sup>; A. Pastorino <sup>4</sup>; J. Ceroni <sup>3</sup>; P. Lima <sup>5</sup>; M. Dorna <sup>4</sup>; R. Honjo <sup>3</sup>; D. Bertola <sup>3</sup>; K. Hamanaka <sup>1</sup>; A. Fujita <sup>1</sup>; S. Mitsuhashi <sup>1</sup>; S. Miyatake <sup>1,6</sup>; A. Takata <sup>1</sup>; N. Miyake <sup>1</sup>; T. Mizuguchi <sup>1</sup>; N. Matsumoto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Yokohama City University, Yokohama, Kanagawa, Japan; 2) Department of Oncology, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 3) Clinical Genetics Unit, Instituto da Crianca, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil; 4) Children's Institute, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil; 5) Patologia Ocular, Divisão de Anatomia Patologica, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil; 6) Clinical Genetics Department, Yokohama City University Hospital, Yokohama, Japan

---

Identification of genetic causes of primary monogenic immunodeficiencies would strengthen the current understanding of their immunopathology. Pathogenic variants in genes in association with tumor necrosis factor  $\alpha$  (TNF $\alpha$ ) signaling including *OTULIN*, *TNFAIP3*, *RBCK1* and *RNF31* cause human congenital autoinflammatory diseases with/without immunodeficiency. *RIPK1*, encoding a receptor interacting serine/threonine kinase 1, is present in protein complexes mediating signal transduction including TNF receptor 1. Biallelic loss-of-function variants in *RIPK1* were recently reported in individuals with primary immunodeficiency with intestinal bowel disease and arthritis. Here we report a novel homozygous *RIPK1* variant in a boy with immunodeficiency and chronic enteropathy. Our patient exhibited severe motor delay and mild intellectual disability, which were previously unknown. The present results are expected to deepen the current understanding of clinical features based on *RIPK1* abnormalities.



## **PgmNr 1151: Unbiased genomic approach to molecular diagnosis in patients with immune system dysfunction leads to unexpected molecular findings, multiple diagnoses and secondary findings.**

### **Authors:**

M. Walkiewicz; M. Similuk; J. Yan; L. Jamal; H. Hullfish; K. Jevtich; C. Lowe; A.P. Hsu; C. Hong; P. Littel; S. Xirasagar; S.M. Holland

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Institute of Allergy and Infectious Diseases, Bethesda, MD.

---

**Background:** In the past decade, dozens of new Mendelian disorders of immunity have been recognized. Yet, the genetic contributions to these disorders remain largely unelucidated. Making progress in this area requires a coordinated, systematic, and transparent approach to clinical genomics and research.

**Methods/Design:** We systematically applied genomic evaluation to research participants, including clinical grade interpretation and reporting of primary and secondary findings. We collect genomic data, family medical history, and standardized phenotype. We interpret and report clinical results in the medical record in conjunction with genetic counseling and shared access to data.

**Results:** We have recruited 1798 participants and completed analysis for 721 individuals and finalized 342 cases. We have issued 249 (72.8%) proband reports; 111 were inconclusive and 138 had at least one molecular finding. Of these, 82 (59.4%) had a pathogenic variant, 16 (11.6%) had a likely pathogenic variant, 39 (28.3%) had a variant of uncertain clinical significance (VUS) and 1 (0.72%) had likely benign variants. Four (2.9%) cases had secondary findings including pathogenic variants in *BRCA1* and *BRCA2*. Two (1.4%) cases had dual molecular diagnoses and 2 (1.4%) cases had a molecular diagnosis outside of “immune” phenotypes. Two (1.4%) cases had revised molecular diagnoses. Lastly, we identified several candidate genes requiring further follow up.

**Conclusion:** Unbiased genomic work-up yields molecular diagnoses, multiple or unexpected diagnoses, and secondary genomic findings, personalizing patient care. Process standardization, data integration, and data sharing policy facilitate research and discovery of candidate genes.

## **PgmNr 1152: Updates to the NIAID Genomic Research Integration System (GRIS): Integrated phenotype-genotype search, laboratory test results integration and visualization, variant tagging, and automating clinical reporting workflows.**

### **Authors:**

S. Xirasagar<sup>1</sup>; A. Oler<sup>1</sup>; Z. Li<sup>1</sup>; X. Cheng<sup>1</sup>; K. Huang<sup>1</sup>; D. Veltri<sup>1</sup>; E. Karlins<sup>1</sup>; J. Barnett<sup>1</sup>; L. Zhang<sup>1</sup>; Y. Fan<sup>1</sup>; K. Kaladi<sup>1</sup>; M. Walkiewicz<sup>2</sup>; M. Similuk<sup>2</sup>; H. Hullfish<sup>2</sup>; J. Yan<sup>2</sup>; G. Ha<sup>3</sup>; U. Shah<sup>3</sup>; A. Long<sup>3</sup>; K. Footer<sup>3</sup>; J. Milner<sup>4</sup>; D. Hurt<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Bioinformatics and Computational Biosciences Branch, Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (NIAID), NIH, Rockville, Maryland.; 2) Division of Intramural Research, National Institute of Allergy and Infectious Diseases (NIAID), NIH, Bethesda, MD; 3) Software Engineering Branch, Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases (NIAID), NIH, Rockville, Maryland.; 4) Division of Intramural Research, Laboratory of Allergic Diseases. National Institute of Allergy and Infectious Diseases (NIAID), NIH, Bethesda, MD

---

The NIAID Genomic Research Integration System (GRIS) has established an interactive web-based application to integrate patient clinical phenotype, pedigree, and genomic sequencing data. GRIS is used by clinical molecular geneticists and novice and expert researchers to identify putative causal variants for disorders and explore relationships between diseases and genes. GRIS is being continuously enhanced to integrate new types of content and broaden the use of existing content. A substantive enhancement, the integrated phenotype-genotype search enables identification of variants that segregate in families comprised of individuals with selected phenotypes. With rapid increases in the volume of next generation sequencing, this feature can be especially helpful for identifying causal and modifier variants for Mendelian immunologic disorders as well as elucidating polygenic disorders, which constitute the majority of immune disease cases. Recent enhancements also include the integration of laboratory test results and data visualizations which provide more granular phenotypic context alongside summary Human Phenotype Ontology (HPO) terms. Lab test results are being integrated at the individual patient level and used to compare patients and build cohorts. Integration of these results with clinical notes provides a rich data source for the creation of phenotypic profiles associated with genetic variants. To enhance user collaboration and productivity centered around genetic variants, GRIS has been extended to include variant tagging by categories including biological significance, clinical variant interpretation, and workflow status tracking. Clinical molecular geneticists use variant tags as a mechanism to communicate their clinical variant analysis results to researchers for review and investigation. Data filters have been expanded in GRIS for more robust filtering of variants including additional public and internal reference databases for population frequencies, gene- and sample-level depth of coverage QC metrics, and impact-based filtering options such as pLI and missense Z scores and extensions to HGMD matching. GRIS enables automated generation of draft clinical reports based on the analyzed and tagged data, which has greatly assisted

the task of clinical variant result reporting by the clinical team. As of June 2019, GRIS includes over 6000 records linked to 1200+ whole exomes. The system is currently being updated to integrate and analyse whole genomes.

## PgmNr 1153: Identification of prion protein as a novel human blood group system antigen.

### Authors:

Y. Omae<sup>1</sup>; S. Ito<sup>2</sup>; M. Takeuchi<sup>3</sup>; K. Isa<sup>4</sup>; K. Ogasawara<sup>4</sup>; K. Kawabata<sup>5</sup>; A. Oda<sup>6</sup>; S. Kaito<sup>6</sup>; H. Tsuneyama<sup>6</sup>; M. Uchikawa<sup>6</sup>; I. Wada<sup>3</sup>; H. Ohto<sup>5</sup>; K. Tokunaga<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Medical Science Project (Toyama), National Center for Global Health and Medicine (NCGM), Shinjuku-ku, Tokyo, Japan; 2) Department of Laboratory Testing, Japanese Red Cross Tohoku Block Blood Center, Miyagi, Japan; 3) Institute of Biomedical Sciences, Department of Cell Science, Fukushima Medical University, Fukushima, Japan; 4) Department of Research and Development, Japanese Red Cross Central Blood Institute, Tokyo, Japan; 5) Department of Blood Transfusion and Transplantation Immunology, Fukushima Medical University Hospital, Fukushima, Japan; 6) Blood Group Section, Japanese Red Cross Kanto-Koshinetsu Block Blood Center, Tokyo, Japan

Blood group systems consist of red cell surface antigens and incompatible blood lead to severe and potentially fatal transfusion reactions or red cell hemolysis. Thirty-six human blood system genes have been registered in International Society of Blood Transfusion (ISBT) and genetic variations including single nucleotide variation encoding amino acid substitution mostly determine presence or absence of the antigens. Anti-KANNO, a broadly reactive red cell alloantibody, is found among some Japanese pregnant women, but the genetic basis of the corresponding antigen remains unclear. Here we integrated a statistical approach to identify the coding gene for KANNO antigen by conducting a genome-wide association study (GWAS) on 4 KANNO<sup>-</sup> individuals and 415 healthy Japanese. The GWAS revealed a genome-wide significant association of chromosome 20p13 locus [ $P = 2.76E-08$ , odds ratio  $> 1,000$  (95% confidence interval = 48–23,674)] and identified single nucleotide polymorphism (SNP) located in an intronic region of prion protein (*PRNP*) gene. We also applied whole exome sequencing to them and revealed a missense variant in the *PRNP* gene (rs1800014, E219K), which is in linkage disequilibrium with the SNP identified in GWAS. Independent 14 KANNO<sup>-</sup> individuals also possessed the homozygous genotype of the missense variant. Monoclonal antibody-specific immobilization of erythrocyte antigens assay using anti-KANNO and mouse anti-human prion protein showed a clear difference between KANNO<sup>+</sup> and KANNO<sup>-</sup> red cells, locating KANNO antigen on red cell specific membrane protein. *In vivo* and *in vitro* binding assays of anti-KANNO were further applied to the cells expressing human prion protein and anti-KANNO showed direct binding to CHO-K1 cells expressing wild type PRNP but not to those expressing E219K PRNP.

In conclusion, we established a novel human blood group system through the identification of encoding gene and causative mutation that affects the seropositivity of blood antigen. The risk allele frequencies of identified causative missense mutation are 4% in Asian, but less than 0.5% in other populations and its homozygous genotype was observed in 17 of 8,236 South Asians and 7 of 4,321 East Asians but unobserved in other populations in ExAC database, indicating that the KANNO-negative blood type is likely more frequent in Asians.

## PgmNr 1154: Genetic analysis of families with a rare platelets function disorder.

### Authors:

M. Ansar<sup>1</sup>; S. Aslam<sup>1</sup>; G. Shaheen<sup>1</sup>; M. Rashid<sup>1</sup>; R. Basharat<sup>1</sup>; D. Shehwar<sup>1</sup>; M.R. Alam<sup>1</sup>; R. Wang<sup>2</sup>; X. Zhang<sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan; 2) State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, McKusick-Zhang Center for Genetic Medicine, School of Basic Medicine Peking Union Medical College, Beijing, China; 3) The Research Center for Medical Genomics, China Medical University, Shenyang, China

---

Glanzmann Thrombasthenia (GT) is a rare autosomal recessive disorder that is mainly caused by defects in the platelet function. Mutations in *ITGB2A* and *ITGB3* genes are known to cause GT. For this study, individuals from 15 families were recruited based on their bleeding time and platelet aggregation profiling. Proband showing the platelet dysfunction were subjected to Sanger and exome sequencing which resulted in the identification of a novel and seven known mutations in *ITGB2A* and *ITGB3* genes. In one of the GT families, a novel variant (c.1512T>A) in *ITGB3* gene was identified in the affected individual and in-silico tools predicted that the variant probably produces a truncated protein (p.Y504\*). However, analysis of the platelets cDNAs from the GT patient and her mother revealed the presence of *ITGB3* transcripts of varying lengths. This finding ruled out the involvement of the previously predicted nonsense mediated decay of transcripts carrying c.1512T>A variant as its sole fate. The cDNA sequencing from this patient and her mother indicated the involvement of alternative splicing due to c.1512T>A variant (via a yet unknown mechanism). Intriguingly, the alternatively spliced transcript present in GT patient and her mother lacked 75 nucleotides around the mutation site in exon 10 of *ITGB3* transcript. Relative expression analysis of these samples demonstrated a reduced level of the alternatively spliced transcript in GT patient as compared to the carrier mother and a healthy control. In line with this data, the immunoblotting confirmed the presence of protein in both the samples with a marked reduction in Itgb3 level in GT patient as compared to the mother. Based on these results it can be concluded that *ITGB3* variant (c.1512T>A) lowered the amount of functional protein in the GT patient which might be a possible reason for the platelet dysfunction and disease phenotype. However additional studies are required to further explore the underlying splicing mechanism and protein function defects.

## PgmNr 1155: A single center experience in Turkey in the molecular diagnosis of hemophilia B.

### Authors:

H. Onay <sup>9</sup>; E. Isik <sup>1</sup>; K. Kavakli <sup>2</sup>; F. Sahin <sup>3</sup>; M. Sezgin Evim <sup>4</sup>; C. Albayrak <sup>5</sup>; G. Tuysuz Kinturp <sup>6</sup>; B. Antmen <sup>7</sup>; E. Yilmaz Keskin <sup>8</sup>; B. Akgun <sup>1</sup>; F. Ozkinay <sup>1,9</sup>; T. Atik <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Pediatric Genetics, Department of Pediatrics, School of Medicine, Ege University, Izmir, Turkey; 2) Division of Pediatric Hematology, Department of Pediatrics, School of Medicine, Ege University, Izmir, Turkey; 3) Division of Hematology, Department of Internal Medicine, School of Medicine, Ege University, Izmir, Turkey; 4) Division of Pediatric Hematology, Department of Pediatrics, School of Medicine, Uludag University, Bursa, Turkey; 5) Division of Pediatric Hematology, Department of Pediatrics, School of Medicine, Ondokuz Mayıs University, Samsun, Turkey; 6) Division of Pediatric Hematology, Department of Pediatrics, School of Medicine, Akdeniz University, Antalya, Turkey; 7) Division of Pediatric Hematology, Department of Pediatrics, Acibadem Hospital, Adana, Turkey; 8) Division of Pediatric Hematology, Department of Pediatrics, School of Medicine, Suleyman Demirel University, Isparta, Turkey; 9) Department of Medical Genetics, School of Medicine, Ege University, Izmir, Turkey

---

Hemophilia B (HB) is an X linked recessive bleeding disorder with a prevalence of 1 in 30,000 live male births. Bleeding symptoms vary according to the level of coagulation factor IX activity. To date, more than 1200 mutations have been defined in the *F9* gene. The aim of this study was to determine the mutation spectrum of the *F9* gene in HB patients from Turkey and to establish a phenotype-genotype correlation.

Thirty-four HB patients, who were molecularly analyzed for *F9* gene mutations in Ege University Pediatric Genetics Laboratory, were included in the study. Clinical and laboratory findings were obtained from hospital records. The Factor IX Gene Variant Database and Human Gene Mutation Database were searched for the identified mutations. Pathogenicity of the variants was classified in accordance with ACMG criteria.

From the 34 HB patients 18 (52.9%) were classified as severe, 13 (38.2%) moderate and 3 (8.8%) mild. Missense, nonsense, splice site, and frameshift mutations were identified in 19 (55.8%), 8 (23.5%), 4 (11.7%) and 1 (2.9%) of the study group, respectively. In one patient, the large deletion was identified. In only one patient, there was no mutation found via sequencing analysis. Five mutations (c. 89-2\_89-1insT, c.521-1G>A, c. 839-1G>T, c. 1088G>T, and c. 1238G>T) were novel. Two of the 5 patients with novel mutations had moderate phenotype while the remaining 3 had the severe phenotype.

In conclusion, in our study group, mutations in the *F9* gene could be found anywhere throughout the entire gene; as in previous studies. This study may contribute to the phenotype-genotype correlation of HB due to the 5 novel mutations mentioned herein.

## PgmNr 1156: A novel molecular indicator for inhibitor development in hemophilia A.

### Authors:

T. Atik <sup>1</sup>; H. Mehdiyeva <sup>1</sup>; B. Akgun <sup>1</sup>; F. Ozkinay <sup>1</sup>; K. Kavakli <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatric Genetics, Ege University, School of Medicine, Izmir, Turkey; 2) Pediatric Hematology, Ege University, School of Medicine, Izmir, Turkey

The development of neutralizing antibodies, which is called inhibitor, against factor VIII (FVIII) is a serious complication of the early stages of replacement therapy in hemophilia A (HA). Risk factors for inhibitor development that allow risk stratification are classified simply into two groups: modifiable and unmodifiable. The position and type of substitution of missense mutations may influence the risk of inhibitor development. In this study, we investigated the inhibitor risk of missense mutations and affecting factors including changing in the type of amino acid group.

We evaluated the *F8* gene variant database (<http://www.factorviii-db.org/>) by EAHAD to collect mutation data and clinical findings including the severity of HA and inhibitor development in the patient. The data in the cases with missense mutations recorded in the database were then selected. Twenty amino acids are known to be classified according to their side chains into 2 major groups: nonpolar hydrophobic and polar. Otherwise, the polar group may subclassify into 3 subgroups: basic, acidic and polar uncharged (Table). Each mutation data were noted to be changed or unchanged according to whether substitution caused changes in the amino acid class using these two classifications.

In this study, otherwise, we used CADD, REVEL, M-CAP, and DANN scores to find significant cut-off value and a specific marker indicating inhibitor development.

Among all data in the *F8* gene variation database (<http://www.factorviii-db.org/>), 3248 different case data with 954 different missense mutations were excluded. The data with both clinical phenotype and inhibitor information were in 2207 cases. Of these 2207 cases, 153 (%6.9) had been reported to be positive. Evaluation according to whether substitution caused changes in the amino acid class or not using 4 groups classification showed that unchanged variants are more associated with inhibitor development in overall ( $p=0.012$ ). This result was more significant when the cases differed between clinical severity. While 9.8% of cases with unchanged variants ( $n=44/451$ ) have been reported to have inhibitor of FVIII, this value was 4.8% ( $n=34/709$ ) in the cases with changed variants in mild HA patients ( $p=0.001$ ). In mild HA cases, CADD and REVEL scores have been found to associated with inhibitor development.

In mild HA cases with a missense mutation, amino acid substitution type (unchanged) might be a novel indicator for inhibitor development in hemophilia A.



## PgmNr 1157: Multiple genetic diagnoses in a cohort of patients with Cryopyrin Associated Periodic Syndrome (CAPS).

### Authors:

A. Almeida de Jesus <sup>1</sup>; S. Torreggiani <sup>1</sup>; M. Garg <sup>2</sup>; S. Alehashemi <sup>1</sup>; K. Johnson <sup>1</sup>; J. Wade <sup>1</sup>; L. Bichell <sup>1</sup>; G.A. Montealegre Sanchez <sup>1</sup>; M. Walkiewicz <sup>3</sup>; R. Goldbach-Mansky <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Translational Autoinflammatory Diseases Section (TADS), NIAID/NIH, Bethesda, MD; 2) Rochester Regional Health, Rochester, NY; 3) Division of Intramural Research (DIR), NIAID/NIH, Bethesda, MD

**Background.** Cryopyrin associated periodic syndrome (CAPS) is an autosomal dominant autoinflammatory disease caused by mutations in *NLRP3*. CAPS comprises 3 clinical phenotypes of increasing severity: familial cold autoinflammatory syndrome (FCAS), Muckle-Wells syndrome (MWS) and neonatal onset multisystem inflammatory disease (NOMID). Patients with CAPS present with variable severity of systemic inflammation, neutrophilic urticaria, hearing loss and aseptic meningitis. We aimed to identify additional Mendelian diagnoses that may explain phenotypic variability not attributable to CAPS.

**Methods.** Whole exome sequencing (WES) and chromosomal microarray analysis (CMA) were performed.

**Results.** Of the 11 CAPS patients analyzed, 4 (36.4%) had a Mendelian genetic diagnosis in addition to CAPS. **Patient 1** is a 10-year-old female with a somatic mutation in *NLRP3* (NM\_001079821)(c.925G>T, p.G309C, ~12% mosaicism) and classic NOMID manifestations. She was noted to have developmental delay, upslanting palpebral fissures and sparse eyebrows, which were explained by CMA findings compatible with 12q21.2q22 deletion syndrome. **Patient 2** is a 10-year-old female with a germline *NLRP3* mutation (c.916G>A, p.E306K) and clinical features of FCAS/MWS. She presented with clinodactyly, scoliosis, and broad great toes. Height was between the 5<sup>th</sup> and 10<sup>th</sup> percentile. WES revealed a novel heterozygous variant in *NPR2* (c.2T>A, p.M1?), gene associated with epiphyseal chondrodysplasia (MIM#615923). **Patient 3** is an 8-year-old female with a germline *NLRP3* mutation (c.1437C>A, p.N479K) and a NOMID phenotype. She presented with brachydactyly, a round face and a broad nasal bridge. WES showed a novel heterozygous variant in *FBN1* (c.7876C>G, p.L2626V), gene associated with several musculoskeletal syndromes, including Marfan syndrome (MIM#154700) and Weill-Marchesani syndrome 2 (MIM#608328). **Patient 4** is a 19-year-old male with a germline mutation in *NLRP3* (c.983G>A, p.G328E) and a clinical picture of NOMID. WES showed a novel heterozygous variant (c.821T>G, p.M274R) in *NLRC4*, gene associated with macrophage activation syndrome (MIM#616050).

**Conclusions.** Of 11 CAPS patients with pathogenic *NLRP3* mutations, 4 had a secondary genetic diagnosis, which explained clinical features that were not attributable to CAPS. Our data suggest that further genetic testing should be considered to explain “phenotypic variability”.

Funding for this study was provided by the Division of Intramural Research, NIAID, NIH



## PgmNr 1158: Novel *ELANE* gene mutation in a Vietnamese boy with severe congenital neutropenia.

### Authors:

C. Bui <sup>1</sup>; T. Pham <sup>2</sup>; L. Pham <sup>2</sup>; D. Le <sup>2</sup>; A. Nguyen <sup>3</sup>; T. Nguyen <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Functional genomics, Centre for molecular biomedicine, University of medicine, Hochiminh, Cantho, VietNam; 2) DNA Medical Technology, UMP Hospital, Hochiminh city, Vietnam; 3) PIDs unit, Children Hospital 1, Hochiminh city, Vietnam

---

Severe chronic neutropenia is a congenital condition defined as an absolute neutrophil count of less than 500/ $\mu$ L for at least three months. There are three categories according to the onset time or pattern of variation of neutrophil levels: congenital, cyclic, and idiopathic. Among them, severe congenital neutropenia is an inborn disorder with maturation arrest of the early stage of granulopoiesis associated with various genetic abnormalities that has been classified with 26 genes. We here report a 4 year-old boy was admitted due to fever, sepsis, *skin abscesses and gingivitis*. He had a past history of febrile illnesses including septic shock, pulmonary *tuberculosis*, recurrent pneumonia, cutaneous abscesses, oral thrush and isolated neutropenia from birth (  $0.1 - 0.4 \times 10^9/L$  ). Family history was nonspecific and he had no siblings. Initial laboratory revealed severe neutropenia (white blood cells, neutrophils, lymphocytes, monocytes was  $4.73 \times 10^9/L$ ,  $0.17 \times 10^9/L$ ,  $3.68 \times 10^9/L$ ,  $0.68 \times 10^9/L$ , respectively) and increasing of erythrocyte sedimentation rate (112 mm/hr) and C-reactive protein (121 mg/L). A gram-stain and culture of blood showed *Ralstonia pickettii* and culture of abscess fluid showed negative. Bone marrow findings showed normal myeloid cells, reduced granulocyte cell line and no malignant cells. Test of HIV and TB profile were negative. Immunoglobulin quantification results showed normal range of IgG, IgA and IgM. Lymphocyte immuno-phenotyping showed normal count of T lymphocytes, TCD4, TCD8, TCD4, B lymphocytes and natural killer cells. No duplication or deletion in the CGH. Of note, novel mutation of the *ELANE* gene encoding human neutrophil elastase is found via whole exome sequencing and validated by Sanger sequencing. Functional motif Patient were managed with prompt administration of *combination intravenous antibiotic therapy* in two week without granulocyte colony stimulating factor. When he discharged from hospital, white blood cells and neutrophils was  $9.94 \times 10^9/L$  and  $0.43 \times 10^9/L$ , respectively. He received antimicrobial prophylaxis with trimethoprim-sulfamethoxazole and did not accept high cost of *G-CSF* prophylaxis in 6- month follow- up. During the follow-up observation period, he suffered from two times of skin abscesses.

## **PgmNr 1159: Integrative exome analysis of primary immunodeficiency diseases reveals novel pathogenic variants correlated with specific clinical findings among Vietnamese patients.**

### **Authors:**

T. Pham <sup>1</sup>; T. Vu <sup>1</sup>; L. Pham <sup>1</sup>; D. Le <sup>1</sup>; A. Nguyen <sup>3</sup>; C. Bui <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Precision Genomics Unit, DNA Medical Technology, Ho Chi Minh city, Ho Chi Minh city, Vietnam; 2) Functional genomics, Centre for Molecular Biomedicine, University of Medicine, Ho Chi Minh city, Vietnam; 3) Haematology & Immunology Unit, Children Hospital 1, Ho Chi Minh city, Vietnam

---

The diagnostic landscape of genetically heterogeneous primary immunodeficiency diseases (PID) has been transformed remarkably by the advent of sequencing. However, no genetic study of PID applying whole exome sequencing has been conducted in Vietnam. This study aims to build effective approach for identifying pathogenic variants in PID exomes and characterizing distinct mutations among Vietnamese patients, which contributes to the genomic database of Asian population. 36 patients with suspected PID and 65 controls were recruited. Mapping, variant calling and annotation was conducted applying BWA, GATK Best Practices and ANNOVAR. Extensive panel of 410 PID genes was developed following Clinical Genomic Database 2019 and IUIS 2017. We applied an in-house bioinformatics pipeline for WES analysis which integrates collaboration between geneticists and physicians. Criteria for variant prioritization were: rare variants with MAF<0.01, predicted deleterious in numerous functional prediction algorithms from dbNSFPv3.5a database. Downstream pathogenicity assessment of top candidates follows 3 conditions: (1) matched correlation of disease causing genes with clinical presentation including key symptoms, (2) comprehensive database search to determine the impact of detected mutations on expression and function of DNA, RNA and encoded protein, (3) internal control database was screened to ensure the absence of concerned variants. The family segregation of all detected variants was validated by Sanger sequencing. A diagnostic yield of 30% was achieved. We identified up to 67% novel pathogenic variants in PID patients: 3 cases with de-novo dominant mutations, 4 cases with compound heterozygous mutations and 4 cases with hemizygous mutations, leaving 25 cases undiagnosed. Causative variants in 10 genes were highly associated with clinical findings of 4 main PID groups: immunodeficiencies affecting cellular and humoral immunity (*ADA*, *CD40LG*, *IL2RG*, *RAG1*), diseases of immune dysregulation (*AIRE*, *LRBA*, *SH2D1A*), congenital defects of phagocyte number or function (*ELANE*, *ITGB2*), and predominantly antibody deficiencies (*NFKB2*). Further functional validation is required to prove the pathogenicity of detected mutations, and hence to share these data in genetic databases. Our results suggest that WES holds great promise as an effective diagnostic strategy for PID and other rare diseases in developing countries like Vietnam.

## PgmNr 1160: *UNC13D* deficiency associated with epileptic seizures and antibody deficiency: The first case from the Iranian national registry.

### Authors:

P.A. Amirifar <sup>1,2</sup>; R.E. Yazdani <sup>2</sup>; M.R. Ranjouri <sup>3</sup>; H.A. Abolhassani <sup>2,4</sup>; G.O. Azizi <sup>5</sup>; N. Parvaneh <sup>2</sup>; N. Rezaei <sup>2</sup>; A.S. Aghamohammadi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, Tehran university of medical science, Tehran, Tehran, Iran; 2) Research Center for Immunodeficiencies, Pediatrics Center of Excellence, Children's Medical Center, Tehran, and the University of Medical Science, Tehran, Iran; 3) Genetics, Zanzan university of medical sciences, Zanzan, Iran; 4) Division of Clinical Immunology, Department of Laboratory Medicine, Karolinska Institute at Karolinska University Hospital Huddinge, Stockholm, Sweden; 5) Non-Communicable Diseases Research Center, Alborz University of Medical Sciences, Karaj, Iran; Department of Laboratory Medicine, Imam Hassan Mojtaba Hospital, Alborz University of Medical Sciences, Karaj, Iran

---

Familial hemophagocytic lymphohistiocytosis (FHL) is a severe life-threatening immunodeficiency that is characterized by proliferation and infiltration of hyperactivated macrophages and T-cells. FHL is a heterogeneous autosomal recessive disorder with five subtypes. Certain mutations in some genes encoding perforin (*PRF1*: FHL2) [11], munc13-4 (*UNC13D*: FHL3), syntaxin-11 (*STX11*: FHL4), and munc18-2 (also known as syntaxin-binding protein 2) (*STXBP2*: FHL5) could result in FHL disease. Perforin is a pore-forming protein that is crucial for target cell lysis, while *UNC13D*, *STX11*, and *STXBP2* encode proteins that are in charge of fusion and trafficking of the granule exocytosis towards the target cells [1, 15, 16]. The clinical symptoms result from decreased or absent cytotoxic T lymphocytes and natural killer cells activities.

Here, we described a fatal case of Iranian infant with prolonged fever, hepatosplenomegaly, cytopenias, hemophagocytosis, skin rash and neurological symptoms. The diagnosis of FHL type3 was confirmed by the identification of homozygous c.1208T>C (p.Leu403Pro) missense mutation in exon 14 of *UNC13D* gene. In contrast to previously reported *UNC13D* deficient patients, immunologic findings indicated a significant decrease in CD19<sup>+</sup> and CD20<sup>+</sup> cells in our patient. It is noteworthy that major humoral immunodeficiency has been reported in FHL so far. Based on our data, *UNC13D* deficiency should be considered in children with phenotype resembling antibody deficiency and lymphoproliferative disorders.

## PgmNr 1161: Biallelic loss-of-function *JAK1* variants detected by whole exome sequencing in a patient with immunodeficiency and multiple congenital anomalies.

### Authors:

J. Ji<sup>1,2</sup>; R.J. Schmidt<sup>1,2</sup>; G. Raca<sup>1,2</sup>; J.A. Biegel<sup>1,2</sup>; J.A. Church<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA; 2) Department of Pathology, Keck School of Medicine of the University of Southern California, Los Angeles, CA; 3) Department of Pediatrics, Children's Hospital Los Angeles, Los Angeles, CA; 4) Department of Pediatrics, Keck School of Medicine of the University of Southern California, Los Angeles, CA

The *JAK1* gene encodes Janus kinase 1, a ubiquitously expressed protein-tyrosine kinase. Previous animal model experiments suggest that *JAK1* plays an essential role in interferon-alpha/beta and interferon-gamma signal transduction. *Jak1* <sup>-/-</sup> cells do not manifest biologic responses to a subset of cytokine receptors. To date, a single patient with recurrent mycobacterial infection and early-onset metastatic bladder carcinoma has been reported to harbor biallelic pathogenic variants in *JAK1* leading to *JAK1* deficiency. However, the *JAK1* gene-disease association has not yet been well-established.

Here, we report a male patient with newborn screen TREC = 1/mcL (reference range >18); CD3+ = <20/mL, CD19+ = 538/mcL, NK = <20, consistent with T-B+NK- severe combined immunodeficiency disease. Other clinical features included hypotonia, pulmonary hypertension, respiratory distress, pneumothorax, patent foramen ovale, ventriculomegaly, tricuspid insufficiency, dysmorphic facial features, wide-spaced nipples, arthrogryposis, club foot, sclerosis of the bones, bowing of the femurs, mesomelia, rhizomelia, single palmar crease, brachydactyly, elevated prothrombin time, elevated partial thromboplastin time, and abnormal brain imaging.

Exome sequencing for the patient detected an apparent homozygous c.1613dupT (p.Met539Hisfs\*21) variant in exon 11 of the *JAK1* gene (NM\_002227.3). Due to the lower depth coverage for this *JAK1* frameshift variant, copy number variant detection was performed using NxClinical. A 58 kb overlapping deletion encompassing exons 9-25 of the *JAK1* gene was identified. This deletion was confirmed by retrospective review of the chromosomal microarray data for this patient. Follow-up parental studies demonstrated that the frameshift variant was inherited from the father and the multi-exon deletion was inherited from the mother, consistent with compound heterozygosity. There were no other clinically significant variants identified from exome sequencing and chromosomal microarray analysis.

This is the second reported patient with inherited bi-allelic *JAK1* loss-of-function variants and human immunodeficiency. Although the bi-allelic *JAK1* mutations are also likely related to this patient's other clinical findings, including congenital cardiac, pulmonary, skeletal and neurological anomalies, the presence of additional pathogenic alterations not detected by whole exome sequencing cannot be ruled out.

## **PgmNr 1162: Pyrin-associated Autoinflammation with Neutrophilic Dermatosi (PAAND): Novel autosomal recessive MEFV mutation and genotype-guided colchicine treatment.**

### **Authors:**

L. Youssefian <sup>1</sup>; M. Faghankhani <sup>1</sup>; H. Vahidnezhad <sup>1</sup>; V. Ziaee <sup>2</sup>; S. Zeinali <sup>3</sup>; J. Uitto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Thomas Jefferson University, Philadelphia, PA, USA; 2) Tehran University of Medical Sciences, Tehran, Iran; 3) Pasteur Institute of Iran, Tehran, Iran

---

PAAND is a recently described monogenic auto-inflammatory disorder which shares clinical features with Familial Mediterranean Fever (FMF). PAAND unlike FMF, is not associated with serositis or amyloidosis. Although, different recessive and dominant mutations in *MEFV* can cause FMF, up to now, all reported cases of PAAND were associated with a dominant p.S242R mutation in *MEFV*. The proband was a 5-years old female born to healthy, consanguineous parents. She developed an oral aphthous ulcer at 20 days of age, followed by a maculo-papular and pustular skin rash on the perioral region and upper and lower extremities. Other symptoms included perineal mucosal inflammatory lesions, recurrent episodic fever and abdominal pain. She had failure to thrive and developmental delay, and manifested with episodic myalgias and arthralgias. Her younger brother presented with similar symptoms. High-dose corticosteroid therapy failed to cause symptomatic relief, but both the patient and her similarly affected brother showed dramatic improvement with Colchicine treatment, with complete resolution of symptoms within several months. We sought the genetic cause of this auto-inflammatory disease with whole exome sequencing (WES) of the trio. We detected in the patient a novel homozygous missense mutation, in *MEFV*, which was heterozygous in the parents. Genome-wide homozygosity mapping revealed a number of regions of homozygosity (ROH, <2Mb), in the patient's DNA, including one on chromosome 16 harboring *MEFV*. PAAND is an auto-inflammatory disease with *MEFV* mutation, our family being the first one with documented autosomal recessive inheritance. Colchicine, an effective treatment for FMF, showed promising results in the treatment of PAAND.

## PgmNr 1163: Duplication that involve the *F9* gene such as cause of thrombophilia X-Linked due to factor IX.

### Authors:

M. Manotas <sup>1,2</sup>; R. Quero <sup>1,2</sup>; J. Villalta <sup>1,2</sup>; G. Mariana <sup>1,2</sup>; J. Rojas <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Instituto Genética Humana, Facultad de Medicina, Pontificia Universidad Javeriana, Bogotá, Colombia; 2) Hospital Universitario San Ignacio, Bogotá, Colombia

---

Thrombophilia is defined as a predisposition to thrombosis. It is caused by inherited or acquired causes. Establish genetic causes include antithrombin III, protein C, protein S, factor V Leiden, prothrombin 20210A allele, and MTHFR mutations. Other genetic factors causing thrombosis or pulmonary embolism have been identified in recent studies, including elevated factor VIII and plasminogen activator inhibitor-1 (PAI-1) deficiency. However elevated factor IX is actually considered such an indeterminate cause of genetic thrombophilia.

We present the case of a family suffering from recurrent thrombosis and pulmonary embolism due to hereditary thrombophilia. The proband is a 35-year-old male who experienced ten episodes of pulmonary embolism. Thrombophilia was suspected and screening tests for establish genetics factors were negative. Then exome in trio further sequencing analysis identified a duplication that involved the *F9* gene in the patient such as his mother, this duplication was confirmed with comparative genomic hybridization. The mother suffered from one episodes of deep venous thrombosis, pulmonary embolism and ischemic stroke and one sister experienced one episode of pulmonary embolism at 35 years-old.

The present report highlights a very rare case of thrombophilia related with duplication that involve the *F9* gene. Factor IX is a circulating serine protease that serves as an essential component of the blood coagulation pathway. The gain of function is the possible pathogenic mechanism that associates factor IX as a cause of thrombophilia. Our findings suggest the duplication of *F9* gene such as cause of thrombophilia, X-linked, due to factor IX defect.

## PgmNr 1164: *De novo* loss-of-function variants in *CAPRIN1* in patients affected by severe autism and language delay.

### Authors:

A. Brusco<sup>1</sup>; L. Pavinato<sup>1</sup>; E. Giorgio<sup>1</sup>; D. Carli<sup>2</sup>; V. Salpietro<sup>3</sup>; I.M.B.H. van de Laar<sup>4</sup>; E. Di Gregorio<sup>5</sup>; J.L. Howe<sup>6</sup>; E. Agolini<sup>7</sup>; A. Novelli<sup>7</sup>; S.W. Scherer<sup>6,8,9,10</sup>; G.B. Ferrero<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Sciences, University of Turin, Turin, Italy; 2) Department of Public Health and Pediatrics, University of Turin, Turin, Italy; 3) Department of Molecular Neuroscience, UCL Institute of Neurology, London, United Kingdom; 4) Department of Clinical Genetics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands; 5) Medical Genetics Unit, Città della Salute e della Scienza University Hospital, Turin, Italy.; 6) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada; 7) Laboratory of Medical Genetics, Bambino Gesù Children's Hospital, Rome; 8) Genetics & Genome Biology Program, The Hospital for Sick Children, Toronto, Canada; 9) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 10) McLaughlin Centre, University of Toronto, Toronto, Canada.

---

We report the description of six patients carrying heterozygous *de novo* nonsense or splicing mutations in *CAPRIN1*, identified by whole-genome or whole-exome sequencing, and a case with a 1 Mb *de novo* deletion encompassing the gene.

All cases are characterized by severe non-syndromic autism spectrum disorder and severe language delay without any skeletal anomalies or characteristic facial dysmorphisms.

Caprin-1 is an ubiquitous protein highly expressed in brain and involved in essential cellular processes, such as cell proliferation and migration. The protein localizes at neuronal RNA granules, that are fundamental structures for the proper transport and translation of many mRNAs of key proteins involved in synaptic plasticity. Caprin-1 is able to directly interact with the Fragile X Mental Retardation protein (FMRP) and with the Ras GTPase-activating protein-binding protein 1 (G3BP1). Interestingly, *CAPRIN1*<sup>-/-</sup> mice die perinatally, while *CAPRIN1*<sup>+/-</sup> are characterized by a reduction in social interactions and lower response to novelty (Ohashi *et al.* 2016). In rats and mice, Caprin-1 was shown to be important in the development of dendrites and dendritic spines. Behavioural studies on these models have also shown that the protein is involved in long-term memory formation, through a mechanism mediated by AMPA Receptors.

Given the interesting available information on the gene, we aim at establishing an international collaborative network aims at demonstrating the role of Caprin-1 haploinsufficiency as a mechanism involved in the onset of a specific form of neurodevelopmental disorder.



## PgmNr 1165: Family-based whole exome sequencing allows a 20% diagnostic yield in patients with isolated autism spectrum disorder.

### Authors:

D. Carli <sup>1</sup>; E. Giorgio <sup>2</sup>; E. Riberi <sup>1</sup>; L. Pavinato <sup>2</sup>; S. Cardaropoli <sup>1</sup>; A. Mussa <sup>1</sup>; A. Pelle <sup>3</sup>; G. Mandrile <sup>3</sup>; V. Antona <sup>4</sup>; L. Sorasio <sup>5</sup>; E. Biamino <sup>1</sup>; E.F. Belligni <sup>1</sup>; F. Sirchia <sup>6</sup>; I. Bagnasco <sup>7</sup>; C. Arduino <sup>8</sup>; G. Zacchetti <sup>8</sup>; G. Gai <sup>8</sup>; E. Grosso <sup>8</sup>; C. Davico <sup>9</sup>; R. Keller <sup>10</sup>; S. De Rubeis <sup>11</sup>; A. Brusco <sup>2,8</sup>; G.B. Ferrero <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Public Health and Pediatrics, University of Turin, Turin, Italy; 2) Department of Medical Sciences, University of Turin, Turin, Italy; 3) Medical Genetics, San Luigi Gonzaga University Hospital, Orbassano, Italy; 4) Department of Sciences for Health Promotion and Mother and Child Care "G. D'Alessandro", University of Palermo, Palermo, Italy; 5) Pediatrics, Santa Croce e Carle Hospital, Cuneo, Italy; 6) Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy; 7) Neuropsichiatria infantile Torino Sud- ospedale Martini- Asl Città di Torino, Turin, Italy; 8) Medical Genetics Unit, Città della Salute e della Scienza University Hospital, Turin, Italy; 9) Division of Child and Adolescent Neuropsychiatry, Department of Public Health and Pediatric Science, University of Turin, Turin, Italy; 10) Local Health Unit ASL City of Turin, Department of Mental Health, Turin, Italy; 11) Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York

**Introduction:** Neurodevelopmental disorders (ND) are highly heterogeneous conditions, including autism spectrum disorders (ASD) and intellectual disability (ID), characterized by extensive clinical and molecular overlap. Previous studies by whole exome sequencing (WES) in patients with ND indicated a diagnostic yield ranging widely (6-45%), with higher rates in patients with additional clinical features compared to patients with isolated ND.

**Methods:** Trio/quad based WES was performed at ASC consortium on 78 families with one or more probands with ND. Forty-two patients presented ID (10 isolated and 32 syndromic) and 36 ASD (23 isolated and 13 syndromic). ID and ASD were considered syndromic when associated with at least one of the following features: facial dysmorphisms, microcephaly/macrocephaly, growth delay/overgrowth, epilepsy, congenital malformations/skeletal abnormalities.

**Results:** We identified 22 pathogenic or likely pathogenic variants in genes associated with monogenic Mendelian disorders (28.2%): 10 in patients with ASD (28%) and 12 in patients with ID (29%). Five patients presented syndromic ASD with mutations in *ARID1B*, *DSCAM*, *HNRNP2*, *NAA15*, *SRCAP* genes respectively; 5 isolated ASD with mutations in *DEAF1*, *DLG4*, *FOXP1*, *PTEN* genes, 11 syndromic ID with mutations in *BRAT1*, *CDK5RAP2*, *CEP152*, *DDX3X*, *GRIN1*, *H3F3B*, *PIGC*, *PPM1D*, *PRR12*, *PTPN11*, *SON* genes and one isolated ID with mutation in *IQSEC2*. Detection rates in the four subgroups were 38.5%, 21.7%, 34.4% and 8.3%, respectively. All 10 ASD patients presented de novo variants, while patients with ID presented 7 de novo variants and 4 biallelic variants consistent with an autosomal recessive inheritance.

**Conclusion:** These data confirm that trio based WES has a high diagnostic rate in ND. Notably, more than 20% of patients with isolated ASD disclosed a monogenic cause, indicating that trio based WES is a powerful strategy for the molecular diagnosis of isolated ASD. Moreover, these results demonstrate



that several genetic syndromes associated with ND are characterized by a wide phenotype spectrum ending with isolated ASD in the absence of other obvious clinical signs.

## PgmNr 1166: Exome sequencing of a Finnish cohort reveals known and candidate genes for syndromic neurodevelopmental disorders.

### Authors:

T. Hadley <sup>1</sup>; Z.H. Coban Akdemir <sup>1</sup>; S. Jhangiani <sup>2</sup>; D. Muzny <sup>2</sup>; T. Määttä <sup>3</sup>; I. Järvelä <sup>4</sup>; J. Lupski <sup>1,2,5,6</sup>; J. Posey <sup>1</sup>; Baylor Hopkins Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, 77030, USA; 3) Disability Services, Joint Authority for Kainuu, Kajaani, Finland; 4) Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, 00251, Finland; 5) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA; 6) Texas Children's Hospital, Houston, Texas, 77030, USA

---

**Background:** Developmental delay/intellectual disability (DD/ID) is a heterogeneous group of disorders marked by limits in cognitive function, adaptive behaviors, and/or physical functioning. Identification of the genetic contributions to DD/ID is complicated by its extreme genetic heterogeneity. Recessive pathogenic variants contribute to DD/ID, particularly in isolated populations (i.e. bottleneck) due to founder alleles and in populations that have high coefficients of consanguinity and identity-by-descent (IBD) within a clan; both result in absence of heterozygosity (AOH) observable in exome sequencing (ES) data from personal genomes.

**Method:** Samples were collected from 5 Finnish families from a subisolate population in Kainuu with syndromic DD/ID. ES and family-based rare variant analysis included two parent-child trios, one parent-child-sibling quad, one parent-child duo, and one sibship with an affected first cousin. Parental consanguinity (first-cousins) was reported in 1 family.

### Results:

At least 1 known or candidate gene was identified in each of 7 affected individuals from 5 families, and included rare pathogenic or likely damaging variants in established DD/ID genes *ANKRD11* (*de novo*) and *GRIN2A*, consistent with autosomal dominant (AD) inheritance of an apparently sporadic trait, and *KIA2022* consistent with an X-linked (XL) trait in a male proband. Variants in previously reported candidate genes included *SPTBN4* (*de novo*) and *DHX58* (*de novo*). Novel candidates include homozygous variants in *ERGIC3* located in 7.1 Mb of AOH and a heterozygous variant in *NPAT*. Two families were identified to have rare variants in 2 genes: supporting a potential dual diagnosis (*GRIN2A*, *NPAT*) in one proband, and providing an explanation for apparent intrafamilial phenotypic variation between another proband (*ERGIC3*) and her cousin (*KIAA2022*). Despite a high proportion of variants (5/7) suggestive of AD disease, the average total AOH is 172.7 Mb (154.5 – 188.2 Mb), with 29.4 (25 – 36) long AOH tracts (>1.606 Mb) per proband, supporting a high rate of IBD in this small cohort.

### Discussion:

Among the 5 families, most variants were *de novo* (3) or consistent with AD inheritance (2). This observation is notable, given the high average AOH observed in the studied probands (172.7 Mb). Our findings provide additional evidence to support 2 previously reported candidate disease genes (*SPTBN4* and *DHX58*) and reveal *ERGIC3* and *NPAT* as potential novel candidate disease genes.

## PgmNr 1167: Haploinsufficiency of *SPTBN1* is associated with syndromic intellectual disability and autism.

### Authors:

J. Posey<sup>1</sup>; R. Xiao<sup>1,2</sup>; F. Kanani<sup>3</sup>; M.J. Parker<sup>4</sup>; M.K. Koenig<sup>5</sup>; A. van Haeringen<sup>6</sup>; F. Xia<sup>1,2</sup>; J.A. Rosenfeld<sup>1</sup>; M.R. Bekheirnia<sup>1,7</sup>; C.M. Eng<sup>1,2</sup>; B.H. Lee<sup>1</sup>; Y. Yang<sup>1,2</sup>; Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human & Molec Genetics, Baylor College of Medicine, Houston, Texas.; 2) Baylor Genetics Laboratories, Houston, Texas, 77030, USA; 3) Sheffield Clinical Genetics Service, Sheffield Children's NHS Foundation Trust, Sheffield, UK; 4) The Wellcome Centre for Ethics and Humanities/Ethox Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK; 5) Department of Pediatrics, University of Texas Health Science Center, Houston, Texas, 77030, USA; 6) Department of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 7) 7.Renal Section, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030 USA

Spectrins are a common component of cell cytoskeletons and function to bind to cytoskeletal elements such as microtubules, and active and intermediate filaments. They are each composed of  $\alpha$  and  $\beta$  subunits, encoded by *SPTA1* and *SPTAN1* ( $\alpha$ ) and *SPTB*, *SPTBN1*, *SPTBN2*, *SPTBN4* and *SPTBN5* ( $\beta$ ). One  $\alpha$  and one  $\beta$  subunit each interact in an antiparallel manner to form rod-shape heterodimers that further organize into tetramers. Spectrins are broadly expressed, notably in the brain, kidney, and muscle. To date, spectrin genes have been associated with erythroid cell disorders (*SPTA1*, *SPTB*) and neurologic disorders (*SPTAN1*, *SPTBN2*, *SPTBN4*), but no disease traits have been identified in association with *SPTBN1* or *SPTBN5*.

Clinical exome sequencing (ES) performed on the index case, a 25-year-old female with autism spectrum disorder, revealed a *de novo* missense variant (c.613G>A [p.G205S]) in the calponin homology domain of *SPTBN1*. Review of the local diagnostic laboratory database and DECIPHER identified 4 additional unrelated individuals with apparently sporadic overlapping neurodevelopmental phenotypes and *de novo* *SPTBN1* variants, including 1 missense variant (c.5641G>A [p.D1881N]) located in the spectrin repeat domain, two canonical splice site variants (c.567-2\_584delins17; c.647+1G>T), and one stopgain variant (exon 5/36). The stopgain variant and both splice site variants are predicted to result in a loss of function. All variants are novel and located at positions that are evolutionarily conserved.

Here, we describe 5 unrelated individuals with *de novo* variants in *SPTBN1* in association with a neurodevelopmental phenotype, supporting the candidacy of *SPTBN1* as a novel neurodevelopmental disorder gene. The observation of a stopgain variant and 2 variants impacting the canonical splice site suggest a loss-of-function mechanism for disease. Genotype-phenotype correlations are limited by the small cohort size, and larger allelic series will further inform the phenotypic spectrum of disease associated with *SPTBN1*.

## PgmNr 1168: Heterozygous loss-of-function variants in *SFMBT1* are associated with intellectual disability.

### Authors:

V. Rossi <sup>1</sup>; B. Weimin <sup>1,2</sup>; M. Nizon <sup>3</sup>; B. Isidor <sup>3</sup>; R. Colombo <sup>4</sup>; A. Lewis <sup>1,5</sup>; P. Ward <sup>1,2</sup>; C. Bénéteau <sup>3</sup>; J. Rosenfeld <sup>1</sup>; B. Lee <sup>1,5</sup>; R. Marom <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX, USA; 2) Baylor Genetics, Houston TX, USA; 3) CHU Nantes, Service de Génétique Médicale, Nantes, France; 4) Faculty of Medicine and Surgery, Catholic University, IRCCS Policlinico Agostino Gemelli, Rome, Italy; 5) Texas Children's Hospital, Houston TX, USA

---

Chromatin remodeling proteins that regulate gene transcription have a well-established role in the pathogenesis of intellectual disability syndromes. These include BAF-related genes such as *ARID1B*, histone demethylases such as *KDM5C*, and components of the mediator complex such as *MED12*, among others. We identified heterozygous, loss-of-function single nucleotide variants or copy number losses encompassing *SFMBT1*, encoding a histone-binding protein, in four unrelated families with intellectual disability of variable severity.

Two subjects had a *de novo* variant leading to a frameshift (c.1445delG, p.Gly482fs) or nonsense (c.1518C>G, p.Tyr506\*) mutation. Two subjects had copy number losses encompassing *SFMBT1*, one of which was found to be *de novo* and the other was a familial variant, inherited from a mildly affected parent and grandparent. Review of the DECIPHER database (<https://decipher.sanger.ac.uk/>) revealed an additional case of chromosome 3p21.1 microdeletion with similar features, further refining the smallest region of overlap to 415 kb in size (chr3: 52811541-53226664). This region includes 7 genes, of which *SFMBT1* is the most compelling candidate to explain the phenotype. Clinically, our subjects share phenotypic features in addition to developmental delay and intellectual disability including mild dysmorphic facial features, short stature in 2/6 subjects, and scoliosis in 3/6 subjects. Because of developmental and speech delay, brain MRI (performed in 3/6 subjects) and audiology exam (performed in 4/6 subjects) were pursued but resulted normal.

*SFMBT1* belongs to a group of highly conserved MBT domain-containing proteins and functions as part of transcriptional corepressor complexes. It is predicted to be intolerant to loss-of function (pLI score 0.94, <https://gnomad.broadinstitute.org>). Interestingly, pathogenic variants in *KDM1A*, encoding a protein known to interact with *SFMBT1*, are the cause of syndromic intellectual disability, further supporting the important role of this complex in neurodevelopment. We suggest that loss-of-function variants in *SFMBT1* should be considered in the diagnostic work up of patients presenting with developmental delay and intellectual disability.

## PgmNr 1169: A novel mutation of *JMJD1C* is involved in epileptic encephalopathy of childhood.

### Authors:

H. Murakami <sup>1</sup>; Y. Enomoto <sup>2</sup>; N. Nishimura <sup>1</sup>; T. Kumaki <sup>1</sup>; Y. Kuroda <sup>1</sup>; K. Kurosawa <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, Kanagawa Children's Medical Center, Minami-ku, Yokohama, Japan; 2) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan

---

Epileptic encephalopathy is characterized by frequent epileptic seizures associated with developmental delay or regression in infancy or childhood. We present here the first patient with epileptic encephalopathy associated with a novel missense mutation in *JMJD1C* (jumonji domain-containing protein 1C) [MIM 604503] (NM\_032776:exon12:c5347G>A;p.A1783T). The patient was 5-year-old female. She was born at 37 weeks of gestation as a first child to non-consanguineous parents by cesarean section after an uncomplicated pregnancy. Her parents were both healthy with no history of seizures. Her birth parameters were normal. Her developmental milestones were not delayed at 1 year and 6 months: controlling her head at 3 months, sitting at 6 months, walking alone at 12 months. She could speak meaningful words at 12 months. At 1 year and 6 months, she developed a status epilepticus with generalized tonic-clonic seizures and transferred to our hospital. Although her clinical examination indicated to a kind of metabolic encephalopathies, there were no clues of diagnosis after the analysis of amino acids in her blood and cerebrospinal fluid and organic acids in her urine. After this event, her development was regressed. Her seizures remained refractory to treatment with multiple anti-epileptic drugs. She was referred to us for the genetical assessment at the age of 5 years. We performed exome sequence and identified a novel heterozygous mutation in *JMJD1C* that was absent in the gnomAD database. Sanger sequencing confirmed that this mutation was *de novo*. This mutation affects evolutionary highly conserved amino acid and is predicted damaging to protein function. *JMJD1C* encodes a histone demethylase shown to be widely expressed in the brain. Only less than ten pathogenic mutations in *JMJD1C* were identified, which were implicated to cause autism spectrum disorder, intellectual disability or Rett like syndrome. This is the first case of epileptic encephalopathy caused by *JMJD1C* mutation and will expand the phenotypic spectrum of *JMJD1C*.

## PgmNr 1170: Biallelic variants in *IQSEC1* cause intellectual disability, developmental delay, and short stature.

### Authors:

S.A. Paracha<sup>1</sup>; M. Ansar<sup>2</sup>; H.L. Chung<sup>3,4,5</sup>; A. Al-Otaibi<sup>6</sup>; T.A. Ravenscroft<sup>3,4</sup>; M.N. Elagabani<sup>7</sup>; R. Scholz<sup>8</sup>; T.A. Magid<sup>6</sup>; M.T. Sarwar<sup>1</sup>; S.F. Shah<sup>9</sup>; A.A. Qaisar<sup>10</sup>; P. Makrythanasis<sup>2,11</sup>; E. Falconnet<sup>2</sup>; E. Ranza<sup>2,12,13</sup>; F.A. Santoni<sup>2,14</sup>; H. Aldhalaan<sup>15</sup>; A. Al-Asmari<sup>16</sup>; E.A. Fageih<sup>16</sup>; J. Ahmed<sup>1</sup>; H.C. Kornau<sup>7,8</sup>; H. Bellen<sup>3,4,5,17</sup>; S.E. Antonarakis<sup>2,12,18</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Basic Medical Sciences, Khyber Medical University, Peshawar, Pakistan; 2) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, United States of America; 4) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, United States of America; 5) Howard Hughes Medical Institute, Houston TX, United States of America; 6) King Fahad Medical City, National Neuroscience Institute, Riyadh, Saudi Arabia; 7) Neuroscience Research Center (NWFZ), Charité - Universitätsmedizin Berlin, Berlin, Germany; 8) Center for Molecular Neurobiology (ZMNH), University of Hamburg, Hamburg, Germany; 9) Department of Medicine, KMU Institute of Medical Sciences, Kohat, Pakistan; 10) Radiology department, Lady reading Hospital, Peshawar, Pakistan; 11) Current address, Biomedical Research Foundation of the Academy of Athens, Athens, Greece; 12) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 13) Current address, Medigenome, The Swiss Institute of Genomic Medicine, Geneva, Switzerland; 14) Current address, Department of Endocrinology Diabetes and Metabolism, University Hospital of Lausanne, Switzerland; 15) King Faisal Specialist Hospital & Research Centre, Riyadh, Saudi Arabia; 16) Medical Genetics Department, Children's Hospital, King Fahad Medical City, Riyadh, Saudi Arabia; 17) Department of Neuroscience and Program in Developmental Biology, Baylor College of Medicine, Houston, TX, United States of America; 18) iGE3 Institute of Genetics and Genomics of Geneva, Geneva, Switzerland

---

We report two consanguineous families with probands that exhibit intellectual disability, developmental delay, short stature, aphasia, and hypotonia in which homozygous non-synonymous variants were identified in *IQSEC1* (NM\_001134382.3). In a Pakistani family the *IQSEC1* segregating variant is c.1028C>T:p.(Thr343Met), while in a Saudi family the variant is c.962G>A:p.(Arg321Gln). *IQSEC1*, 2, and 3 encode ARF-GEFs for ARF6 and their loss affects a variety of actin dependent cellular processes, including AMPA receptor trafficking at synapses. The ortholog of *IQSEC1* in the fly is *schizo* and its loss affects ROBO mediated growth cone guidance at the midline in the CNS, an actin dependent process. Ubiquitous overexpression of the reference *IQSEC1* cDNA in wild-type flies is lethal. However, overexpression of the two variant *IQSEC1* cDNAs does not affect viability, suggesting that they are loss of function alleles. Loss of *schizo* in flies causes embryonic lethality that can be rescued to 2<sup>nd</sup> instar larvae by moderate expression of the human reference cDNA. However, the p.(Arg321Gln) and p.(Thr343Met) variants fail to rescue this embryonic lethality. These data indicate that the variants behave as loss of function mutations in two independent assays. We also show that *Schizo* is expressed and required in photoreceptors for proper function based on Electroretinograms. Finally, mice with a conditional *lqsec1* deletion in cortical neurons exhibited an increased density of dendritic spines with an immature morphology. The phenotypic similarity of the patients and the

functional experiments in *Drosophila* and mice indicate that *IQSEC1* variants are the cause of a recessive disease with intellectual disability, developmental delay and short stature, and that glial growth cone, axonal guidance and dendritic projection defects as well as dendritic spine dysgenesis may underlie disease pathogenesis.



## PgmNr 1171: Genes contributing to brain malformation and neurodevelopmental syndromes in a Turkish population.

### Authors:

D. Pehlivan<sup>1,2</sup>; T. Mitani<sup>1</sup>; J. Punetha<sup>1</sup>; D. Marafi<sup>1</sup>; Z. Coban Akdemir<sup>1</sup>; A. Gezdirici<sup>3</sup>; S. Isikay<sup>4</sup>; E. Yilmaz-Gulec<sup>3</sup>; J.M. Fatih<sup>1</sup>; A. Ayaz<sup>3</sup>; I. Akalin<sup>5</sup>; S. Yilmaz<sup>5</sup>; E. Karaca<sup>1</sup>; Y. Bayram<sup>1,6</sup>; S. Jhangiani<sup>7</sup>; E. Boerwinkle<sup>7</sup>; R. Gibbs<sup>1,7</sup>; J.E. Posey<sup>1</sup>; J.R. Lupski<sup>1,7,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA; 3) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, 34303, Turkey; 4) Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, School of Health Sciences, Gaziantep, 27000, Turkey; 5) Department of Medical Genetics, Istanbul Medeniyet University, Faculty of Medicine, Istanbul, Turkey; 6) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029; 7) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, 77030, USA; 8) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA; 9) Texas Children's Hospital, Houston, Texas, 77030, USA

**Background:** Approximately 1 in 6 children have a Neurodevelopmental Disorder (NDD) resulting in significant sociocultural and economic burden to the society. We expanded a previously reported study of 128 Turkish families with brain malformations (Phase I of this study had an >85% solve rate and revealed 41 novel disease genes) to determine contributing genes and genetic architecture underlying NDD.

**Material and Methods:** We now report data from 270 families including 142 newly recruited Turkish families (phase II) in addition to the phase I cohort. We implemented family-based exome sequencing (ES) of probands and informative affected and unaffected family members based on pedigree structure. Potentially pathogenic, likely damaging variants of interest underwent Sanger confirmation and segregation analysis. Array Comparative Genomic Hybridization was performed in individuals found to have potential pathogenic copy number variants from ES data. Additionally, 13 of the families who remained undiagnosed during phase I of the study underwent exome re-analyses with expansion of ES to include additional family members.

**Results:** Of the 142 newly enrolled families, we achieved a potential molecular diagnosis in 83 families (58.4%) thus far including 19 new candidate disease genes (13.3%). Amongst the candidate disease genes, we have identified  $\geq 2$  affected unrelated individuals for five genes: *TUBGCP2*, *GRM7*, *PLXNA1*, *NTNG2*, and *TRAPPC4* either within this cohort, through collaborations via GeneMatcher or surveys of the literature. Multilocus pathogenic variation was observed in 8 of the 83 families, driven by absence of heterozygosity (AOH), i.e. AR + AR trait loci, apparently due to the high identity-by-descent (IBD) observed in this consanguineous cohort. Three of 13 families undiagnosed from phase I were resolved with reanalysis and expansion to family-based ES.

**Conclusion:** Our approach allowed us to propose 19 novel candidate NDD genes (phase II) and reiterate the importance of reanalysis for unsolved exome cases and the impact of oligogenic inheritance in a subset of cases with this complex trait. Also, our study further expands the clan genomics hypothesis to multi-locus models, confirming our observations in the arthrogryposis

complex trait, by revealing accumulation of rare homozygous deleterious variants within the clan is driven by IBD in consanguineous populations.

## PgmNr 1172: Copy number variation detection using NGS-based data: Statistics and intellectual disability case report.

### Authors:

K.M. Rocha <sup>1</sup>; J.R.M. Ceroni <sup>1,2</sup>; M. Lazar <sup>1</sup>; M.C. Varela <sup>1</sup>; G.L. Yamamoto <sup>1,2</sup>; V.N. Takahashi <sup>1</sup>; K. Bagatini <sup>1</sup>; D.G. Franco <sup>1</sup>; M.O. Scliar <sup>1</sup>; J.Y.T. Wang <sup>1</sup>; M. Zatz <sup>1</sup>; M.R. Passos-Bueno <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome and Stem Cell Research Center (HUG-CELL), University of São Paulo, Brazil Rua do Matão, travessa 13, 106 - 05508-090. São Paulo, SP, Brazil; 2) Department of Clinical Genetics, Children's Hospital, Medical School, University of São Paulo, Brazil Av. Dr. Enéas Carvalho de Aguiar, 647 - 05403-000. São Paulo, SP, Brazil

Deletions, and to a lesser extent, duplications involving a single or multiple exons, or even encompassing several genes, may be the molecular cause of several diseases. However, apart from well studied genes, such as *DMD* or *PMP22*, the frequency of these events is unknown or underestimated. To analyze CNV on NGS data we used XHMM (eXome Hidden Markov Model) and Nextgene (Softgenetics) applications. Some regions were validated by MLPA and currently, both applications are being used in our routine. From 2017 to 2019 we analyzed 578 patients that were divided in three major groups: neuromuscular disorders (NMD), skeletal disorders (SD) and developmental disorders (DD). Based solely on sequencing analysis, we have concluded the diagnosis in 43.9% of the cases (40.6%, NMD, 52.0%, SD and 44.9%, DD). The inclusion of CNV analysis has increased our yield to 53.3% (51.9%, NMD, 60.0%, SD and 51.9%, DD). This considerably improvement in our diagnosis yield is due to the detection of deletions/duplications not only in specific genes (*ATP7A*, *DYSF*, *LAMA2*, *NF1*, *SPG11*), but also in regions not primarily considered (18q deletion, 22q11 deletion).

As an example, we describe a *WAC* gene partial deletion detected in an 11yo girl referred for intellectual disability (ID) genetic testing. She was born to non-consanguineous parents and presented delayed neuropsychomotor development, mild-moderate ID (IQ=50), anxiety, underweight, hearing loss, hypertrichosis, synophrys, macrostomia and mild phenotypic abnormalities. DNA sequencing analysis using a panel containing ~6700 genes (TruSight One Expanded – Illumina) did not reveal SNVs that could explain her clinical features. On a subsequent CNV analysis using NGS data, two CNVs were detected (hg19): an heterozygous deletion encompassing exons 7-14 of *WAC* gene (minimal size chr10:28884662-28908535) and an heterozygous duplication involving exons 30-84 of *HUWE1* gene (minimal size chrX:53560270-53622363). Loss-of-function alterations in *WAC* gene (MIM \*615049) were identified in patients with ID and, to our knowledge, only one partial deletion was described in this gene (exons 5-14). On the other hand, the *HUWE1* partial duplication was not clearly associated with ID and was considered VUS. The patient's phenotype was attributed to *WAC* deletion. Though not all identified CNVs could be validated, we consider an essential step the use of at least two independent tools to detect these alterations.

Financial support: Fapesp/CEPID; CNPq/INCT

## PgmNr 1173: Complex genetic network underlying the convergent of Rett Syndrome like (RTT-L) phenotype in neurodevelopmental disorders.

### Authors:

S. Rangasamy <sup>1,2</sup>; E. Frankel <sup>1</sup>; J. Dodson <sup>1</sup>; R. Gupta <sup>1</sup>; M. Brzezinski <sup>1</sup>; P. Venugopal <sup>1</sup>; L. Llaci <sup>1</sup>; B. Gerald <sup>1</sup>; G. Mills <sup>1</sup>; K. Ramsey <sup>1,2</sup>; N. Belnap <sup>1,2</sup>; M. Sanchez-Castillo <sup>1,2</sup>; W. Jepsen <sup>1,2</sup>; M. Naymik <sup>1,2</sup>; C. Balak <sup>1,2</sup>; S. Szelinger <sup>1,2</sup>; I. Schrauwen <sup>1,2</sup>; I. Piras <sup>1,2</sup>; D. Craig <sup>1,2</sup>; M. Huentelman <sup>1,2</sup>; V. Narayanan <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Rare Childhood Disorders, Translational Genomics Research Institute, Phoenix, Arizona, United States; 2) Neurogenomics, Translational Genomics Research Institute (TGen), Phoenix, Arizona

---

Mutations of the X-linked gene encoding methyl-CpG-binding protein 2 (*MECP2*) cause classical forms of Rett syndrome (RTT) in females. Clinical diagnosis of typical (classical) and atypical forms of RTT requires a period of developmental regression followed by recovery or stabilization. Typical RTT is further characterized by meeting four necessary criteria and all exclusion criteria. Atypical RTT requires two to four necessary criteria, all exclusion criteria, and five of supportive criteria (Neul et al 2010). It is recognized that individuals with other phenotypes (autism, developmental delay, intellectual disability) may harbor *MECP2* mutations or duplications/deletions. Mutations in genes coding for cyclin-dependent kinase-like 5 (*CDKL5*) and forkhead box G1 (*FOXG1*) are more commonly found in patients with atypical Rett syndrome. There does exist a subset of patients who may have features of RTT, without meeting all diagnostic criteria for typical or atypical RTT, and are lacking a mutation in *MECP2*, *NTNG1*, *CDKL5*, or *FOXG1*. We refer to these individuals as having Rett syndrome like (RTT-L) disorder. Whole Exome Sequencing (WES) of eight patients in our cohort with RTT-L disorder revealed mutations in the genes *GABRG2*, *GRIN1*, *ATP1A2*, *KCNQ2*, *KCNB1*, *TCF4*, *SEMA6B*, and *GRIN2A*, which are seemingly unrelated to Rett-related genes. We hypothesized that the phenotypic overlap in RTT and RTT-L is caused by mutations in genes that affect common cellular pathways critical for normal brain development and function. We annotated a list of mutations in genes identified from peer-reviewed articles as causing RTT-L and its associated clinical phenotype, then performing a protein-protein interaction (PPI) network analysis of these genes. We also investigated their molecular interaction of RTT-L genes with *MECP2*, *CDKL5*, *NTNG1*, and *FOXG1*. We find evidence for a significant number of the molecular interaction among the RTT-L genes and with RTT genes. Genes linked to RTT and RTT-L disorder were highly interconnected and enriched in the biological pathways of circadian entrainment, the CREB pathway, and RET signaling, and neuronal processes like synaptic transmission, and transcription. We conclude that mutations in genes that significantly interact with the PPI network established by Rett-related genes often cause RTT-L, explaining the considerable feature overlap between RTT-L genes and Rett-related genes.

## PgmNr 1174: Second *Fxr2* knockout mouse demonstrates more severe *Fmr1/Fxr2* double KO phenotypes: Implications for Fragile X syndrome and *FMR1* function.

### Authors:

Y. Gu; D.L. Nelson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human and Molecular Genetics, Baylor College of Medicine, Houston, Texas.

Fragile X syndrome results from loss of *FMR1* function due to CGG repeat expansion and loss of mRNA expression. We previously demonstrated that mutation of the *Fmr1* paralog, *Fxr2*, enhanced phenotypes found in either *Fxr2* or *Fmr1* single KO mice, indicating likely compensation that could be relevant to severity of symptoms in fragile X syndrome and potentially to therapeutic strategies. We now report that a more complete ablation of *Fxr2* with a second KO allele provides more severe double KO phenotypes, suggesting a more significant role for *Fxr2* in the compensating for *Fmr1* mutation.

*Fxr2* KO1 showed hyperactivity and impairment of rotarod performance, while *Fmr1/Fxr2* double KO displayed additional phenotypes including aberrant circadian behavior and molecular rhythms and metabolic disturbances. The *Fxr2* KO1 allele has now been shown to encode an *Fxr2* protein lacking 39 amino acids due to an in-frame deletion; residual activity has confounded analysis.

We generated a second *Fxr2* KO (KO2) allele while developing a conditional KO allele. The mRNA is predicted to encode only the N-terminal 79 amino acids of *Fxr2* and appears to represent a *bona fide* *Fxr2* null. Interestingly, *Fxr2* KO2 shows more severe phenotypes than *Fxr2* KO1, supporting the notion that KO1 was incomplete. Of particular note, recovery of *Fmr1/Fxr2* double KO male mice is severely reduced from expectations and fewer than half of the expected *Fxr2* KO2/KO2; *Fmr1* KO/+ females are recovered. These difficulties are much more severe than those found with the *Fxr2* KO1 allele. Efforts to determine where embryos are lost are underway.

We characterized *Fxr2* KO2 along with intermediate *Fxr2* mutant lines (generated to create the conditional KO) for weight, fecundity and behavior. All lines showed reduced weight and all but *Fxr2* KO2 produced fewer than half the expected homozygous pups. Molecular analysis confirmed that the intermediate lines: *Fxr2* Tm2a, Tm2b, Tm2f and Tm2g encode mRNAs that produce truncated *Fxr2* proteins containing the original Tudor domain 1 coupled to a novel chimeric Tudor domain 2. We hypothesize that truncated proteins encoded by the intermediate lines not only have lost *Fxr2* function but also interfere with the function(s) of *Fmr1* and/or *Fxr1*.

Understanding the functions and capacities of all three members of the *FMR1/FXR1/FXR2* family to cross-compensate is important for determining the effects of loss of *FMR1* function in Fragile X syndrome and could suggest therapeutic interventions.

## PgmNr 1175: Fragile X syndrome: Genetic testing in a large Cameroon family.

### Authors:

K. Kengne Kamga <sup>1</sup>; S. Nguetack <sup>2</sup>; M. Khuthala <sup>1</sup>; E. Wonkam Tingang <sup>1</sup>; S. Nchangwi Munung <sup>1</sup>; A. Wonkam <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology, Division of Human Genetics, University of Cape Town, South Africa; 2) Department of Paediatrics, University of Yaounde 1, Cameroon

---

Fragile X Syndrome(FXS) is known to be the main monogenic cause of Intellectual Disability(ID). Despite reported cases of FXS in some African countries, genetic testing remains largely unavailable in most healthcare settings in Africa. We report here FXS in a large polygamous family from Cameroon, with confirmed molecular diagnosis.

We conducted a cross-sectional study in the proband's extended family. Snowball sampling was used to identify and recruit potential participants for FXS testing. Demographic information such as age, sex, clinical characteristics were collected. Pedigrees were drawn using Cyrillic 3.0. DNA were extracted from peripheral blood of any consented family members, or their parents. Molecular diagnosis of FXS was performed at the National Health Laboratory Services, Cape Town, South Africa. Data was analysed using Epi-info7.2.

A total of 46 individuals were tested for FXS. Of these, 58.7%(n=27/46) were females. The mean age was 9.4(±5.0) years for children and 45.9(±15.9) years for adults. Up to 8.7%(n=4) were clinically affected with severe ID, all were males with full mutation (more than 200CGG). 13.0%(n=6) presented with mild ID of which, 3 were females with a full mutation, 2 were females with a premutation(101, 91 CGG) and a normal male(42 CGG). Study of the family pedigree suggested that the founder of these families had more than 20 wives and was likely a FXS normal transmitting male.

No case of Fragile-X-Tremor-Ataxia syndrome was reported. An affected male with a confirm diagnosis of FXS had 3 daughters(112, 120, 120 CGG), while two women aged 63years (101CGG) and 64years (91CGG) reported premature menopause. Among females(n=27), 14.8% had a full mutation, 37.0% a permutation and 48.2% were normal. Moreover, 21.1%(n=4) males had a full nutation and the rest were normal. Being a member of the first normal transmitting male's lineage is a risk factor for having a premutation or full mutation for FXS(P=0.0004).

This study describes the pattern of a genetic transmission of FXS in an exceptionally large Cameroon family, and will inform further retrospective counselling and testing for reproductive risk, and risk for developing premature ovarian insufficiency. The family will also help in exploring challenges associated with the return of an individual genetic results in a rural setting in Africa. The large family could help in exploring potential genetic modifiers of FXS, including differential methylation status in female carriers.

## PgmNr 1176: *In vivo* role of FMRP phosphorylation in fragile X syndrome.

### Authors:

Y. Wen <sup>1</sup>; K. Zhang <sup>1</sup>; S. Warren <sup>1,2</sup>; P. Jin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA; 2) Department of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta, GA.

---

Fragile X syndrome (OMIM 309550), the most frequent cause of inherited intellectual disability and a well-known cause of autism, is due to the loss of functional fragile X mental retardation protein (FMRP), an RNA-binding protein. Previous studies have shown that phosphorylated FMRP can inhibit the translations of its mRNA targets while unphosphorylated FMRP allows translation to proceed. Serine499 is the primarily conserved phosphorylated residue in mouse FMRP, and the mutation *p.Ser499Asp* (S499D) is associated with stalled ribosomes that inhibit downstream translation, thereby mimicking the phosphorylation of FMRP *in vivo*. To further investigate the function and mechanism of FMRP phosphorylation in *FMR1*-related intellectual disabilities, we generated an *Fmr1* S499D (c.1495\_1497TCT>GAC) mutant mice using the CRISPR/Cas9 Engineering System. No phosphorylated FMRP was detected in the mutant line. Using the fibroblasts derived from this mutant line, we determined the half-life of FMRP and found that the mutant S499D FMRP displayed a longer half-life compared to wildtype FMRP. Furthermore, we observed increased anxiety/stress in the *mutant* S499D males. These data together suggest that FMRP phosphorylation could play an important role in FMRP-mediated regulation of neurodevelopment and behaviors.



## PgmNr 1177: Paralog studies augment gene discovery: *DDX* and *DHX* genes.

### Authors:

I. Paine<sup>1</sup>; J. Posey<sup>1</sup>; C. Grochowski<sup>1</sup>; K. Wang<sup>2</sup>; M. Pinelli<sup>3,4</sup>; Z. Coban Akdemir<sup>1</sup>; J. Hui<sup>5</sup>; W. Lan Yeung<sup>6</sup>; B. Wong<sup>7,8</sup>; L. Ortega<sup>9</sup>; M. Reza Bekheirnia<sup>1,10,11</sup>; T. Bierhals<sup>12</sup>; J. Johannsen<sup>13</sup>; S. Bozdogan<sup>14</sup>; H. Aydin<sup>15</sup>; E. Karaca<sup>16</sup>; H. Ityel<sup>17</sup>; J. White<sup>18</sup>; E. Wilichowski<sup>19</sup>; S. Wortmann<sup>20,21,22</sup>; E. Casella<sup>23</sup>; F. Kok<sup>24,25</sup>; N. Brunetti-Pierri<sup>3,4</sup>; F. Hildebrandt<sup>17</sup>; A. Brautbar<sup>26</sup>; I. Van den Veyver<sup>1,7</sup>; I. Glass<sup>27</sup>; D. Lessel<sup>12</sup>; G. Lyon<sup>28,29,30</sup>; J. Lupski<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, Texas.; 2) Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA; 3) Department of Translational Medicine, University of Naples "Federico II", 80138 Napoli, Italy; 4) Telethon Institute of Genetics and Medicine (TIGEM), 80078 Pozzuoli, Italy; 5) Department of Pediatrics, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong SAR, China; 6) Department of Pediatrics and Adolescent Medicine, Alice Ho Miu Ling Nethersole Hospital, Hong Kong SAR, China; 7) Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX, 77030, USA; 8) The Jan & Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas, 77030, USA; 9) Cook Children's Genetics, Fort Worth TX, 76104, USA; 10) Department of Pediatrics, section of Pediatric Renal, Baylor College of Medicine, Houston, TX, 77030, USA; 11) Department of Genetics, Texas Children's Hospital, Houston, TX, 76104, USA; 12) Institute of Human Genetics, University Medical Center Hamburg-Eppendorf, 20251 Hamburg, Germany; 13) Department of Pediatrics, University Medical Center Hamburg-Eppendorf, 20251 Hamburg, Germany; 14) Department of Medical Genetics, Cukurova University Faculty of Medicine, 01330 Adana, Turkey; 15) Department of Medical Genetics, Medical Faculty of Namik Kemal University, Tekirdag 59100, Turkey; 16) Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, 35294, USA; 17) Division of Nephrology, Department of Pediatrics, Harvard Medical School, Boston MA, 02115, USA; 18) Department of Pediatrics, University of Washington, Seattle, WA, 98195, USA; 19) Department of Pediatrics and Pediatric Neurology, Georg-August-Universität Göttingen, 37075 Göttingen, Germany; 20) Institute of Human Genetics, Technical University München, 81675 Munich, Germany; 21) Institute of Human Genetics, Helmholtz Zentrum München, 85764 Neuherberg, Germany; 22) University Children's Hospital, Paracelsus Medical University, 5020 Salzburg, Austria; 23) Children's Institute, Hospital das Clinicas, University of Sao Paulo, 05405-000 Sao Paulo, Brazil; 24) Mendelics Genomic Analysis, 04013-000 Sao Paulo, Brazil; 25) Department of Neurology, University of Sao Paulo School of Medicine, 01246-903 Sao Paulo, Brazil; 26) Medical Genetics Department, Cook Children's Hospital, Fort Worth TX, 76104, USA; 27) Division of Genetic Medicine, University of Washington, Seattle, Washington, 98195, USA; 28) Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, NY, 11724, USA; 29) 5.Utah Foundation for Biomedical Research, Salt Lake City, UT, 84107, USA; 30) Institute for Basic Research in Developmental Disabilities (IBR), Staten Island, New York, 10314, USA

---

Members of a paralogous gene family in which variation in one gene is known to cause disease, are eight times more likely to also be associated with human disease. Recent studies have elucidated *DHX30* and *DDX3X* as genes for which pathogenic variant alleles are involved in neurodevelopmental



disorders. We hypothesized that variants in paralogous genes encoding members of the DExD/H-box RNA helicase superfamily may also underlie developmental delay/intellectual disability (DD/ID) disease phenotypes. Here we describe 18 unrelated individuals with DD/ID, CNS dysfunction, vertebral anomalies, and dysmorphic features who were found to have likely damaging variants in DExD/H-box RNA helicase genes. In addition, these individuals exhibit a variety of other tissue and organ system involvement including ocular, outer ear, hearing, cardiac, and kidney tissues. Five individuals with homozygous (1), compound heterozygous (2), or *de novo* (2) missense variants in *DHX37* were identified by exome sequencing. We identified ten total individuals with missense variants in three other DDX/DHX paralogs, *DHX16* (4 individuals), *DDX54* (3 individuals), and *DHX34* (3 individuals). Most identified variants are rare, predicted to be damaging, and occur at conserved amino acid residues. Taken together, these 15 individuals implicate the DExD/H-box helicases in both dominantly and recessively inherited neurodevelopmental phenotypes and highlight the potential for more than one disease mechanism underlying these disorders.

## PgmNr 1178: Solving the unsolved: Integrated analysis of genome and transcriptome sequencing data identifies structural variants responsible for rare developmental delay syndromes.

### Authors:

A. Vitobello<sup>1</sup>; A. Plagos<sup>1</sup>; F. Tran Mau-Them<sup>1</sup>; E. Tisserant<sup>1</sup>; Y. Duffourd<sup>1</sup>; P. Callier<sup>2</sup>; A.L. Bruel<sup>1</sup>; A.S. Denomme-Pichon<sup>1</sup>; C. Philippe<sup>1</sup>; B. Isidor<sup>3</sup>; S. Heide<sup>4</sup>; A. Afenjar<sup>5</sup>; D. Rodriguez<sup>6</sup>; C. Mignot<sup>4</sup>; D. Heron<sup>4</sup>; M. Vincent<sup>3</sup>; P. Charles<sup>4</sup>; S. Moutton<sup>7</sup>; S. Odent<sup>8</sup>; N. Jean<sup>7</sup>; C. Dubourg<sup>9</sup>; A. Faudet<sup>4</sup>; B. Keren<sup>4</sup>; J. Thevenon<sup>7</sup>; A. Boland<sup>10</sup>; R. Olaso<sup>10</sup>; L. Faivre<sup>7</sup>; C. Thauvin-Robinet<sup>7</sup>; J.F. Deleuze<sup>10</sup>; A. Vitobello<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UMR1231 GAD, Inserm - University of Burgundy-Franche Comté, Dijon, France; 2) UF Innovation en diagnostic génomique des maladies rares, CHU Dijon, France; 3) CHU de Nantes, Service de Génétique Médicale, Nantes, France; 4) AP-HP, Département de Génétique, Hôpital de la Pitié-Salpêtrière, 75013, Paris, France et Centre de Référence "déficiences intellectuelles de causes rares", 75013 Paris, France.; 5) APHP, Département de génétique et embryologie médicale, Hôpital Armand Trousseau, F-75012 Paris, France; 6) Centre de Référence de Neurogénétique et Service de Neurologie Pédiatrique, AP-HP, Hôpital Armand Trousseau, GHUEP; Sorbonne Université, GRC n°19, pathologies Congénitales du Cervelet-LeucoDystrophies, Paris, France; 7) Centre de Référence maladies rares « Anomalies du développement et syndromes malformatifs », centre de génétique, FHU-TRANSLAD, Dijon University Hospital, Dijon, France; 8) CHU de RENNES, Genetique; Univ Rennes, CNRS, IGDR (Institut de Génétique et Développement de Rennes) UMR 6290, Rennes, France; 9) Service de Génétique Moléculaire et Génomique, BMT-HC « Jean Dausset », CHU Pontchaillou, 2 rue Henri Le Guilloux, 35033 RENNES Cedex 9, France; 10) Centre National de Recherche en Génomique Humaine (CNRGH), Institut de Biologie François Jacob, CEA, Université Paris-Saclay, F-91057, Evry, France, and LabEx GENMED

---

Exome sequencing (ES) represents the first-tier diagnostic test in patient presenting with intellectual disability (ID)/developmental delay (DD) with suspected monogenic etiology and negative array-CGH results. Yet, almost 50% of these patients remain unsolved, at the molecular lever, arguing the potential interest to extend the genetic investigations beyond the protein-coding genome. We deployed trio genome (GS) followed by, when available, blood transcriptome sequencing (RNA-seq) as part of an integrated strategy to evaluate the utility of molecular approaches beyond the exome, to solve a cohort of 41 patients unexplained by standard clinical ES and array-CGH analyses. We identified *de novo* balanced inversions disrupting the Topologically Associating Domain regulating the neural expression of *MEF2C*, responsible for the epileptic encephalopathy presented by one patient, and disrupting *BCL11A* in a patient presenting with syndromic DD respectively. Two complex SVs consisting of deletions-inversions affecting *WWOX* and *CASK* were detected in patients presenting with epileptic encephalopathy and syndromic DI respectively. *WWOX* patient was compound heterozygous for the SV inherited from the unaffected father and a pathogenic missense inherited from the mother. The SV caused the inversion of exon 5, which underwent exon skipping as showed by RNA-seq results. No *WWOX* protein was detected by western blot analysis. *CASK* patient was a

female carrying a *de novo* 9.4 Kb variant responsible for the loss of 3 exons. RNA-seq demonstrated the creation of a shorter transcript altering *CASK* reading frame. One patient, clinically compatible with a Simpson-Golabi-Behmel syndrome 1, carried a deep intronic hemizygous deletion of *GPC3*, identified by array-CGH. GS identified a more complex SV involving an intra-chromosomal insertion-deletion disrupting *GPC3* transcriptional unit. RNA-seq demonstrated the absence of *GPC3* expression and a transcriptional profile compatible with a perturbed Glypican pathway. A patient presented a *de novo* partial deletion of *GATAD2B*. Lastly GS identified two *de novo* pathogenic variants affecting nucleotides not covered by ES in *PURA* and *FOXG1*.

Overall, GS followed by non-systematic RNA-seq, allowed the identification of the causative events underlying unexplained DI/DD patients in 8/41 (20%) cases, highlighting the interest in deploying multi-omics approaches to solve unsolved Mendelian disorders.

## **PgmNr 1179: Expanding the AUTS2-associated syndrome: Novel variants in AUTS2 are associated with a severe neurodevelopmental disorder that overlaps with Rubenstein-Taybi syndrome.**

### **Authors:**

K. Aldinger<sup>1</sup>; C. Cheng<sup>1</sup>; Z. Thomson<sup>1</sup>; S. Michels<sup>1</sup>; M. Dave<sup>1</sup>; S. Caraffi<sup>2</sup>; I. Ivanovski<sup>2</sup>; E. Errichiello<sup>3</sup>; L. Garavelli<sup>2</sup>; O. Zuffardi<sup>3</sup>; C. Zweier<sup>4</sup>; M. Schneider<sup>5</sup>; A. Papavasiliou<sup>6</sup>; S. Bakhtiari<sup>7</sup>; M. Kruer<sup>7</sup>; H. Magee<sup>7</sup>; T. Leis<sup>8</sup>; M.S. Perry<sup>9</sup>; J. Humberson<sup>10</sup>; J. Rotenberg<sup>11</sup>; M.T. Cho<sup>12</sup>; W. Dobyns<sup>1,13</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington.; 2) Struttura Semplice Dipartimentale di Genetic Medica, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy; 3) Dipartimento di Medicina Molecolare, Università di Pavia, Pavia, Italy; 4) Institute of Human Genetics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 5) Carle Physicians Group, St. Christopher's Hospital for Children, Urbana, Illinois; 6) Neurology, IASO Children's Hospital, Athens, Greece; 7) Barrow Neurological Institute, Phoenix Children's Hospital, Phoenix, Arizona; 8) Department of Pediatrics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany; 9) Comprehensive Epilepsy Program, Jane and John Justin Neuroscience Center, Cook Children's Medical Center, Fort Worth, Texas; 10) Division of Genetics, Department of Pediatrics, University of Virginia Children's Hospital, Charlottesville, Virginia; 11) Houston Specialty Clinic, Houston, Texas; 12) GeneDx, Inc., Gaithersburg, Maryland; 13) Department of Pediatrics and Neurology, University of Washington, Seattle, Washington

---

The AUTS2-associated neurodevelopmental syndrome is characterized by mild to moderate intellectual disability, mild microcephaly, variable facial dysmorphism, and feeding difficulty in individuals with chromosomal arrangements, especially deletions, that impact the *AUTS2* gene. We report 8 new individuals with heterozygous *de novo* variants in *AUTS2*, including 5 with missense or indel variants in exon 9 who have more severe clinical features than any of the individuals previously reported with the AUTS2-associated syndrome. Clinical and neuroimaging review uncovered two distinct presentations. The first group included 3 individuals with predicted truncating mutations and clinical features similar to published reports. The second group included 5 individuals with missense mutations or in-frame deletions of exon 9 downstream of a second translational start site for this bicistronic protein and clinical features distinct from the published AUTS2-associated phenotype. This second group included one individual with a severe epileptic encephalopathy and four individuals with a novel, recurrent mutation that have normal head size and dysmorphic facial features similar to Rubinstein-Taybi syndrome, as well as partial agenesis of the corpus callosum and mild to moderate cerebellar hypoplasia. The clinical features observed among individuals in group 2 were much more severe than patients with AUTS2 predicted loss-of-function, suggesting the exon 9 mutations have severe functional consequences, possibly by selectively affecting expression of the short isoform. The spectrum of disease severity, type of AUTS2 aberration, and phenotypic features further delineates the AUTS2-associated syndrome.

## PgmNr 1180: Phenotypic expansion and neuroimaging findings in *CSNK2A1*: Okur-Chung neurodevelopmental syndrome, case series report.

### Authors:

S. Bachir <sup>1</sup>; S. Poskanzer <sup>2,3</sup>; K. Manickam <sup>2,3</sup>; K. Abell <sup>1</sup>; B. Simpson <sup>1</sup>; C. Rajakaruna <sup>1</sup>; N. Weaver <sup>1</sup>; R.J. Hopkin <sup>1</sup>; C. Prada <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics , Cincinnati Children's Hospital Medical Center, Cincinnati , OH.; 2) Division of Genetic and Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; 3) Department of Pediatrics, The Ohio State University College of Medicine, OH

---

Okur-Chung Neurodevelopmental syndrome (OCNDS) is a recently described autosomal dominant condition caused by mutations in *CSNK2A1*, with clinical heterogeneity comprising of intellectual disability, global developmental delays and facial features. Brain MRI anomalies are present in two thirds of published cases, these have included: cortical malformations, delayed myelination, vermian hypoplasia and hydrocephalus. Of the total 27 published cases, 25 are missense mutations while two are splice site in the *CSNK2A1* gene. We identify the first nonsense variant along with novel neuroimaging findings in three out of four newly reported cases of OCNDS.

Our cases range between 22 months to 32 years old (y/o). They all presented with developmental delays, intellectual disability, dysmorphic features and all but one with seizures. Diagnoses were made via Whole Exome Sequencing or Autism/Intellectual Disability panel. Case 1 is a 32 y/o male who is nonverbal with autism and ataxia, the latter correlating with his brain MRI finding of cerebellar vermian hypoplasia. Serial brain images revealed occipital volume loss, which may correlate with his bilateral astigmatism. Case 2, is a 7y/o female with bilateral astigmatism and cataracts requiring surgery, her brain MRI revealed hydrocephalus resulting in optic tortuosity and required shunting.

Case 3 is a 9 y/o female with anxiety and seizures. Brain MRI revealed delayed myelination. She has a missense mutation in p.K198R, which is in a highly conserved region in the activation segment of *CSNK2A1*. Interestingly another published case with the same mutation has similar neuroimaging findings. Case 4 is a 22-month-old female with the first documented nonsense variant in p.R306X. She presented with severe oral motor dysphagia, cardiac malformations and dysmorphic features; cleft lip, hypertelorism with depressed nasal bridge.

Our cases expand the phenotype for ophthalmologic and neuroimaging findings seen in OCNDS. Also for the first time we illustrate genotype correlation with neuroimaging phenotype. Recent evidence indicates that CK2 kinase regulates eye morphogenesis in *Drosophila* indicating a possible role in eye pathogenesis. Since the syndrome is characterized by developmental delays, it may be difficult to ascertain visual changes, thus physicians should be aware. Furthermore, given the recent discovery and paucity of cases, management guidelines such as the need for neuroimaging in OCNDS will need to be developed.

# PgmNr 1181: Exploration of proteasome-related pathomechanisms involved in *USP7*-associated intellectual disability.

## Authors:

W. Deb <sup>1</sup>; S. Küry <sup>1,2</sup>; B. Isidor <sup>2</sup>; F. Ebstein <sup>3</sup>; B. Cogné <sup>2</sup>; E. Krüger <sup>3</sup>; R. Redon <sup>1</sup>; T. Besnard <sup>1,2</sup>; V. Vignard <sup>1,2</sup>; S. Mercier <sup>1,2</sup>; S. Béziau <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) L'institut du thorax, INSERM, CNRS, UNIV Nantes, Nantes, France.; 2) L'institut du thorax, CHU Nantes, Service de Génétique Médicale, Nantes, France.; 3) Institut für Medizinische Biochemie und Molekularbiologie, University of Greifswald, Germany.

---

**Introduction.** Ubiquitin Specific Peptidase 7 (USP7) is a deubiquitinating enzyme involved in syndromic intellectual disability. Its interaction with ubiquitin-tagged proteins and its once described bond with the proteasome strongly suggest a deeper link, but no evidence of direct interaction has been validated and characterized. Several proteins of the Ubiquitin Proteasome System have already been shown as causative of neurodevelopmental diseases. Our goal was to investigate relation between USP7 and the proteasome structure and function, to guide us in the understanding of USP7 role in neurodevelopmental diseases.

**Material and Methods.** We studied 2 families: one patient and one control (father) in the first, one patient and two controls (father/mother) in the second. Both patients carry a de novo missense mutation in USP7. The first lies in the catalytic domain, and the second in a structurally important domain: UBL-2. We investigated the proteasome using SDS and Native-Page Western Blots as well as chymotrypsin-like activity assays, and begun exploring key protein in neurodevelopmental pathways.

**Results.** USP7 levels differed between our patients: they were normal in patient 1 with a variant in the catalytic domain, and reduced in patient 2 with a variant in UBL-2. Proteasomal catalytic activity was reduced in patient 1, but normal in patient 2. We also identified dysregulated proteins in the Sonic HedgeHog pathway in patient 1. Any such findings were not noted in family 2.

**Discussion.** Interestingly, we highlighted a potential proteasomal dysfunction in a patient with a pathogenic variant in the catalytic domain of USP7, as well as a dysregulated Shh pathway that could both be involved in the neurodevelopmental part of the disease. The absence of similar findings in patient 2 likely relates to its different type of variant and effect on protein levels of USP7. This observation should help to further understand physio/pathological functions of USP7 and establish genotype-phenotype correlations.

## PgmNr 1182: Neuropsychological study in 14 French patients with White-Sutton syndrome and *POGZ* mutations.

### Authors:

A. Garde <sup>1,2</sup>; J. Cornaton <sup>1</sup>; A. Sorlin <sup>1,2</sup>; S. Moutton <sup>1,2</sup>; C. Nicolas <sup>1</sup>; D. Geneviève <sup>3</sup>; L. Perrin <sup>4</sup>; P. Khau-Van-Kien <sup>5</sup>; T. Smol <sup>6</sup>; C. Vincent-Delorme <sup>6</sup>; B. Isidor <sup>7</sup>; B. Cogne <sup>7</sup>; A. Afenjar <sup>8</sup>; B. Keren <sup>9</sup>; C. Coubes <sup>3</sup>; F. Prieur <sup>10</sup>; A. Toutain <sup>11</sup>; A. Piton <sup>12</sup>; B. Gérard <sup>12</sup>; A. Vitobello <sup>2</sup>; C. Philippe <sup>2</sup>; C. Christel-Thauvin <sup>1,2</sup>; L. Faivre <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre de Référence Anomalies du Développement et Syndromes Malformatifs, FHU TRANSLAD, Hôpital d'Enfants, CHU Dijon, Dijon, France; 2) Inserm UMR1231 GAD, Génétique des Anomalies du Développement, Université de Bourgogne, Dijon, France; 3) Département de Génétique Médicale, Maladies rares et Médecine Personnalisée, CHRU de Montpellier, Montpellier, France; 4) AP-HP, Department of Genetics, Hôpital Robert Debré, Paris, France.; 5) Unité de Génétique Médicale et Cytogénétique, CHU de Nîmes, Nîmes, France.; 6) Service de génétique clinique Guy Fontaine, CHRU de Lille-Hôpital Jeanne de Flandre, Lille, France; 7) Service de Génétique Médicale, Hôpital Hôtel-Dieu, Centre Hospitalier Universitaire de Nantes, 44093 Nantes, France.; 8) Medical Genetics Department, AP-HP, University hospital Armand Trousseau, Paris, France.; 9) APHP, Hôpital Pitié-Salpêtrière, Département de Génétique, Centre de Référence Déficience Intellectuelle de Causes Rares, GRC UPMC «Déficience Intellectuelle et Autisme», Paris, France.; 10) Service de Génétique Médicale, CHU de St Etienne, Hôpital Nord, Saint Etienne, France.; 11) Service de Génétique Clinique, CHRU de Tours, Tours, France; 12) Laboratoire de Diagnostic Génétique, Hôpitaux Universitaires de Strasbourg, Strasbourg, France.

**Introduction** – White-Sutton syndrome is a rare developmental disease characterized by a global developmental delay, intellectual disabilities, and neurobehavioral abnormalities secondary to pathogenic *POGZ* variant. Although *POGZ* is described as an intellectual disability gene, some patients may only present learning disability. Thus, the purpose of our study was to describe the neurocognitive phenotype in patients with identified *POGZ* pathogenic variants.

**Materials and methods** – This study is based on a national collaboration through the AnDDI-Rares network, and includes 14 patients from 13 families (10 females, 4 males) with *POGZ* pathogenic variants, with a mean age of 11.7 years. All clinical data and neuropsychological tests have been collected from medical files by the referring physician.

**Results** – In our cohort, we observed 13 *POGZ* variants, including 5 frameshift, 3 nonsense, 2 splice variants and 3 missense. All cases were sporadic except a family following autosomal dominant inheritance. Among 14 patients, 10 patients exhibit intellectual disability: 5 mild ID, 2 moderate ID and 3 severe ID. The 4 remaining patients have learning disabilities with a similar neurocognitive profile, including dysexecutive syndrome, attentional disorders, slowness and difficulties on the affect recognition. 10/14 patients also exhibit a behavioral disorder, including anxiety in 6 patients, stereotypies in 5 patients, inappropriate behavior in 4 patients, and social withdrawal in 5 patients. Sleep difficulties were reported in 4 patients.

**Conclusion** – This study reveals that the cognitive phenotype of patients with *POGZ* mutation may range from learning disabilities to severe intellectual disabilities. This point is of importance for newly diagnosed young patients. It also highlights that same genes can be reported in a large spectrum of

neurocognitive profile, and that children with learning disabilities could benefit from next generation sequencing techniques such as exome sequencing.



## PgmNr 1183: Cause or coincidence: Explore the pathogenicity of *de novo* missense variants in *AHDC1*.

### Authors:

Y. Jiang<sup>1,2</sup>; D.R. Murdock<sup>1,2</sup>; M.F. Wangler<sup>2,3</sup>; G. Herman<sup>4</sup>; J. Friedman<sup>5</sup>; J.A. Mokry<sup>2</sup>; H. Li<sup>1,2</sup>; J. Traynelis<sup>1,2</sup>; A.W. Hansen<sup>1,2</sup>; M.M. Khayat<sup>1,2</sup>; K. McWalter<sup>6</sup>; J.R. Lupski<sup>1,2,3,7</sup>; R.A. Gibbs<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College Medicine, Houston, Texas.; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 4) Department of Pediatrics, Nationwide Children's Research Institute, Columbus, OH; 5) Department of Neurosciences and Pediatrics, University of California San Diego, San Diego, CA; 6) GeneDx, Gaithersburg, Maryland; 7) Texas Children's Hospital, Houston, TX

**Background:** Xia-Gibbs Syndrome (OMIM: 615829: XGS) is a Mendelian disorder discovered in 2014 and characterized by hypotonia, sleep apnea, intellectual disability and developmental delay. The disorder can result from *de novo* truncating mutations in the *AHDC1* gene located on chromosome 1p36.11 (Xia *et al.*, 2014). Many missense variants in *AHDC1* were reported in public databases of putatively healthy people and it is unclear whether some missense alleles in *AHDC1* can be pathogenic. Here, we have aggregated data from eight cases with *de novo* *AHDC1* missense mutations and provide guidelines for consideration of missense mutations at this locus, for possible assignment as disease causing variation.

**Methods:** Patients were recruited through the XGS Registry which was hosted in a secure, HIPAA compliant server (Jiang *et al.*, 2018). Families of patients submitted clinical and genetic reports allowing detailed assessment of genotypic and phenotypic information. Additional cases were identified through searching in public databases and literature review. Multiple comprehensive computational algorithms were used to predict the protein damage from missense alleles.

**Results:** We identified eight cases bearing seven different *de novo* missense mutations in *AHDC1* (two were previously reported (Gumus, 2019; Guipponi *et al.*, 2014)). The phenotypes of these missense cases showed overlapping features with classic XGS syndrome, including speech delay, intellectual disability, hypotonia. There are also non-overlapping phenotypes observed such as hypercalcuria, 'molar tooth sign' in MRI images. Multiple computational algorithms predict that three of the *de novo* mutations that are clustered in a region with an AT-Hook binding motif (p.Gly537Asp, Gly792Arg and Asp607Asn) have a high potential of being deleterious. In contrast, one of the *de novo* variants (p.Gly792Arg) is predicted to be not-damaging. That case also is suspected of a diagnosis of Joubert Syndrome based on her phenotype, although not genetically confirmed.

**Conclusion:** The pathogenicity of *de novo* missense mutations in *AHDC1* is still uncertain. However, a preponderance of *de novo* mutations, together with computational models, have suggested that mutations occurring in some regions of *AHDC1* may be damaging. Aggregating case reports of individuals carrying *de novo* missense mutations in the *AHDC1* gene, cellular functional studies and mouse models will lead to better understanding of the pathogenicity of such mutations.

## PgmNr 1184: Pathogenesis of *CDK8*-family-associated mediatoropathy: Three patients with *CDK8* or *CDK19* variants and functional analyses using zebrafish model.

### Authors:

T. Uehara <sup>1</sup>; K. Abe <sup>2</sup>; M. Oginuma <sup>2</sup>; N. Okamoto <sup>3</sup>; H. Yoshihashi <sup>4</sup>; S. Mizuno <sup>5</sup>; K. Kosaki <sup>1</sup>; T. Ishitani <sup>2,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Tokyo-to, Japan; 2) Department of Molecular Medicine, Institute for Molecular & Cellular Regulation, Gunma University, Gunma, Japan; 3) Department of Medical Genetics, Osaka Women's and Children's Hospital; 4) Department of Medical Genetics, Tokyo Metropolitan Children's Medical Center; 5) Department of Pediatrics, Aichi Developmental Disability Center Central Hospital, Aichi, Japan; 6) Department of Homeostatic Regulation, Research Institute for Microbial Diseases, Osaka University

---

CDK8 is a conserved serine/threonine kinase and is a member of the mediator complex that regulates a various transcription regulators including  $\beta$ -catenin and Notch. Mutations in *MED12* and *MED13* as other members of the complex are associated with syndromic intellectual disability (ID) and collectively referred to as “mediatopathies”. As expected, 12 patients with syndromic ID and missense mutations in *CDK8* or one patient with syndromic ID and haploinsufficiency of *CDK19* gene (a homolog of *CDK8*) were reported. The association between *CDK8*-family genes mutations and mediatopathy-resembled disorders is yet to be delineated. Here we report three patients with *de novo* missense mutations in *CDK8*-family genes and the mechanistic basis of *CDK8*-family-associated disorders with zebrafish model. Patient 1 was a 2-year-old male with ID, microcephaly, and congenital heart defects. Hypertelorism, low-set ears, a short philtrum, a high nasal bridge, a high arched palate, and a micrognathia were existed. Patient 2 was an 8-year-old male with ID. Hypertelorism, a short philtrum, a wide nasal bridge, and bifid uvula were existed. Patient 3 was a 12-year-old male with ID and microcephaly. Sparse hair, arched eyebrow, hypertelorism, low-set ears, a short philtrum, a wide nasal bridge, and a broad columella were existed. Exome analysis showed patient 1 and patient 2 had *de novo* missense variants in *CDK8* (p.G28A and p.N156S, respectively), and patient 3 had a *de novo* missense variant in *CDK19* (p.Phe197Leu). Pathogenicity of the variants were evaluated by various assays. 1) *In vitro* kinase assay of human CDK8 showed that enzymes with p.G28A or p.N156S substitution exhibited decreased kinase activity. 2) Zebrafish overexpression analyses showed that p.G28A or p.N156S alleles were hypomorphic alleles *in vivo*. Importantly, inhibition of CDK8 and CDK19 kinase activity in zebrafish embryos using morpholino antisense oligonucleotide or specific chemical inhibitor induced the craniofacial and heart defects related to the patients' phenotype. We suggest inhibition of Wnt/ $\beta$ -catenin signaling in developing neural crest cells may play a role in the pathogenesis of *CDK8*-family-associated disorders because of phenotypic overlap between inhibition of Cdk8 and Cdk19 activity and Wnt/ $\beta$ -catenin signaling. In conclusion, zebrafish studies showed that missense mutations in kinase domain of *CDK8* or *CDK19* cause human congenital defects as hypomorphic alleles.

## PgmNr 1185: Truncated AHDC1 protein induces a nucleolar stress response in Xia-Gibbs syndrome.

### Authors:

M.M. Khayat <sup>1,2</sup>; Y. Jiang <sup>1,2</sup>; V. Chander <sup>1,2</sup>; Q. Meng <sup>1,2</sup>; H. Shen <sup>1,2</sup>; S. Li <sup>1,2</sup>; J. Traynelis <sup>1,2</sup>; H. Li <sup>1,2</sup>; A.C. Salvador <sup>6</sup>; A. Hansen <sup>1,2</sup>; A. Sabo <sup>1,2</sup>; J. Posey <sup>2</sup>; J.R. Lupski <sup>1,2,4,5</sup>; M. Wangler <sup>2,3</sup>; D.R. Murdock <sup>1,2</sup>; R.A. Gibbs <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 3) Department of Pediatric Radiology, Texas Children's Hospital, Houston, TX, 77030; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX, 77030; 5) Texas Children's Hospital, Houston, TX, 77030; 6) Texas A&M University Health Science Center, College Station, TX, 77843

---

Xia-Gibbs syndrome (XGS, MIM: 615829) is an emerging cause of intellectual disability characterized by delayed speech, autistic features, sleep apnea, and hypotonia. More than 130 cases have been identified worldwide with a wide spectrum of predicted pathogenic mutations in the AT-Hook DNA Binding Motif Containing 1 (AHDC1) protein. We have established a secure patient registry, recording detailed clinical surveys for more than 45 families. The overwhelming majority (113 of 120) of cases are born with a dominant *de novo* truncating mutation in *AHDC1* at chr1p36.11, with age range at diagnosis from birth to 57 years old. RNA expression analysis using two permissive patient samples shows bi-allelic expression of the truncated and wildtype RNA transcripts providing supporting evidence for evasion of NMD. Interestingly, large deletions encompassing *AHDC1* have been reported as pathogenic with no cases providing sufficient evidence of its single involvement. *AHDC1* has three putative AT-hook binding motifs and homology to the catalytic subunit of DNA polymerase zeta (*REV3L*) and the Neurite Extension and Migration Factor (*NEXMIF*), both of which are implicated in developmental and intellectual disability disorders. We investigated a set of 7 clinically predicted pathogenic *de novo* missense or truncating mutations in *AHDC1* in a transfection assay using fluorescent tags. Independent expression and colocalization of wildtype and mutant variants were measured. We observed that the localization patterns of 3 transiently expressed truncated AHDC1 mutants (C791WFs\*57, R925\*, Q970\*) produce hallmarks of nucleolar stress. Specifically, mislocalized protein is shuttling from nucleoplasmic foci to the nucleoli. Such patterns are typically indicative of protein migration from heterochromatic sites surrounding the nucleoli to transcriptionally active euchromatic structures of rDNA transcription and suggest that the truncated proteins produce a disruption in this pattern. Shorter and longer truncation alleles (Q262\*, Q1270RFs\*75) display a spectrum in their localization pattern. This cellular phenotype is not observed with the tested missense mutations thus far (D607N, G792R.) In aggregate, these results are suggestive of a gain of function mechanism underlying cases of Xia-Gibbs syndrome in which the abnormal truncated AHDC1 protein impacts this cellular process and produces downstream developmental abnormalities, delayed developmental milestones and intellectual disability.

## **PgmNr 1186: Neurodevelopmental proteasomopathies caused by alterations of genes encoding subunits of the 19S regulatory particle.**

### **Authors:**

S. Küry<sup>1,2</sup>; F. Ebstein<sup>3</sup>; G.M. van Woerden<sup>4,5</sup>; T. Besnard<sup>1,2</sup>; C. Rosenfelt<sup>6,7,8</sup>; A. Hajdukowicz<sup>3</sup>; D. Baldrige<sup>9</sup>; J.A. Brault<sup>10</sup>; C. Mignot<sup>11,12</sup>; D. Ortiz<sup>13</sup>; S. Srivastava<sup>14</sup>; L. Ades<sup>15,16</sup>; A. Seeley<sup>17</sup>; J.A. Rosenfeld<sup>18</sup>; K. McDonald<sup>19</sup>; C. Forster<sup>20</sup>; A. Dauber<sup>21</sup>; R. Bernier<sup>22</sup>; T. Turner<sup>23</sup>; C. Costin<sup>24</sup>; J. Pappas<sup>25</sup>; D. Niyazov<sup>26</sup>; D. Beck<sup>27</sup>; D. Adams<sup>28</sup>; R. Redon<sup>2</sup>; P. Stankiewicz<sup>29</sup>; F. Bolduc<sup>6</sup>; Y. Elgersma<sup>4,5</sup>; S. Béziau<sup>1,2</sup>; E. Krüger<sup>3</sup>; PSMC3/PSMC5 consortium

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) CHU Nantes, Service de Génétique Médicale, Nantes, France;; 2) l'institut du thorax, INSERM, CNRS, UNIV Nantes, Nantes, France; 3) Universitätsmedizin Greifswald Institute of Medical Biochemistry and Molekular Biology, Klinikum DZ7, Greifswald, Germany; 4) Department of Neuroscience, Erasmus University Medical Center, 3015 CN, Rotterdam, The Netherlands; 5) ENCORE Expertise Center for Neurodevelopmental Disorders, Erasmus University Medical Center, 3015 CN, Rotterdam, The Netherlands; 6) Department of Medical Genetics, University of Alberta, Edmonton, AB T6G 2H7, Canada; 7) Division of Pediatric Neurology, University of Alberta, Edmonton, AB, Canada; 8) Neuroscience and Mental Health Institute, University of Alberta, Edmonton, AB, Canada; 9) Pediatrics, Washington University in St. Louis, St. Louis, MO, USA; 10) Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA; 11) APHP, Hôpital Pitié-Salpêtrière, Département de Génétique, Centre de Référence Déficience Intellectuelle de Causes Rares, GRC UPMC «Déficience Intellectuelle et Autisme», Paris, France; 12) Sorbonne Universités, Institut du Cerveau et de la Moelle épinière, ICM, Inserm U1127, CNRS UMR 7225, F-75013, Paris, France; 13) UPMC Children's Hospital of Pittsburgh, One Children's Hospital Drive, 4401 Penn Avenue, Pittsburgh, PA 15224, USA; 14) Department of Neurology, Boston Children's Hospital, Boston, Massachusetts, USA; 15) Department of Clinical Genetics, The Children's Hospital at Westmead, Locked Bag 4001, Westmead, NSW, 2145, Australia; 16) Discipline of Genetic Medicine, University of Sydney, Westmead, NSW, 2145, Australia; 17) Genomic Medicine Institute, Geisinger, Danville, PA, USA; 18) Baylor College of Medicine, Houston, TX, USA; 19) University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216, P 601.984.1597, F 601.984.1916, USA; 20) GeneDx, 207 Perry Parkway, Gaithersburg, MD 20877, USA; 21) Division of Endocrinology, Children's National Health System, Washington, DC, USA; 22) Department of Psychiatry & Behavioral Sciences, Center on Human Development and Disability, University of Washington, Box 357920 Seattle, WA 98195, USA; 23) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; Howard Hughes Medical Institute, Seattle, WA 98195, USA; 24) Akron Children's Hospital, 300 Locust Street, Suite 500 Akron, Ohio 44302, USA; 25) Clinical Genetic Services, Department of Pediatrics, NYU School of Medicine, New York, NY, USA; 26) Ochsner Clinic Foundation, University of Queensland, AU; 1315 Jefferson Highway, New Orleans, LA 70121, USA; 27) NIH/NHGRI, 10 Center Drive, Building 10, Room B-34129, Bethesda, MD 20892-1852, USA; 28) NIH Undiagnosed Disease Program, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, USA; 29) Baylor Genetics Laboratories, Houston, TX, USA

---

**Background:** Proteasome-mediated protein degradation is essential to proteostasis in eukaryotic cells. The scarcity of pathogenic variants in the ~50 genes which encode proteasome subunits, assembly helpers and regulatory particle subunits suggests adverse selection pressure associated with compromised cell viability. Whereas recessive variants in genes of the 20S catalytic core particle subunits have been found involved in autoinflammatory diseases, we recently showed that *de novo* variants in *PSMD12* –which encodes the scaffolding subunit Rpn5 of the 19S regulatory particle of 26S proteasome - lead to a syndromic neurodevelopmental disorder, the Stankiewicz-Isidor syndrome (MIM: 617516).

**Methods:** A collaborative international effort allowed us to identify a series of likely pathogenic variants in two additional genes of the regulatory particle 19S, *PSMC3* and *PSMC5*, which encode ATPases involved in unfolding and translocation of the substrate to be degraded into the 20S proteolytic core particle. In cells of affected individuals, we assessed the functional effect of the variants on proteasome 26S function and activity. We examined the possible impact of the variants on main pathways of neurodevelopment by immunoblotting. We used *in vivo* mouse and *Drosophila* models to determine the consequences of *PSMC3* and *PSMC5* variants on neuronal function and behavior.

**Results:** The variants found in *PSMC3* and *PSMC5* are missense, *de novo* and clustered in the ATPase domain. Individuals with *PSMC3* variants present with developmental delay and frequent MRI anomalies, along with cardiac and/or skeletal abnormalities, whereas those with *PSMC5* variants display intellectual disability and behavioral issues. We observed impaired proteasome function with a significant decrease in proteasomal activity, accumulation of ubiquitin conjugates, activation of the unfolded protein response and impaired mTOR activity.

**Conclusion:** We extend the list of neurodevelopmental disorders caused by variants of the 19S proteasome particle. Our preliminary results show that *PSMC3* and *PSMC5* alterations are associated with disruption of protein degradation systems and of the mTOR pathway. Proteomic analyses and development of animal model are in progress and are anticipated to expand our knowledge of the mechanisms of 19S-related disorders.

## PgmNr 1187: *De novo* variant in *FBXO11* gene in a child with neurodevelopmental disorder.

### Authors:

I. Zarante; M. Manotas; R. Quero; M. Guerra; J. Villalta; J. Rojas

[View Session](#) [Add to Schedule](#)

**Affiliation:** Instituto de Genetica Humana, Pontificia Universidad Javeriana, Bogota, Colombia

Neurodevelopmental disorders (NDDs) affect about 400,000 children per year, the novo mutations associated to NDDs have an average prevalence of 1 in 213 to 1 in 448 births according to the age of the parents. Recent studies of complete exome sequencing (WES) in children with undiagnosed developmental disorders is allowing the identification of rare *de novo* heterozygous mutations in genes associated with development as an important cause. A specific diagnosis in neurodevelopmental disorders is and will be useful for the management of the condition, prediction of possible problems, prognosis and genetic counseling.

We report a case of an 5-year-old male child with congenital hypotonia, congenital scoliosis, global developmental delay, sensorineural hearing impairment, dysmorphic features, metatarsus adductus and hypoplasia of the corpus callosum. The karyotype and comparative genomic hybridization which proved to be normal. Due to severe neurodevelopmental delay and facial dysmorphisms we completed the genetic investigations with whole-exome sequencing in trio, showed a probable pathogenic variant heterozygous in the *FBXO11* gene: c.2588C *de novo* and it has not been reported previously as a pathogenic variant nor as a benign variant; 20 cases of patients with neurodevelopmental disorder with variable intellectual disability and various other features associated with pathogenic variants in the *FBXO11* gene have been reported. Clinically, the patient's phenotype is congruent with global syndromic development delay due to allelic mutations in the *FBXO11* gene.

*FBXO11* encodes a member of the F-Box protein family, constituting a subunit of an E3-ubiquitin ligase complex. This complex is involved in ubiquitination and proteasomal degradation and thus in controlling critical biological processes by regulating protein turnover. It is possible that *FBXO11* mutation is the cause of neurodevelopmental disorder in other patients actually without etiology diagnosis.

## **PgmNr 1188: Mitochondrial-processing peptidase subunit beta (*PMPCB*) rare compound heterozygous variants in a patient with developmental delay, neurodegeneration, cerebellar vermis agenesis and cerebellar atrophy.**

### **Authors:**

K. Belhassan <sup>1</sup>; J. Neidich <sup>1</sup>; Y. Cao <sup>1</sup>; T. Toler <sup>2</sup>; J. Granadillo <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Cytogenomics and Molecular Pathology Laboratory, Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, Missouri; 2) Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, Saint Louis, Missouri

---

The *PMPCB* gene encodes the catalytic subunit of the mitochondrial processing protease (MPP). The protein cleaves an N-terminal signal of imported mitochondrial precursor proteins for their maturation. Five individuals from four families with childhood-onset neurodegeneration, absent speech and cerebellar atrophy were reported with pathogenic biallelic variants in *PMPCB*. Functional studies showed the deleterious effect with buildup of mitochondrial precursor proteins and impaired activity in mitochondrial respiratory chain complexes and in mitochondrial/cytosolic enzymes.[PMID: 29576218] This gene was found to be up-regulated and expressed in CD40L-stimulated Acute Lymphoblastic Leukemia (ALL) cells leading to their differentiation to dendritic cells.[PMID: 19580345] *PMPCB* was also up-regulated in colon cancer cells treated with saponin extracts. [PMID: 18425323] A study revealed blocking *PMPCB* lead to tumor suppression and apoptosis of EpCAM+ hepatic carcinoma cells via inactivation of Wnt/ $\beta$ -catenin signaling pathway.[PMID: 30862714]. Our patient is an 18-year-old male with global developmental delay, spastic ataxia and history of a seizure disorder. Brain MRI showed cerebellar vermis aplasia and cerebellar hypoplasia. An EEG revealed slow background for age. Chromosomal microarray was normal, mitochondrial genome analysis negative, and muscle biopsy unremarkable. Whole exome sequencing trio of proband and parents (WES) showed that the patient is a compound heterozygote for two variants, p.R451X and p.I160M, in the *PMPCB* gene (NM\_004279.2), each inherited from one parent. Both parents had learning difficulties as children. No other variants were reported via the WES analysis. Unlike the five cases described in literature, our patient has intelligible speech. His lactate/ pyruvate ratio was normal, suggesting some preserved mitochondrial respiratory function. This contrasts with the abnormal lactate levels and decreased muscular respiratory chain complex and aconitase activities described with the previously reported cases. One previous study showed that biallelic variants in *PMPCB* triggered defects in MPP proteolytic activity leading to mitochondrial dysfunction with early onset childhood neurodegeneration. Since few data are known about this gene beside its significance in neurodevelopment, this report helps to expand the phenotype of *PMPCB*-related disorders. Further studies are warranted to explore the functional impact of these variants.



## PgmNr 1189: Neurochondrin, a novel candidate for autosomal recessive intellectual disability with epilepsy, is required for normal maturation and function of cortical interneurons.

### Authors:

A. Fatima <sup>1</sup>; L. Laan <sup>1</sup>; J. Hoeber <sup>1</sup>; J. Klar <sup>1</sup>; T. Akram <sup>1,2</sup>; Z. Ali <sup>2</sup>; R. Fadoul <sup>1</sup>; C. Castillejo-López <sup>1</sup>; M.A. Ali <sup>3</sup>; S.M. Baig <sup>2</sup>; J. Schuster <sup>1</sup>; N. Dahl <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Box 815, SE-751 08 Uppsala, Sweden; 2) Human Molecular Genetics Laboratory; Health Biotechnology Division, National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, Pakistan; 3) Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, SE-751 08 Uppsala, Sweden

---

The *NCDN* gene encodes neurochondrin, predominantly expressed in neural tissues where it mediates neural growth, synaptic plasticity and signal transduction. Depletion of *Ncdn* in mice causes epileptic seizures, depressive-like behaviors and impaired spatial learning. We identified a rare homozygous *NCDN* missense variant (NM\_001014841.1:c.1246G>C) in three affected family members of a consanguineous Pakistani family characterized by mild intellectual disability (ID) and recurrent seizures. Electroencephalography (EEG) of two affected siblings revealed abnormal patterns. To investigate the cellular and biophysical mechanisms behind neurochondrin deficiency, we applied the CRISPR/Cas9 technology to knockout (KO) *NCDN* in human induced pluripotent stem cells (iPSCs) and neuroblastoma cell line (SH-SY5Y). Both targeted and isogenic wild-type (WT) iPSCs were differentiated into cortical interneurons. Electrophysiological recordings using whole cell patch-clamp analysis revealed that the neurochondrin-deficient neurons exhibited only abortive or single action potentials, with amplitudes significantly reduced, when compared to isogenic WT neurons ( $p=0.0067$ ). Furthermore, cortical interneuron cultures with *NCDN* deletions contained fewer mature neural cells. Moreover, differentiated SH-SY5Y cells with *NCDN* deletions exhibited significantly reduced neurite outgrowth when compared to WT SH-SY5Y cells. Overexpression of WT *NCDN* in neurochondrin-deficient SH-SY5Y cells rescued the neurite outgrowth phenotype. In contrast, overexpression of *NCDN* with the missense variant (*NCDN* c.1246G>C), identified in the three patients, failed to rescue the phenotype. Taken together we identified a *NCDN* missense variant associated with ID and epilepsy. Functional analysis of *NCDN* deficiency in iPSC derived cortical interneuron and neuroblastoma cell lines indicate that the protein is pivotal for neurite outgrowth, neuronal maturation and normal biophysical properties.



## **PgmNr 1190: Alpha-thalassemia X-linked intellectual disability syndrome: A multigenerational family provides additional insights into phenotypic variability within this syndrome.**

### **Authors:**

J. Schymick; N. Dykzeul; D. Matalon

[View Session](#) [Add to Schedule](#)

**Affiliation:** Stanford University, Division of Medical Genetics, Stanford, CA.

---

Alpha-thalassemia X-linked intellectual disability syndrome is a rare disorder with only 168 known cases (Gibbons R, 2006). It is characterized by severe global developmental delay, distinctive craniofacial features, hypotonia, intellectual disability, and a family history suggestive of an X-linked pattern of inheritance. Mild-to-moderate anemia secondary to alpha-thalassemia may also be present. This syndrome results from variants in the *ATR*X gene. Over 90% of pathogenic variants are reported are in regions encoding the zinc finger and helicase domains. Here we report a two generational Hispanic family with three affected individuals carrying a newly described variant in the *ATR*X gene. We describe the clinical characteristics of the affected individuals within our family and review the literature on genotype-phenotype correlations for variants in the *ATR*X gene.

Our proband is a 5-year old male with global developmental delay, microcephaly, short stature, intellectual disability and hypotonia. He has no discernible words and can walk only ten steps independently. Our proband's older brother is an 8-year old male with global developmental delay and hypotonia. He has 20 to 30 words and uses sign language. His first steps were between age 3 and 4 years, and he currently ambulates independently. He does not have microcephaly or short stature. Our proband's maternal uncle is a 39-year old male with a history of intellectual disability and developmental delay. He also has a lack of expressive speech and uses sign language. He ambulates independently. Dysmorphic features include microcephaly and full lips with tented upper vermilion border.

We identified a nonsense variant c.1753G>T (p.Glu585\*) in the *ATR*X gene via an X-linked intellectual disability gene panel. This variant was identified in our proband and subsequently in his older brother and maternal uncle who have relatively milder phenotypes. The glutamine residue at codon 585 is highly conserved and located outside the zinc finger and helicase domains providing additional insights into genotype-phenotype correlations for this gene. Furthermore, our family provides additional evidence for intrafamilial variability and offers additional case examples of milder phenotypes that may be underreported within this syndrome.

## PgmNr 1191: Long-term survival of mice with deletion of BRCA1 as adults.

### Authors:

J.O. Olayiwola <sup>1,2</sup>; J.B. Nguyen <sup>1</sup>; S.G. Rao <sup>1,2</sup>; J.G. Jackson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biochemistry and Molecular Biology, Tulane University School of Medicine, New Orleans, Louisiana.; 2) Graduate Program in Biomedical Sciences, Tulane University School of Medicine, New Orleans, Louisiana

---

BRCA1 faithfully repairs damaged DNA by promoting homology-directed repair (HDR). Loss of BRCA1 and other HDR genes is incompatible with viability and causes severe genomic instability. Homozygous loss of *Brca1* is embryonic lethal in mice, and cells lacking BRCA1 are sensitive to cellular stresses such as DNA damage and ionizing radiation (IR). We thus hypothesized that if *Brca1* is deleted in adult mice, genomic stress that is incompatible with life will ensue within days after deletion.

To test this, we administered doxycycline to 10-week old mice containing alleles of an inducible Cre system (Tet Responsive Element – Cre Recombinase [*TRE-Cre*], and the reverse tet transactivator [*rtTA*]), and floxed *Brca1* alleles (*Brca1*<sup>F22-24/F22-24</sup>). Administration of doxycycline resulted in ~1500 fold increase in expression of Cre recombinase in the liver and mammary gland. PCR of genomic DNA at the *Brca1* locus showed recombination and extensive excision of floxed exons 22 through 24 in liver, intestine, mammary, kidney, muscle, and pancreas. *Brca1* mRNA was reduced 85% to 99% after deletion, and more than half of mice had >95% deletion. Contrary to the embryonic lethality observed in all previously tested *Brca1* deletion mouse models, we found that deletion of *Brca1* in adult mice caused no overt phenotypes, was compatible with life, and that recombined mice (*Brca1*<sup>Δ/Δ</sup>) that have extensive, widespread deletion of *Brca1* survived up to 1 year after *Brca1* recombination. We next examined the survival of mice with *Brca1* loss that were exposed to additional cellular stress in the form of ionizing radiation. Surprisingly, mice that lack *Brca1* and are deficient for HDR showed no increased sensitivity to irradiation unless combined with p53 null alleles. Our surprising results show that mice deficient for HDR are viable and not radiation sensitive. Future studies will examine the compensatory mechanism of DNA repair that preserves viability in mice deficient in *Brca1* and HDR.

## PgmNr 1192: Phenylalanine (Phe) level and compliance for patients with phenylketonuria (PKU) - a Quality Improvement (QI) initiative.

### Authors:

C. Wu <sup>1,2</sup>; A. Kritzer <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA.; 2) Harvard Medical School, Boston, MA

### INTRODUCTION AND OBJECTIVES:

Phenylketonuria (PKU) is an inborn error of metabolism caused by the deficiency of phenylalanine hydroxylase enzyme. It is an autosomal recessive disease caused by pathogenic variants in *PAH* gene. PKU is one of the first biochemical genetic diseases identified (Følling, 1934), and is the first genetic disease treated successfully with diet modification (Bickel, 1953). After Dr. Guthrie developed a screening assay in 1963, newborn screen for PKU was soon implemented in all states in the United States and in many other countries around the world.

Here we investigated phenylalanine (Phe) levels and time elapsed since last Phe level, creating benchmark data for compliance and further quality improvement (QI) initiatives.

### METHODS:

This is a single-institution, retrospective longitudinal project that collects all Phe levels in Boston Children's Hospital (BCH) during an 18-year window (3/21/2001 – 1/7/2019). The Phe level was conducted either on-site at BCH's biochemical genetics laboratory, or by the Massachusetts state lab. We captured and calculated the following 2 metrics:

1. "Last [Phe]": The last Phe level on our record (unit:  $\mu\text{mol/L}$ ). This serves as an endpoint evaluation for how well the patients control their PKU.
2. "Last follow-up": The time period between the collection date of "Last [Phe]" and the last project observation date, 1/7/2019 (unit: days). This metric indicated how long the patient has not been checked for Phe, and represents the compliance.

### RESULTS:

During the 18-year interval, we have 16,007 Phe tests from 395 patients in BCH.

For the "Last [Phe]", there are 186 patients (47.09%) who have a level greater than 360  $\mu\text{mol/L}$ , 117 patients (29.62%) greater than 600  $\mu\text{mol/L}$ , and 39 patients (9.87%) greater than 1,200  $\mu\text{mol/L}$ .

For the "Last follow-up", there are 221 patients (55.95%) who have a "Last follow-up" greater than 182 days (6 months), 182 patients (46.08%) greater than 365 days (1 year), and 144 patients (36.46%) greater than 546 days (1.5 year).

### CONCLUSIONS and FUTURE DIRECTIONS:

Using 360  $\mu\text{mol/L}$  as an optimal Phe level cut-off, 47.09% of the patients were out of range. Using 12 months as a follow-up period cut-off, 46.08% of the patients are not compliant. Almost half of the patients cannot keep up with our treatment goal. We will use these 2 metrics to identify patients and

study the reasons for poor compliance to improve the quality of PKU care.

## PgmNr 1193: Intermittent neurologic decompensations: An underrecognized presentation of tyrosine hydroxylase (TH) deficiency.

### Authors:

M. Champagne <sup>1</sup>; G.A. Horvath <sup>2</sup>; S. Perreault <sup>3</sup>; J. Gauthier <sup>1,4</sup>; G.A. Mitchell <sup>1,4</sup>; J.F. Soucy <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics and Molecular Diagnostic Laboratory, Centre hospitalier universitaire Sainte-Justine and Department of Pediatrics, Université de Montréal, Montréal, Québec, Canada; 2) Division of Biochemical Diseases, Department of Pediatrics, BC Children's Hospital, Vancouver, British Columbia, Canada; 3) Division of Child Neurology, Department of Pediatrics, Centre hospitalier universitaire Sainte-Justine, Montréal, Québec, Canada; 4) Integrated Centre for Pediatric Clinical Genomics, Centre hospitalier universitaire Sainte-Justine, Montréal, Québec, Canada

---

**Introduction.** Tyrosine hydroxylase (TH) deficiency is a treatable autosomal recessive inborn error of dopamine biosynthesis caused by mutations in *TH* and resulting in neurological signs caused by dopamine deficiency. Two main presentations are described. Type A, milder, presents after 12 mo with progressive hypokinesia and rigidity. Type B presents before 12 mo and often neonatally, as a progressive complex encephalopathy.

**Case history.** We report a girl with mild TH deficiency who had recurrent episodes of neurological decompensation following infections or vaccinations. Prior to the first episode, she had normal development except for mild head tremor. The first episode occurred at 12 mo, after 2 weeks of intermittent fever following a vaccination, with marked fatigue, loss of walking/sitting, worsened tremor, severe axial hypotonia, and recuperation over several days. Episodes 2 (19 mo) and 3 (25 mo) followed viral infections by a few days, presenting as lethargy, axial hypotonia, appendicular rigidity and worsening tremor needing several weeks of intensive rehabilitation. She had residual tremor and mild lower limb spasticity. Basal ganglia were normal on cerebral MRI. Exome sequencing revealed two missense variants of unknown significance in *TH*: c.1147G>T (p.G383W, paternal) and c.1084G>A (p.E362K, maternal). Both have gnomAD allele frequencies of <1/50,000 and are classified as deleterious by SIFT, PolyPhen and MutationTaster. CSF analysis showed dopamine metabolite, homovanillic acid (HVA), 160 nmol/L (reference, 233-938) and HVA/5-hydroxyindolacetic acid molar ratio 1.07 (literature reference, 1.5-3.5). The patient responded rapidly to Dopa/carbidopa (currently 2.7 mg Dopa/kg/d) without further episodes.

**Patient series.** Considering this patient plus 4 others discovered by systematic literature review, we identified 7 distinct neurologic episodes. All occurred after infections or vaccinations, between ages 12 mo to 6 yr. Each presented with marked hypotonia and decreased motor function. Tremor occurred in 4/5 patients. Recuperation occurred in all cases. 3/5 patients later developed dystonia.

**Conclusion.** This defines a third clinical category of TH deficiency: episodes of marked hypotonia with recuperation over weeks to months. Because TH deficiency is treatable and important for genetic counselling, this diagnosis should be considered in episodes of marked hypotonia, even after a single episode, especially in patients with dystonia or rigidity.

## PgmNr 1194: Metabolomics analysis of SLC13A5 deficiency in mice.

### Authors:

A.C. Santos de Medeiros <sup>1</sup>; K.E. Ginton <sup>2</sup>; S.H. Elsea <sup>2</sup>; B.H. Graham <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MMGE, IUSM, Indianapolis, Indiana.; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas

---

Early infantile epileptic encephalopathy (EIEE) is a genetically heterogeneous disorder in humans. EIEE is characterized by severe neonate seizures along with developmental delays with intellectual disabilities. One cause of EIEE is mutations in *SLC13A5* (EIEE Type 25), a gene encoding member 5 of solute carrier family 13, a sodium/citrate cotransporter. Reduction of *SLC13A5* expression alters metabolism with weight loss in several species. In humans, *SLC13A5* is expressed predominantly in the liver, brain, salivary gland and testis. Citrate plays essential role in energy homeostasis and is detectable in plasma, cerebral spinal fluid (CSF) and seminal fluid. Citrate is produced in the mitochondrion as a product of oxaloacetate and acetyl-CoA condensation by citrate synthase in the tricarboxylic acid (TCA) cycle. Besides its role in mitochondrial energy metabolism, cytosolic citrate is a substrate for fatty acid and sterol biogenesis, and extracellular citrate plays a role in determining the availability of divalent cations in CSF and seminal fluids through chelation. A recent study suggests that loss of *SLC13A5* disrupts oncogenic signaling in human liver cells. A previous study of a small cohort of patients with *SLC13A5* deficiency demonstrated altered TCA metabolites by mass spectrophotometry in CSF, plasma and urine. These results suggest disruption of TCA cycle/energy metabolism due to *SLC13A5* loss. In the present study, we carried out a pilot experiment examining the metabolites in the brain, liver and plasma of a pair of mice (one male and one female) with *Slc13a5* deficiency. Overall, initial data demonstrates a reduction in amino acid metabolites and elevations of free fatty acids in brain and liver. Interestingly, it also suggests potential sexual dimorphism with significant differences between the male and female samples for several diacylglycerol (DGA) species. Metabolomic evaluation of a larger cohort of animals (12 per genotype, including 6 males and 6 females) is currently underway.

## PgmNr 1195: Multi-omics analysis of a new cellular model of Barth syndrome uncovers mechanisms of mitochondrial dysfunction and highlights the importance of tissue-specific modeling.

### Authors:

A. Anzmann<sup>1</sup>; O.L. Sniezek<sup>1</sup>; S.M. Claypool<sup>2</sup>; H.J. Vernon<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, MD

---

Barth syndrome (BTHS) is an X-linked disorder caused by defects in tafazzin (TAZ), a transacylase involved in the final remodeling step of cardiolipin (CL), which results in increased monolysocardiolipin (MLCL) and decreased remodeled CL on the inner mitochondrial membrane. TAZ is ubiquitously expressed in all tissues, however the CL acyl chain composition is tissue specific; brain CL is characterized by a diversified array of acyl chains including polyunsaturated chains, whereas in cardiac and skeletal muscle CL is predominantly in the tetralinoleoyl form. Clinical features of BTHS include cardiomyopathy, skeletal myopathy, and neutropenia. In order to explore novel areas of cellular dysfunction and identify potential targets for therapeutic intervention we applied a multi-omics discovery approach to a TAZ-deficient HEK293 cell line that was developed and phenotypically validated in our laboratory, TAZ<sup>745</sup>. Lipidomics analysis of TAZ<sup>745</sup> cells showed the characteristic CL abnormalities seen in BTHS including a decrease in remodeled CL, a shift towards unsaturated CL, and an increase in MLCL. Proteomics analysis comparing wild type (WT) and TAZ<sup>745</sup> cells revealed differential expression of proteins involved in the dynamic response to mitochondrial stress (i.e. machinery of fission, fusion, and mitophagy), and of proteins of the mitochondrial respiratory chain, including decreased expression of several subunits of complex I and complex I assembly factors. Metabolomics analysis revealed decreased NAD<sup>+</sup> and increased AMP in TAZ<sup>745</sup> compared to WT, further highlighting respiratory chain dysfunction as a key feature of TAZ deficiency. The most significantly reduced complex I assembly factor in TAZ<sup>745</sup>, NDUFAB1, was also present at lower abundance in BTHS patient-derived lymphoblast cell lines (LCLs) compared to controls. These changes in steady state levels resulted in reduced complex I activity in TAZ<sup>745</sup> and patient derived LCLs compared to their respective controls. Due to the tissue specific-nature of the clinical symptoms in BTHS (i.e. severe cardiac and skeletal muscle dysfunction, with minimal neurological effects), we developed an iPSC model of TAZ deficiency, and differentiated these iPSCs into neural progenitor cells and cardiomyocytes in order to determine if complex I dysfunction is more prominent in tissues with severe clinical effects. These findings have significance for targeted therapeutic development in BTHS aimed at improving complex I function.

## PgmNr 1196: Unraveling AADC deficiency: Clinical, biochemical, and molecular genetics findings in Latin-American patients.

### Authors:

B. Dias <sup>1</sup>; C. Gil <sup>1</sup>; J.H.R. Fonseca <sup>2</sup>; L. Vagnini <sup>3</sup>; J.F. Franco <sup>4,5</sup>; Z.A. Carneiro <sup>1</sup>; C. Lourenco <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centro Universitario Estacio de Ribeirao Preto, Ribeirao Preto, Sao Paulo, Brazil; 2) Bioquimica Genetica - Laboratorio DLE, Rio de Janeiro, Brazil; 3) Centro Paulista de Diagnostico e Pesquisa - CPDP, Ribeirao Preto, Brazil; 4) IPEN/USP, Instituto de Pesquisa Energia Nuclear - Universidade de Sao Paulo, Sao Paulo, Brazil; 5) Pontificia Universidade Catolica de Campinas, Campinas, Brazil

---

**BACKGROUND:**Aromatic L-amino acid decarboxylase (AADC) is central in the synthesis of biogenic monoamine neurotransmitters. AADC deficiency is a severe neurometabolic disorder, usually underrecognized, presenting early in life with hypotonia, hypokinesia, oculogyric crises, autonomic dysfunction, dysphoric mood, and sleep disturbance **METHODS:**Clinical, biochemical and molecular genetics data retrospective analysis from 5 patients with AADC deficiency **RESULTS:**Neurological symptoms became evident in all patients during the first 6 months of life. Severe hypotonia, hypokinesia and hypomimia, dystonic movements and typical oculogyric crises were present in all patients. Ptosis occurred in 4 patients, hyperhidrosis in 3, hypersalivation in 4 and nasal congestion in 4 patients. Two patients suffered from temperature instability in the first year of life and two from hypoglycaemia **DISCUSSION:**AADC deficiency is a neurotransmitter disorder with features presenting as early as in the first month of life. Hypotonia, hypokinesia, oculogyric crises and autonomic are clinical hallmarks of the disease and may lead to clinical suspicion. Clinical phenotype severity is variable, but the majority of patients show minimal motor development in the absence of treatment. Since gene therapy has been shown to bring greater improvement for such patients, early diagnosis in this disease is essential for better clinical outcome



## PgmNr 1197: Characterization of multiple spontaneous novel variants in the human *NPC1* gene.

### Authors:

S. Newman<sup>1,2,3</sup>; R. Dewar<sup>4,5</sup>; C.A. Rupar<sup>1,2,3,4,5</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Pathology and Lab Medicine, Western University, London, Ontario, Canada; 2) Children's Health Research Institute, London Health Sciences Center; Victoria Hospital, London, Ontario, Canada; 3) Lawson Research Institute, London, Ontario, Canada; 4) Biochemical Genetics Laboratory, Pathology and Lab Medicine, London Health Sciences Center; Victoria Hospital, London, Ontario, Canada; 5) Department of Pediatrics, London Health Sciences Center; Victoria Hospital, London, Ontario.

---

Niemann Pick disease type C (NPC) is an autosomal recessive lysosomal storage disease resulting from abnormal cholesterol and sphingolipid trafficking within the endosomal/lysosomal pathway. The disease is due to a variant in the *NPC1* or *NPC2* genes, with both variants presenting identical clinical pathologies. We characterized 29 patients with variants in the *NPC1* gene. Following DNA isolation, exons with flanking intronic regions were sequenced. Our results determined 8 novel unpublished variants;

6 were due to point mutations resulting in codon replacements, 2 were small deletions/insertions leading to frameshift mutations. The novel variants were investigated by bioinformatics using online databases 1000 genomes; SIFT polyphen, and mutation taster. Variants were analyzed computationally using SoftGenetics: Nextgene, Geneticist Assistant, Mutation Surveyor; and Alamut Visual with HGMD to predict potential detrimental effects of each variant. The purpose of this report is to consider these novel variants for relevance for diagnostic purposes, or disease progression in NPC.

## PgmNr 1198: Mouse models of, and AAV gene therapy for, propionic acidemia (PA).

### Authors:

C.P. Venditti; S.N. Smith; R.J. Chandler

[View Session](#) [Add to Schedule](#)

**Affiliation:** NHGRI, NIH, Bethesda, MD

Propionic Acidemia (PA) is a rare autosomal recessive metabolic disorder caused by reduced activity of the enzyme propionyl-CoA carboxylase (PCC). The PCC enzyme is composed of two subunits, alpha and beta, encoded by the *PCCA* and *PCCB* genes, and is responsible for the catabolism of propiogenic amino acids, as well as odd-chain fatty acids and cholesterol. PA is equally likely to be caused by mutations in the *PCCA* or *PCCB* gene. Patients diagnosed with PA typically present in the early newborn period with a metabolic crisis, which can be fatal if not promptly recognized and treated. Hypotonia, lethargy, growth retardation/failure to thrive, cardiomyopathy and seizures are variably present but overall are common complications in the patients. PA is chronically managed via dietary restriction, and antibiotic therapy is used in some patients to decrease propionate derived from gut bacteria. Medical management can decrease the severity of symptoms, but the high rates of morbidity and mortality still persist in many with PA, driving care teams to offer elective liver transplantation as a surgical treatment option. Because of suboptimal outcomes observed in PA, there is a pressing need to develop new therapies.

We have generated new mouse models of *Pcca* and *Pccb* deficiency. Two loss of function alleles in exon 5 of *Pcca*, as well as one knock-in (A134T), and 4 in exon 14 of *Pccb* were recovered after CRISPR Cas-9 gene editing, and bred to homozygosity. Other than the *Pcca*<sup>A134T/A134T</sup> mice, homozygotes for all *Pcca* and *Pccb* mutations displayed neonatal lethality and massively increased 2-methylcitrate, very much like severely affected PA patients. Next, a series of AAV cassettes designed to express the human orthologues of PCCA or PCCB in the liver of the mutant mice were prepared, and pseudoserotyped with an AAV8 or AAV9 capsid. *Pccb*<sup>-/-</sup> or *Pcca*<sup>-/-</sup> mice were treated at birth, and followed for survival, growth and metabolic correction. Both lines of mice could be rescued from lethality with systemic AAV gene therapy, and the rescued mutants had improved metabolic parameters and prolonged survival. In summary, we have created new murine models of severe PA caused by *Pcca* or *Pccb* deficiency, developed a suite of candidate AAV vectors that express the human orthologues, and tested them for efficacy in the mutant mice. Our results provide critical preclinical steps in the pathway to develop AAV gene therapy for patients with propionic acidemia.

## PgmNr 1199: Genotype and phenotype analysis of 12 Chinese children with treatable valine-metabolism-related Leigh syndrome.

### Authors:

J. Wang; F. Fang; X. Han; Z. Liu; M. Xu; D. Shen; L. Dai; W. Zhang; J. Li

[View Session](#) [Add to Schedule](#)

**Affiliation:** Neurology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

**Background** 3-Hydroxyisobutyryl-CoA hydrolase (HIBCH) and short-chain enoyl-CoA hydratase (SCEH) are involved in valine metabolism, encoded by *HIBCH* and *ECHS1*, respectively, and any mutation can cause Leigh syndrome. Mitochondrial function can be improved clinically by limiting valine intake.

**Methods** The clinical data and genetic mutations of patients from 2013 to 2019 were retrospectively analyzed to discuss the genotype and phenotypic characteristics.

**Results** There were 160 genetically diagnosed cases with Leigh syndrome. The *HIBCH* and *ECHS1* mutations were 6 cases respectively, accounting for 7.5% (12/160). The onset age was 0 to 4 years and the median age of onset was 0.33 years. 4 cases were triggered by infection, and 1 was triggered by vaccination, all of which were *HIBCH* mutations. The initial symptoms were mainly developmental regression, acute encephalopathy (mostly in *HIBCH* defect) and dystonia (mostly in *ECHS1* defect). The main clinical symptoms were developmental retrogression (n=9), dystonia (n=8), nystagmus (n=5). Basal ganglia lesions were found in all 12 patients, and 7 were complicated with brainstem lesions. Organic acid screening revealed increased 2,3-dihydroxy-2-methylbutyrate (23DH2MB) in samples from the 10 patients (83%), and the *ECHS1* mutation and *HIBCH* mutation were 5 cases respectively. Blood tandem mass spectrometry analysis showed that 5 patients (42%) of 3-hydroxybutyryl carnitine (C4OH) increased, 3 cases were *HIBCH* mutations. Seventeen mutations were identified in *HIBCH* and *ECHS1*, consisting of 13 novel mutations and 4 mutations reported previously. thirteen novel mutations included 10 missense mutations (*HIBCH*: c.977A>C, p.L326R; c.452C>T, p.S151L; c.1036G>T, p.V346F; *ECHS1*: c.607C>T, p.A203T; c.463G>A, p.G155S; c.557C>T, p.S186L; c.310C>G, p.Q104E; c.554T>C, p.K185R; c.542C>T, p.R181H; c.712G>C, p.A238P)?2 splicing mutations (*HIBCH*: c.750+1G>A; *ECHS1*: c.414+5G>A); 1 nonsense mutation (*HIBCH*: c.469C>T, p.R157X).

**Conclusion** Leigh syndrome caused by *HIBCH* and *ECHS1* mutations was rare and premature. The head MRI showed Leigh-like changes, involving basal ganglia, with or without brainstem involvement. The increase of 23DH2MB in urine may be a biomarker of *HIBCH* and *ECHS1* defects, while the increase of C4OH in blood is more common in *HIBCH* defects. We identified 5 novel *HIBCH* and 8 novel *ECHS1* mutations, enriching the genotype of Leigh syndrome in Chinese children.

## PgmNr 1200: Collaborative development of therapeutics for sialic acid storage disease.

### Authors:

M.E. Hackbarth<sup>1</sup>; M. Huizing<sup>1</sup>; M.C.V. Malicdan<sup>1</sup>; D.R. Adams<sup>1</sup>; K. Dobrenis<sup>2</sup>; M. Wasserstein<sup>3</sup>; W.A. Gahl<sup>1</sup>; S.U. Walkley<sup>2</sup>; SASD Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD, USA; 2) Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY, USA; 3) Children's Hospital at Montefiore, Albert Einstein College of Medicine, Bronx, NY, USA

---

Lysosomal sialic acid (SA) storage disease (SASD) is an autosomal recessive, neurodegenerative, multisystemic disorder caused by defects in the lysosomal SA membrane carrier *SLC17A5* (Sialin). *SLC17A5* defects cause free SA and secondary metabolites to accumulate in lysosomes. The clinical spectrum ranges from severe infantile onset in infantile sialic acid storage disease (ISSD; elevated urine SA; ~40 reported cases), to a mild, adult form, called Salla disease (moderate elevation of urine SA; ~120 cases).

Although sialic acid metabolism, membrane transport, and lysosomal biology have been extensively studied, the pathobiology of SASD remains poorly understood. Moreover, SASD is likely underdiagnosed; known patients have experienced a diagnostic delay due to the rarity of the disorder, non-specific clinical symptoms and absence of routine urine SA testing. There is no approved therapy for SASD.

As is typical for orphan diseases, the small population of patients makes it difficult to motivate industries to invest in performing the pre-clinical and clinical studies necessary to develop therapies. On the other hand, multidisciplinary collaborative efforts involving the NIH, academic clinical scientists, and patient advocacy groups have successfully overcome the scientific, clinical and financial challenges facing the development of new drug treatments for rare diseases.

Encouraged by these successes, we have initiated a collaborative effort for SASD. This has allowed us to start creating cell and mouse models, perform basic/translational research, initiate a natural history study to aid in the identification of biomarkers and treatment endpoints, raise awareness for SASD, and investigate leads on drug candidates. We aim to collect data that incentivize industry to further develop, obtain approval, and commercialize SASD treatments.

## PgmNr 1201: The laboratory approach to the diagnosis of Krabbe disease.

### Authors:

D. Matern <sup>1</sup>; C. Turgeon <sup>1</sup>; K. Nickander <sup>1</sup>; A. White <sup>1</sup>; V.K. Prasad <sup>2</sup>; J. Kurtzberg <sup>2</sup>; J.J. Orsini <sup>3</sup>; M. Gelb <sup>4</sup>; M. Escolar <sup>5</sup>; D. Gavrilov <sup>1</sup>; D. Oglesbee <sup>1</sup>; K. Raymond <sup>1</sup>; S. Tortorelli <sup>1</sup>; P. Rinaldo <sup>1</sup>; A. Guenzel <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biochemical Genetics Laboratory, Mayo Clinic, Rochester, MN.; 2) Department of Pediatrics, Duke University Medical Center, Durham, NC; 3) New York State Department of Health, Albany, NY; 4) University of Washington, Seattle, WA; 5) Children's Hospital of Pittsburgh, Pittsburgh, PA

Newborn screening (NBS) for Krabbe Disease (KD) is currently performed in six US states. High frequency of pseudodeficiency alleles in the galactocerebrosidase (GALC) gene has led to a large number of false positive cases. Psychosine (PSY), a substrate of GALC, has been found to be elevated in the blood in KD patients. To facilitate close monitoring and rapid identification of patients at risk of KD based on an abnormal NBS result or suspicious clinical presentation, we developed assays to measure PSY in dried blood spots (DBS), erythrocytes (RBC) and cerebrospinal fluid (CSF), and adapted a new GALC enzyme assay in leukocytes (WBC).

**Specimens:** Specimens were obtained from at risk patients based on the primary NBS test result or because of a phenotype suggestive of KD.

### Results:

	PSY (DBS) [nmol/L]	PSY (RBC) [pmol/g Hb]	PSY (CSF) [nmol/L]	GALC (WBC) [nmol/h/mg prot]
IKD	25-111 (n=23)	54-356 (n=3)	1.3-4.3 (n=6)	<0.01 (n=12)
LOKD	2.6-9.9 (n=8)	37-43 (n=3)	0.02-0.30 (n=6)	0.03-0.07 (n=5)
GALC Pseudo	0.35-1.3 (n=55)	-	-	0.1-0.3 (n=5)
GALC Carrier	0.36-1.5 (n=27)	-	-	0.09-0.55 (n=14)
Controls	<1.5 (n=205)	<4.5 (n=152)	<0.03 (n=124)	>0.87 (n=130)

IKD, Infantile KD; LOKD, Late Onset KD; ranges represent 1st to 99th percentiles.

**Conclusions:** The new GALC enzyme activity assay along with PSY measurements in various specimen types may allow a differentiated diagnosis of KD which is specifically important when asymptomatic cases are identified by NBS. Testing of PSY in RBC was implemented to allow normalization of PSY values to hemoglobin, thereby avoiding potential collection artifacts common to DBS. The PSY assay in RBC may therefore allow for monitoring of asymptomatic patients with LOKD in preparation for potential hematopoietic stem cell transplantation.

## PgmNr 1202: Extensive Mongolian spots in GM1 gangliosidosis: A pictorial report in Indian population.

### Authors:

S. Mishra; K.M. Girisha

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medical Genetics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, Karnataka, India

---

GM1 gangliosidosis (MIM# 230500) is a lysosomal storage disorder, usually characterized by progressive psychomotor deterioration, visceromegaly, facial coarseness, macular cherry-red spots, and skeletal abnormalities. Cutaneous lesions are not frequently reported in the literature. In this study, we have described five unrelated patients of GM1 gangliosidosis with persistent and extensive Mongolian spots on the trunk and extremities that provided a clue towards their clinical diagnosis. The cutaneous lesions were present at birth, before the appearance of characteristic features of the disease. Four out of five patients were from our previously reported cohort of 46 patients with GM1 gangliosidosis in which 30 patients (65%) had Mongolian spots. Of the five reported patients, four patients presented with psychomotor retardation and one had only extensive Mongolian spots as the presenting complaint. Apart from extensive Mongolian spots, they exhibited coarse facies, hepatosplenomegaly, generalized hypotonia, and features of dysostosis multiplex. Three patients had cherry-red spots on the retina. The condition was confirmed by identification of biallelic pathogenic variants in the *GLB1* gene and very low activity of beta-galactosidase enzyme in peripheral leukocytes. Biallelic pathogenic variants in *GLB1* were identified in all of them. These include one novel missense variant p.(Lys493Asn) and two reported variants, a splicing variant c.75+2dup and a nonsense variant p.(Arg457Ter), in *GLB1*.

We ascertain that extensive Mongolian spots are the unique cutaneous phenotype of GM1 gangliosidosis. Unusual extensive Mongolian spots can be used as a diagnostic handle for early recognition of GM1 gangliosidosis. Extensive Mongolian spots, particularly in the context of neurodevelopmental deterioration and macular cherry-red spots, helped in differentiating GM1 gangliosidosis from other similar neurodegenerative conditions.

## PgmNr 1203: Acid ceramidase deficiency due to mutations in *ASAH1*, presenting as Farber disease: Diagnostic data from the first-ever natural history study.

### Authors:

A. Solyom <sup>1</sup>; C. Grant <sup>2</sup>; C. Ferreira <sup>2</sup>; N. Mungan <sup>3</sup>; F. Bulut <sup>3</sup>; L. Selim <sup>4</sup>; I. Gamal el Din <sup>4</sup>; R. Puri <sup>5</sup>; S. Bijarnia-Mahay <sup>5</sup>; P. Harmatz <sup>6</sup>; C. Lampe <sup>7,8</sup>; A. Hahn <sup>8</sup>; N. Guelbert <sup>9</sup>; E. Sundberg <sup>10</sup>; B. Magnusson <sup>10</sup>; N. Arslan <sup>11</sup>; B. Makay <sup>11</sup>; S. Kapoor <sup>12</sup>; M. DiRocco <sup>13</sup>; S. Ozen <sup>14</sup>; E. Batu <sup>14</sup>; G. Gokcay <sup>15</sup>; M. Torcoletti <sup>16</sup>; A. Kimura <sup>17</sup>; J. Mitchell <sup>18</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Enzyvant, Basel, Switzerland; 2) Children's National Medical Center, Washington, United States; 3) Cukurova University Hospital, Adana, Turkey; 4) Cairo University Children's Hospital, Cairo, Egypt; 5) Sir Ganga Ram Hospital, New Delhi, India; 6) UCSF Benioff Children's Hospital Oakland, Oakland, United States; 7) HSK Wiesbaden, Wiesbaden, Germany; 8) University Hospital Giessen, Giessen, Germany; 9) Children's Hospital of Cordoba, Cordoba, Argentina; 10) Karolinska University Hospital, Stockholm, Sweden; 11) Dokuz Eylul University Hospital, Izmir, Turkey; 12) Lok Nayak Hospital and Maulana Azad Medical College, New Delhi, India; 13) Istituto Giannina Gaslini, Genoa, Italy; 14) Hacettepe University, Ankara, Turkey; 15) Istanbul University, Istanbul, Turkey; 16) University of Milan, Milan, Italy; 17) Enzyvant, Cambridge, USA; 18) Montreal Children's Hospital, Montreal, Canada

**Introduction:** Farber disease is a rare lysosomal storage disorder caused by mutations in both alleles of the *ASAH1* gene. The resulting deficiency of the lysosomal enzyme acid ceramidase, and accumulation of the pro-inflammatory sphingolipid ceramide, causes a broad spectrum of symptoms and disease severity which may delay diagnosis or lead to misdiagnosis. The ongoing study described here is the first comprehensive, systematic clinical study of the natural history of Farber disease.

**Methods:** The Observational and Cross-Sectional Cohort Study of the Natural History and Phenotypic Spectrum of Farber Disease (NCT03233841) is designed to collect retrospective and prospective data including demographics, clinical presentation, phenotype, and diagnostic history of patients diagnosed with Farber disease who have or have not undergone hematopoietic stem cell transplantation (HSCT), along with specific prospective clinical evaluations in living patients.

**Results:** 44 patients, including 24 living non-HSCT, 3 living HSCT, 13 deceased non-HSCT, 4 deceased HSCT, have been enrolled in the study since November 2017. Diagnostic modalities included genetic testing (95%), enzyme activity testing (14%), and nodule biopsy (36%). The average age of the living non-HSCT patients at enrollment was 6.7 years (range 1 to 28 years). Average time from onset of first symptoms to diagnosis was 2 years (range < 1 to 12 years). Patient countries of origin included Afghanistan, Argentina, Canada, Egypt, Germany, India, Iraq, Italy, Mexico, Sweden, Syria, Turkey, and the USA.

**Conclusions:** Patients representing the breadth of the phenotypic spectrum of Farber disease, from rapidly progressive (severe) to slowly progressive (attenuated), have been enrolled from 15 centers in 9 countries. The majority of patients were diagnosed by *ASAH1* mutational analysis. Demographic data and numbers of patients enrolled indicated that Farber disease is likely not as rare as previously thought.

## PgmNr 1204: The prevalence of *GALM* mutations that cause galactosemia: A database of functionally evaluated variants.

### Authors:

A. Kikuchi <sup>1</sup>; S. Iwasawa <sup>1</sup>; Y. Wada <sup>1</sup>; N. Arai-Ichinoi <sup>1</sup>; O. Sakamoto <sup>1</sup>; G. Tamiya <sup>2</sup>; S. Kure <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Tohoku University Hospital, Sendai, Miyagi, Japan; 2) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan

---

Galactosemia is a metabolic disorder that affects the appropriate metabolism of  $\beta$ -D-galactose. Deficiencies in three of the enzymes of the Leloir pathway, namely, GALT, GALK1, or GALE, are characterized as type I, II, and III galactosemia, respectively. Recently, we reported a novel type of galactosemia (type IV galactosemia) due to biallelic *GALM* mutations. Genetic diagnosis is indispensable for diagnosing *GALM* deficiency because no biochemical diagnosis method has been established. Given that apparently pathogenic variants in *GALM* are found in public variant databases, we presumed the presence of pathogenic variants that have not been reported. In this study, we explore 67 *GALM* variants that are prevalent in the ExAC database, including 57 missense variants, 7 stop-gain variants, 2 frameshift variants, and 1 splice-site variant. We performed an in vitro expression assay and an enzyme activity assay. Among the 66 variants except for 1 splice-site variant, 29 produced no or faint protein expression and were judged as pathogenic variants. Furthermore, the remaining 37 variants were evaluated by enzyme activity assay. Two showed mildly reduced enzyme activity and were classified as benign. Based on our study, the estimated incidence of *GALM* deficiency is 1:228,411 in all populations, 1:10,388 in the African population, and 1:80,747 in the Japanese population. Our *GALM* mutation database is useful for the genetic diagnosis of *GALM* deficiency.



## PgmNr 1205: 1-<sup>13</sup>C-Propionate oxidative capacity predicts clinical outcomes in propionic acidemia.

### Authors:

O.A. Shchelochkov<sup>1</sup>; A. Pass<sup>1</sup>; J. Myles<sup>2</sup>; S. Ferry<sup>1</sup>; C. Van Ryzin<sup>1</sup>; M. Schoenfeld<sup>2</sup>; L. Fletcher<sup>3</sup>; T. Cassimatis<sup>3</sup>; K.Y. Chen<sup>3</sup>; C. Chlebowska<sup>4</sup>; J. Sloan<sup>1</sup>; A. Thurm<sup>4</sup>; I. Manoli<sup>1</sup>; C.P. Venditti<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MGMGB, National Human Genome Research Institute, NIH, Bethesda, MD, USA.; 2) Nutrition Department, Clinical Center, NIH, Bethesda, MD, USA; 3) NIDDK, NIH, Bethesda, MD, USA; 4) NIMH, NIH, Bethesda, MD, USA

---

Propionic acidemia (PA) is a severe metabolic disorder characterized by significant multiorgan involvement. There are no FDA-approved genomic therapies or validated disease biomarkers. We explored the role of *in vivo* whole-body oxidation of 1-<sup>13</sup>C-propionate as a surrogate endpoint for clinical outcomes in propionic acidemia

All study participants were enrolled in a natural history protocol (The Natural History, Physiology, Microbiome and Biochemistry Studies of Propionic Acidemia, ClinicalTrials.gov ID NCT02890342). To identify parameters associated with 1-<sup>13</sup>C-propionate oxidation, we selected variables meeting a univariate criterion of coefficient  $r < -0.4$  or  $> 0.4$  (99 of 343 parameters). Of these, a subset of 9 variables representative of key disease characteristics (height z-score, full scale IQ (FSIQ), optic nerve atrophy, sensorineural hearing loss, left ventricular ejection fraction (LVEF%), cystatin C-based estimated glomerular filtration rate (eGFR), serum erythropoietin, complete and incomplete protein as a percent of recommended daily allowance) were used to specify a clinical prediction model of propionic acidemia in non-transplanted patients.

We enrolled 34 participants, ages 3-53 years (mean age 16 years, 56% females, including 1 kidney- and 3 liver-transplanted individuals); 30 completed the stable isotope study. Participants with biallelic null variants (frame shift, premature stop codons, and exon deletions) in *PCCA* or *PCCB* had lower 1-<sup>13</sup>C-propionate oxidation (mean = 6.2%) compared to those with at least one non-null variant (mean = 20.2%,  $P < 0.0001$ ). Oxidation was normal in liver transplant recipients (mean = 64.7%,  $P < 0.0001$ ). In non-transplanted participants, 120-minute 1-<sup>13</sup>C-propionate oxidation was associated with height z-scores ( $r = 0.48$ ,  $P < 0.05$ ), FSIQ ( $r = 0.74$ ,  $P < 0.0001$ ), sensorineural hearing loss (odds ratio = 0.8, 95% CI 0.67 – 0.98), cystatin C eGFR ( $r = 0.46$ ,  $P < 0.05$ ), erythropoietin ( $r = 0.55$ ,  $P < 0.01$ ), complete protein intake ( $r = 0.43$ ,  $P < 0.05$ ), and incomplete protein intake ( $r = -0.44$ ,  $P < 0.05$ ). LVEF% ( $P = 0.06$ ) and optic nerve atrophy (odds ratio = 0.73, 95% CI 0.5 – 1.04) showed statistical trends.

Lower 1-<sup>13</sup>C-propionate oxidation was associated with severe genotypes and more adverse clinical outcomes helping define the clinical prediction model of propionic acidemia. These data support the role of *in vivo* metabolic profiling as a surrogate endpoint in clinical trials aimed at restoring propionate oxidation in the liver.

## PgmNr 1206: The *Ornt1* KO mouse: A disease model for the HHH (Hyperornithinemia-Hyperammonemia-Homocitrullinuria) syndrome.

### Authors:

J. Camacho <sup>1,2</sup>; N. Rioseco-Camacho <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Loma Linda Children's Hospital, San Diego, California.; 2) Pediatrics, University of California Irvine, Orange, CA

---

**Introduction:** The Hyperornithinemia-Hyperammonemia-Homocitrullinuria (HHH) syndrome is a disorder of the urea cycle (UC), ornithine degradation pathway and cationic amino acid transport caused by mutations in the mitochondrial ornithine transporter, *ORNT1*. In contrast to most disorders of the UC that present as neonatal hyperammonemic coma (i.e. OTC deficiency), HHH syndrome has a later onset and more variable and mild clinical phenotype. We hypothesize that the HHH syndrome phenotype is influenced by the existence of two additional mitochondrial cationic amino acid transporters, *ORNT2* (non-functional in mouse) and *ORNT3* (SLC25A29). To understand the mechanisms of disease for HHH syndrome, we studied an available *Ornt1* KO mouse that was generated by replacing a region of intron 4 and exon 5 with a neomycin  $\beta$ -galactosidase cassette.

**Results:** Two phenotypes of homozygous (HMZ) pups (n=65) were observed: one group expired by 24-36 h (45%) and the second group developed severe neurological disease by 14-15 days of life before dying (55%). HMZ pups, significantly smaller in size at birth, exhibited marked growth delay. Progressive neurological symptoms (ataxia, loss of righting reflex, spasticity) started at day 10 of life. Hyperammonemia was observed in 10-14 d/o pups (Wild-type 258  $\pm$  43  $\mu$ m vs HMZ 872  $\pm$  182  $\mu$ m). Fatty liver was evident in  $\sim$ 50% of 10 d/o HMZ pups. Significant elevation in plasma ornithine (5X) and urinary excretion of cationic amino acids, homocitrulline, lactate and citrate were also observed. At non-saturating ornithine concentration, residual ornithine transport was measured in skin fibroblasts from HMZ pups when compared to WT.

**Conclusions:** The *Ornt1* KO mouse presents a phenotype similar to patients with HHH syndrome that includes spasticity, liver disease, growth delay and mild hyperammonemia. Despite a non-functional mouse *Ornt2*, *in vivo* gene redundancy is evident in the *Ornt1* KO mouse implying that *Ornt3* (Slc25a29) or another mitochondrial cationic amino acid transporter is responsible for the residual ornithine transport and variable phenotype in this murine disease model. Previous *in vitro* studies showed that SLC25A29 acts as an ornithine transporter whereas SLC25A20 (carnitine/acylcarnitine transporter) and SLC25A45 (two members of the ORNT1 subfamily) did not rescue the deficient ornithine transport in ORNT1 deficient fibroblasts.

## PgmNr 1207: High-dose hydroxocobalamin injection achieves biochemical correction and clinical improvement in adults with late-onset cobalamin C disease.

### Authors:

T. Higashimoto <sup>1</sup>; A.Y. Kim <sup>1</sup>; J.T. Ogawa <sup>1</sup>; M.A. Almuqbil <sup>1,2,3</sup>; J. Sloan <sup>4</sup>; I. Manoli <sup>4</sup>; C.P. Venditti <sup>4</sup>; M. Gunay-Aygun <sup>1</sup>; T. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine and Department of Pediatrics, Johns Hopkins University, Baltimore, Maryland; 2) Division of Pediatric Neurology, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia; 3) King Abdullah International Medical Research Center, King Abdullah Specialist Children's Hospital – Ministry of National Guard, Riyadh, Saudi Arabia; 4) Medical Genomics and Metabolic Genetics Branch, National Institute of Human Genome Research, National Institutes of Health, Washington DC

---

Cobalamin C (cblC) disease is the most common inborn error of intracellular cobalamin metabolism caused by pathogenic variants in *MMACHC*. It is characterized biochemically by accumulation of homocysteine (Hcy) and methylmalonic acid (MMA). cblC disease is classified into early- and late-onset forms. The late-onset form is rare and could present from late childhood to adult with neurological deteriorations, psychiatric disturbances, and thromboembolic complications. Hydroxocobalamin injection is a standard treatment. Its optimal dosing has not been systematically evaluated in adults. We report three adult siblings with a biochemically and molecularly confirmed cblC disease. The 28-year-old proband presented with severe psychosis, deep venous thrombosis with pulmonary embolism, and progressive neurological deterioration associated with extensive lesions in the spinal cord, periventricular white matter as well as basal ganglia. Serum total homocysteine and methylmalonic acid were markedly elevated. Serum Hcy and MMA levels were markedly elevated. Hydroxocobalamin injections, together with betaine and folic acid, were prescribed at standard (1 mg/day) and escalated (25 mg/day) doses in an attempt to maximize therapeutic benefits. This regimen led to a rapid normalization of serum Hcy and MMA levels, resolution of psychiatric symptoms, and improvement of neurological functions and brain and spinal cord lesions. Two biochemically affected but clinically asymptomatic siblings received hydroxocobalamin injections at 25 mg/day, which normalized serum MMA and Hcy in 2-3 weeks. Their injections were successfully spaced out to 25 mg/week for maintenance therapy. All three patients are compound heterozygotes for a known variant, Arg91LysfsX14, and a novel variant, Tyr130Cys, in *MMACHC*. This case highlights a need to evaluate defects of intracellular cobalamin metabolism in adults with severe neuropsychiatric manifestations. A longer follow up is needed to determine the optimal dosing regimen for acute (25 mg/day) and maintenance (25 mg/week) therapy in adult patients with late-onset cblC disease but the clinical and biochemical responses we have documented suggest that high doses hydroxocobalamin injections are beneficial.

# PgmNr 1208: Newborn screening program by tandem mass spectrometry for inborn metabolic disease in Hunan Province: Incidence, genetic characteristics, and clinical outcomes.

## Authors:

H.M. Yan <sup>1</sup>; Z.J. Jia <sup>1</sup>; J. Liu <sup>2</sup>; Y.H. Zhang <sup>1</sup>; H. Tang <sup>1</sup>; J. Chen <sup>1</sup>; J.Q. Fang <sup>2</sup>; D.H. Xie <sup>3</sup>; H. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Genetic Medicine, Maternal and Child Health Hospital of Hunan Province, Changsha, Hunan, China; 2) Department of Health Care, The Maternal and Child Health Hospital of Hunan Province, Changsha, Hunan, China;; 3) Department of Health Care, The Maternal and Child Health Hospital of Hunan Province, Changsha, Hunan, China

This study sought to investigate the prevalence, mutation characteristics and clinical outcomes of inborn metabolic disorders (IMD), detected by tandem mass spectrometry (MS/MS) newborn screening in Hunan province. We studied 565,182 newborns who underwent MS/MS screening for 26 IMDs, including fatty acid oxidation disorders (FAODs), amino acid disorders (AAs), and organic acidemias (OAs) between March 2013 and September 2017. For patients with positive results, a recall screening test was performed, and the results were further confirmed by specific biochemical and genetic analysis. For all patients with IMD, guideline-directed medical treatment was administered, and the follow-up outcomes were evaluated. A total of 107 newborns were diagnosed with IMDs, with an overall prevalence of 1:5,282, including 65 newborns with FAODs (1: 8,695), 29 with AAs (1:19,489), and 13 with OAs (1:43,476). The Primary carnitine deficiency (PCD) (1:12,845), Hyperphenylalaninemia (HPA) (1:33,246), Short-chain acyl-CoA dehydrogenase deficiency (SCADD) (1:47,099), Citrine deficiency (NICCD) (1:94,197) were the four most common IMDs in Hunan province. The hotspot mutation in SLC22A5 gene of PCD were c.51C>G (25.3%), c.1400C>G (23.0%), and c.760C>T (13.8%); in PAH gene of HPA were c.728G>A (22.2%) and c.721C>T (14.8%); in ACADS gene of SCADD was c.1031A>G (38.9%); and in SLC25A13 gene of NICCD was c.851\_854delGTAT (50.0%). The remaining IMDs were rare, and the hotspot mutations were unclear right now. During a mean follow-up of 26.1±5.6 months, 7 patients died, 4 were suffered an intelligent disability, whereas the remaining 96 subjects had normal physical and intelligent development. **Conclusion** The overall prevalence of IMDs is not fairly low in Hunan province. The spectrum of these disorders presents with typical regional characteristics, and the hotspot mutation profile is consistent with that detected in the Chinese population. Newborn screening by MS/MS enables the early detection of IMDs. Early appropriate management can significantly improve the outcomes of these patients.

## PgmNr 1209: Novel mutation in *MT-ND1* m.3955G>A related to neonatal onset Leigh syndrome with spinal lesion.

### Authors:

T. Kumaki <sup>1</sup>; H. Shimbo <sup>2</sup>; T. Goto <sup>2</sup>; Y. Enomoto <sup>1</sup>; N. Aida <sup>3</sup>; K. Murayama <sup>4</sup>; I. Nishino <sup>5</sup>; Y. Goto <sup>6</sup>; K. Kurosawa <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2) Division of Neurology, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 3) Division of Radiology, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 4) Center for Medical Genetics and Department of Metabolism, Chiba Children's Hospital, Chiba, Japan; 5) Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan; 6) Medical Genome Center, National Center of Neurology and Psychiatry, Tokyo, Japan

---

**Introduction** Leigh syndrome is subacute necrotizing encephalomyelopathy characterized by symmetrical lesion of basal ganglia and brain stem. *MT-ND1* gene mutation causes the functional decline of respiratory chain complex I activity and is associated with Leber hereditary optic atrophy, MELAS and Leigh encephalopathy. *MT-ND1* mutation related to neonatal onset lethal lactic acidosis was also reported. **Case report** ??The patient was born at 38 weeks gestation to nonconsanguineous healthy parents. At the age of 2 weeks she came into coma and was admitted to the hospital. Laboratory test revealed elevated serum lactate (9.5 mmol/L), and pyruvate (0.27 mmol/L). Cerebrospinal fluid lactate and pyruvate also elevated. MRI showed restriction of diffusion bilaterally in the cervical spinal cord, brain stem and cerebral white matter. Magnetic resonance spectroscopy showed elevation of lactate peak. At the age 3 weeks, MRI showed the diffusion restriction area were extended and cerebellar tonsillar herniation was present. She didn't recover from coma and couldn't breathe. She received a tracheostomy at the age 8 of weeks. Her condition came stable. At 3 months old, infection exacerbated her condition and she died. Whole mitochondrial genome sequencing detected homoplasmic *MT-ND1* m.3955G>A, p.Ala217Thr in her blood and muscle, which is not detected in her mother's blood. Whole exome sequencing detected no mutations associated with her symptoms. Enzyme activity of respiratory chain complex I was decreased in skin fibroblasts and muscle. **Discussion** We concluded *MT-ND1* variant was the cause of Leigh syndrome because decreased enzyme activity of respiratory chain complex I was consistent with *MT-ND1* mutation. Spinal cord and brain were involved in our case. Spinal lesion involvement has been reported in some patients with Leigh syndrome, but genetic causes have been identified in only a few cases including *NDUFA1*, *MTTK*, *MTRNR2*, *SURF1* mutation. This is the first case of *MT-ND1* mutation related to Leigh syndrome with spinal lesion. These results suggest that *MT-ND1* could be one of the causes of Leigh syndrome with spinal lesion.

## PgmNr 1210: Next generation sequencing analysis for mitochondrial disorder in Japanese patients.

**Authors:**

Y. Okamoto; A. Yoshimura; M. Ando; Y. Higuchi; A. Hashiguchi; H. Takashima

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dept Neurology & Geriatrics, Kagoshima Univ Grad Sch Medical & Dental Sci, Kagoshima City, Kagoshima, Japan

---

Mitochondrial disorders have a reported incidence of 1 in 5000 live births and are clinically and genetically heterogeneous group of diseases. A correct diagnosis is challenging, mainly because of the absence of clear phenotype-genotype correlations, and these diseases are attributed to mutations in the 16.6 kb mitochondrial genome and approximately 1500 genes encoded in the nuclear genome. The development of Next Generation Sequencing (NGS) has revolutionized the diagnostic approach. New approaches make genetic analyses much faster, more sensitive and more efficient. We have developed a targeted NGS-based assay of the mitochondrial DNA (mtDNA), which include 179 nuclear genes reference to the Human Gene Mutation Database (HGMD®) associated with mitochondrial disorders. Specifically, we designed and performed a 3-step strategy, comprising of clinical and pathological screening, long-range PCR using muscle DNA and whole mtDNA sequencing with Illumina MiSeq, and gene-targeted analysis following whole-exome sequencing (WES). A total of 150 suspected mitochondrial diseases cases were analyzed in five years from 2014 to 2019. Mutations were found in mtDNA in 15 of the 150 analyzed cases. Miseq was performed for 144 cases, and deletion of mtDNA was observed in 85 cases out of 144 cases (59%). WES was performed in 65 cases, 17 of which revealed mitochondrial related nuclear gene abnormalities (25.7%). Multiple cases have mild renal failure, external ophthalmoplegia, and neuropathy. Further study should investigate more nuclear genes and pathological study to clarify the pathophysiology. These symptoms were underestimated or undetected before diagnostic confirmation. Our results indicate that neuropathy might be an important symptom of adult mitochondrial disease caused by mutations in nuclear genes. We speculate that a noticeable number of mildly symptomatic patients with an adult mitochondrial disease existed.

## PgmNr 1211: The phenotypic spectrum of 43 Czech patients with single mitochondrial DNA deletion.

### Authors:

M. Tesarova; N. Anteneova; K. Berankova; H. Kolarova; A. Vondrackova; J. Zeman; T. Honzik

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

---

The single mitochondrial DNA deletion (del-mtDNA) is associated with sporadic clinical syndromes of varying severity ranging from Pearson syndrome, Kearns-Sayre syndrome (KSS) spectrum to isolated progressive external ophthalmoplegia (PEO). We present clinical and molecular data of Czech patients with single del-mtDNA. Results: Since 1992, a total of 43 patients have been diagnosed. Single del-mtDNA was detected in 85 DNA samples obtained from our patients (more than one tissue was tested in each patient), mtDNA deletion break-points were unequivocally characterized in 29 samples. The single del-mtDNA lengths were 1835-7827 bp, the single del-mtDNA heteroplasmy levels ranged from 31% up to 90%. The age of onset varied from weeks after birth to 79 years, 67% patients developed symptoms before the age of 20. Ptosis was the most common presenting symptom (36 patients), following by PEO, muscle weakness and dysarthria. Five patients were classified as having Pearson syndrome, 25 patients KSS spectrum (58%) and seven PEO. Six patients suffered from atypical phenotype manifesting myalgia and muscle weakness (3/6), isolated dysarthria or ptosis and muscle atrophy (2/6) and dysarthria and epilepsy (1/6). Conclusion: Six (14%) individuals manifested different clinical features further expanding the clinical spectrum associated with SMD phenotype. *Supported by AZV17-30965A and RVO-VFN64165.*



## PgmNr 1212: Genetic diagnosis has been provided for 260 patients with inherited metabolic diseases positively screened by newborn screening in Japan since 2014.

### Authors:

T. Fukao <sup>1,2</sup>; H. Sasai <sup>1,2</sup>; Y. Ago <sup>1,2</sup>; H. Matsumoto <sup>1</sup>; H. Otsuka <sup>1,3</sup>; J. Hosokawa <sup>4</sup>; R. Fujiki <sup>4</sup>; O. Ohara <sup>4</sup>; Y. Nakajima <sup>5</sup>; T. Ito <sup>5</sup>; K. Hara <sup>6</sup>; M. Kobayashi <sup>7</sup>; G. Tajima <sup>8</sup>; N. Ichinoi <sup>9</sup>; J. Kido <sup>10</sup>; S. Matsumoto <sup>10</sup>; K. Nakamura <sup>10</sup>; T. Hamazaki <sup>11</sup>; H. Kobayashi <sup>12</sup>; Y. Hasegawa <sup>12</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Pediatr, Gifu Univ, Gifu, Japan; 2) Div of Clin Genet, Gifu Univ Hosp, Gifu, Japan; 3) Dept of Neonat Med, Gifu Pref Gen Med Centr, Gifu, Japan; 4) Kazusa DNA Res Inst, Kisarazu, Japan; 5) Dept of Pediatr, Fujita Medical Univ, Toyoake, Japan; 6) Dept of Pediatr, Kure Med Centr, Kure, Japan; 7) Dept of Pediatr, Jikei Univ Med, Tokyo, Japan; 8) Div of Neonatal Screening, NCCHD, Tokyo, Japan; 9) Dept of Pediatr, Tohoku Univ, Sendai, Japan; 10) Dept of Pediatr, Kumamoto Univ, Kumamoto, Japan; 11) Dept of Pediatr, Osaka City Univ, Osaka, Japan; 12) Dept of Pediatr, Shimane Univ, Izumo, Japan

**BACKGROUND:** Newborn screening (NBS) using tandem mass started all over Japan in 2014. Since then the number of NBS target diseases increased from 6 to 20. We have been conducting a follow-up project of patients with inherited metabolic diseases (IMD), who were positively screened in NBS, after confirming their genotypes since 2014. This will provide us phenotype/genotype correlations in these NBS target IMD and important evidence for evaluation of NBS efficacy and genotypes-adjusted clinical guidelines.

**METHODS:** We designed a gene panel covering the NBS target IMD and the related diseases. Gene panel analysis was performed at Kazusa DNA Research Institute. In 2014-2016, we did mutation analysis on research basis with parents' informed consent, after 2017 mutation analysis for a large part of NBS-target IMD were covered by public medical insurance. We obtained written consents from parents for follow-up study. These studies were approved by the Ethical Committee of our institutions. This research was supported in part by AMED under Grant Number JP17ek0109276 and Health and Labour Sciences Research Grants (H29-nanchitou(nan)-ippan-051) for Research on rare and intractable diseases.

**RESULTS:** More than 340 cases were enrolled in this gene panel study for five years (January 2014 to March 2019). Among them, the number of patients with IMD detected by NBS were 260, as follows: Propionic acidemia (49), Hyperphenylalaninemia (37), Methylcrotonylglycinuria (27), VLCAD deficiency (19), Methylmalonic acidemia (18), Galactosemia (17), MCAD deficiency (16), Maple syrup urine disease (14), primary systemic carnitine deficiency (12), Glutaric acidemia type 2 (9), Homocystinuria(8), Others (34). Most of other 70 cases were symptomatic patients who were previously diagnosed. In most cases, we could find the gene mutations in their corresponding genes. We also started clinical follow-up studies.

**DISCUSSION:** In some NBS-target IMD, we found specific mutations only identified in NBS. For example Y435C was predominantly identified in Japanese patients with propionic acidemia who were identified in NBS. Clinical phenotype of patients with Y435C in PCCB (propionic acidemia) has been asymptomatic or mild. Similar findings were also seen in VLCAD deficiency. It is challenging to distinguish patients with high risk from those with no symptom or low risk and make genotype-



adjusted clinical guidelines at this moment. We need further long-term follow-up studies.

## PgmNr 1213: Modulation of the gut microbiome in a mouse model of isolated methylmalonic acidemia induces a therapeutic biochemical and hepatic response.

### Authors:

E. Isko <sup>1</sup>; A. Lesser <sup>1</sup>; C. Deming <sup>2</sup>; S. Conlan <sup>2</sup>; I. Manoli <sup>1</sup>; O. Shchelochkov <sup>1</sup>; J. Segre <sup>2</sup>; C. Venditti <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genomics and Metabolic Genetics Branch, NHGRI, NIH; 2) Microbial Genomics Section, NHGRI, NIH

---

Cobalamin A deficiency (cblA), caused by pathogenic variants in *MMAA*, is a form of isolated methylmalonic acidemia (MMA), a severe disorder of intermediary metabolism. MMA is characterized by multiorgan pathology including hepatic mitochondrial dysfunction. Clinical management of MMA involves a protein restricted diet, aggressive management of disease related symptoms, and, in some patients, intermittent antibiotic treatment to minimize the production of propionate. Although previous reports suggest antibiotic treatment could reduce serum methylmalonic acid concentrations by as much as 25% in human subjects with MMA, there is no experimental evidence to prove the beneficial effects of antibiotic treatment.

In the current study, we assessed the effects of antibiotic therapy in a mouse model of cblA deficiency. Vancomycin was administered in the drinking water of affected (*Mmaa*<sup>-/-</sup>) and unaffected (*Mmaa*<sup>+/-</sup>) mice for two weeks and resulted in drastic improvements in the mutant phenotype: weight increased 2-fold (multiple t-test:  $p < 0.05$ ), plasma MMA decreased 4-fold (Mann-Whitney test:  $p < 0.01$ ), and plasma Fgf21 decreased nearly 30-fold (Mann-Whitney test:  $p < 0.05$ ).

To explore the mechanisms underlying antibiotic effects, we performed RNA-seq on liver mRNA and bacterial 16S rRNA metagenomic sequencing on stool samples from affected and unaffected mice. Comparison of affected and unaffected liver transcriptomes revealed 9,632 differentially expressed genes in the disease state, of which vancomycin treatment restored expression in 83% (8,007) of these dysregulated genes (Fisher's exact test:  $p < 0.001$ ). Comparison of dysregulated genes to mRNA and proteomic datasets from MMA patients and MMA murine models identified a core set of 4 dysregulated pathways including mitochondrial energy homeostasis and proliferation, xenobiotic metabolism, and host-bacterial interactions ( $p < 0.001$ ). 16S sequencing of stool samples collected during and after vancomycin treatment revealed a loss of *Bacteroidetes*, *Firmicutes*, and *Actinobacteria* communities in both affected and unaffected mice.

This study is the first to clearly establish a therapeutic effect of antibiotics in isolated MMA and explore the gut microbiome effects on the liver transcriptome. The understanding of the gut-liver axis in MMA can point to new targets for small molecule therapeutics and disease-specific probiotics.

## PgmNr 1214: Novel mutations identified by whole exome sequencing in *ECHS1* from three Chinese patients with Leigh syndrome.

### Authors:

M. Wu <sup>1</sup>; Z. Liu <sup>2</sup>; Y. Liu <sup>3</sup>; D. Sun <sup>4</sup>; F. Fang <sup>2</sup>; Z. Liu <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Medicine, Jiangnan University, Wuhan 430056, China; 2) Department of Neurology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, 100045, China; 3) Aegicare (Shenzhen) Technology Co., Ltd., Shenzhen, 518110, China.; 4) Department of Pediatric Neurology, Wuhan Children's Hospital, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430016, China

**Background:** Mitochondrial short-chain enoyl-CoA hydratase-1 deficiency (ECHS1D, OMIM 616277) is a rare inborn metabolic disorder whose clinical presentations are consistent with Leigh Syndrome (LS, OMIM 256000), a clinically and genetically heterogeneous disorder that commonly onset within the first months or years of life. It was first characterized to be caused by biallelic mutations of gene *ECHS1* (OMIM 602292) in 2014. The genetic errors in *ECHS1* gene affects the fourth step of valine degradation and leads to the accumulation of toxic intermediates, methacrylyl-CoA and acryloyl-CoA, to further cause brain pathology. To date, more than 30 different pathogenic mutations have been identified from over 40 patients.

**Methods:** We implemented whole-exome sequencing in three Chinese patients diagnosed as LS to precisely identify the disease-causing variants. During the diagnosis, other examinations and biomedical measurements such as metabolic measurement, mitochondrial respiratory chain (MRC) enzyme activity measurement, oxygen consumption rate (OCR) measurement and brain magnetic resonance imaging (MRI), were also carried out.

**Results:** WES analysis of three patients identified four distinct mutations in *ECHS1* to be the pathogenesis. Among the detected mutations, only one, c.583G>A (p.Gly195Ser), has been reported in previous studies; the other three, c.463G>A (p.Gly155Ser), c.557C>T (p.Ser186Leu) and c.476\_c.477delAGinsGGCATAGA (p.Gln159delinsLeuTyrAla), are all novel. It is worth noting that one of the novel variants, c.463G>A (p.Gly155Ser), was detected in all three patients from unrelated families, indicating a potential founder effect, which has already been previously reported to be existing in LS in other mutations. Clinically, despite different initial presentations, all studied patients were with similar clinical syndromes such as development regression, paroxysmal exercise-induced dystonia, as well as common radiological features like symmetrical bilateral brain abnormalities, similar metabolic results like consistently elevated plasma. The presence of the marker metabolite 2-methyl-2,3-hydroxybutyric acid in urine organic acids have also been detected.

**Conclusion:** Our findings enriched the pathogenic mutation spectrum of *ECHS1* gene and confirmed the clinical and phenotypic presentations of ECHS1D in LS. Further studies are required to look into the noteworthy novel mutation c.463G>A (p.Gly155Ser), especially in Chinese origin population.

## PgmNr 1215: The G6PD-mediated mechanism of inhibiting the invasion and proliferation of *plasmodium falciparum* by dehydroepiandrosterone and nicotinamide.

### Authors:

W. Jiang; P. Du; C. Jiang; J. Yan; Q. Hu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medical Genetics, Zhongshan School of Medicine, Sun Yat-Sen University, 74# Zhongshan Road, Guang Zhou, Guangdong, China

**Background:** New drugs are urgently needed to control artemisinin-resistant *Plasmodium* strains. Our previous study revealed that plasmodium multiplication would understandable inhibited in G6PD deficient erythrocytes because the plasmodium organelles could not obtain enough NADPH. However, we rare know the effect of insufficient ribose-5-phosphate on plasmodium multiplication in G6PD deficient erythrocytes and the potential mechanism of dehydroepiandrosterone and nicotinamide decreases G6PD activity to inhibit the invasion and proliferation of *Plasmodium falciparum*.

**Methods:** Late parasites were isolated from infected RBCs using flow cytometry. The R-5-P was transferred to erythrocyte by electrorotator, followed R-5-P was determined quantitatively by mass spectrometry. The infection and proliferation of *P. falciparum* were detected by Giemsa staining and real-time FQ PCR. After the infected erythrocytes were treated by DHEA and nicotinamide, G6PD activity was assayed by standard method, Young's modulus of erythrocyte membranes was measured to observe the mechanical properties by atomic force microscopy and the invasion and proliferation of *Plasmodium falciparum* were observed by Giemsa staining and real-time FQ PCR.

**Result:** The invasion and proliferation of *P. falciparum* in G6PD deficient erythrocyte were significantly reduced. When ribose-5-phosphate was transferred into G6PD deficient erythrocyte, proliferation of *P. falciparum* will be rescued. After the infected erythrocytes were treated by DHEA and nicotinamide, G6PD activity decreased and Young's modulus of erythrocyte membranes decreased with the altered mechanical properties. The minimum effective concentration of DHEA to inhibit *P. falciparum* was 12.5  $\mu\text{mol/L}$ , and nicotinamide was 2mM/L.

**Significance:** G6PD activity decreased after dehydroepiandrosterone and niacinamide treated, which mediated to inhibit invasion and proliferation of *P. falciparum* in host erythrocyte. The mechanism is not only involved in modulation of the redox state due to reduced NADPH, but also in decreased nucleotide synthesis due to reduced ribose-5-phosphate in pentose phosphate pathway, which is main source for nucleotide synthesis of *P. falciparum* in host erythrocyte. G6PD may be the target of the new drug for fighting against malaria. Our results also showed that both dehydroepiandrosterone and niacinamide have anti-malaria effect mediated by G6PD and the drug combination more effectively and less adverse effects.

## PgmNr 1216: Natural course of auditory function in Takenouchi-Kosaki syndrome.

**Authors:**

T. Takenouchi; K. Kosaki

[View Session](#) [Add to Schedule](#)

**Affiliation:** Keio Univ, Tokyo, Japan

---

**Background:** We recently established a novel multi-system disorder caused by mutations in *CDC42*, which was eponymized as Takenouchi-Kosaki syndrome (OMIM# 616737). All reported cases exhibited sensorineural hearing loss.

**Methods:** We retrospectively analyzed auditory brainstem response (ABR) and imaging data in two patients with Takenouchi-Kosaki syndrome. Both patients had hearing aids and had a de novo heterozygous mutation in *CDC42*, i.e., p.Tyr64Cys.

**Results:** patient 1: ABR at the age of four years showed clear waves I, III and V at 80dB, whereas ABR at the age of 21 years showed a blurred wave I at 80dB. Compared to audiometry, the patient showed a decreased response in ABR and otoacoustic emission. Patient 2: ABR at the age of seven years showed a clear wave V at 60dB on the right and at 70dB on the left. At the age of twelve years, her ABR showed a clear wave V at 80dB bilaterally, and a barely recognizable wave I at 90dB. The results were compatible with those in audiometry. A computed tomography scan of the inner ear showed a decreased number of rotations of cochlear and was compatible with 'incomplete partition type I', according to the classification of cochleovestibular malformations by Sennaroglu and Saatci.

**Discussion:** Our retrospective analysis on the natural course of auditory function in two patients with Takenouchi-Kosaki syndrome showed a possible progressive hearing loss from childhood through adulthood. These clinical observations appeared compatible with existing findings in *cdc42* knock-out mice that exhibited progressive hearing loss and loss of inner hair cells. The anatomical changes of the inner ear of the patient 2 suggested developmental arrest during the fetal period.

## PgmNr 1217: Metaphyseal enchondromatosis with D-2-hydroxyglutaric aciduria and variants in *IDH1*, *IDH2*, and *EXT2*.

### Authors:

E.S. Partan <sup>1,2</sup>; L. Kratz <sup>3</sup>; G.S. Gottesman <sup>4</sup>; D. Valle <sup>1,5</sup>; N. Sobreira <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Predoctoral Training Program in Human Genetics, Johns Hopkins School of Medicine, Baltimore, MD; 3) Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, MD; 4) Center for Metabolic Bone Disease and Molecular Research, Shriners Hospitals for Children, St. Louis, MO; 5) Department of Pediatrics, Johns Hopkins School of Medicine, Baltimore, MD

---

Metaphyseal enchondromatosis with D-2-hydroxyglutaric aciduria (ME-HGA; OMIM 614875) is a rare disorder characterized by enchondromas and elevated levels of D-2-hydroxyglutaric acid (D-2HG). Vissers et al. (2011) reported the somatic mosaic *IDH1*-R132 variant in two unrelated patients with ME-HGA. Heterozygous, somatic, gain-of-function (GOF) variants of *IDH1* (R132C, R132G, R132H, R132L, and R132S) and *IDH2* (R172S) have been reported in a variety of cancers as well as Ollier disease (OD; OMIM 166000) and Maffucci syndrome (MS; OMIM 614569). OD and MS are related conditions characterized by enchondromas with a 30% risk of chondrosarcoma transformation. While wild-type *IDH1* and *IDH2* catalyze the oxidative carboxylation of isocitrate to  $\alpha$ -ketoglutarate ( $\alpha$ KG), the GOF variants instead cause  $\alpha$ KG to be converted to D-2HG. The conversion of  $\alpha$ KG to D-2HG leads to inhibition of  $\alpha$ KG-dependent reactions promoting increased histone and DNA methylation as well as impaired cell differentiation. Here, we present a 4-year-old boy with delayed tooth eruption, bilateral multiple enchondromas (shoulders, arms, elbows, wrists, hands, ribs, spine, hips, knees, ankles, and feet), and D-2HG aciduria. He had a normal CGH array. We performed WES of his blood and two enchondroma biopsies and identified a germline, heterozygous *IDH2*-T435M variant in all samples tested. We also identified a somatic mosaic, heterozygous *IDH1*-R132C GOF variant in both enchondromas but not in the blood sample. The proband also had a germline, heterozygous *EXT2*-T620M variant in all samples tested. Parents' samples are now being tested to determine if the germline *IDH2*-T435M and *EXT2*-T620M variants are *de novo*. The *IDH2*-T435M variant is of unknown significance with a gnomAD MAF of 3.61e-3. Loss of function variants in *EXT2* are known to cause autosomal dominant multiple exostoses (OMIM 133701). The *EXT2*-T620M variant is of unknown significance but is predicted to be pathogenic by DANN, dbNSFP, FATHMM, GERP, LRT, MetaLR, MetaSVM, MutationAssessor, MutationTaster, PROVEAN, and SIFT and has a gnomAD MAF of 6.62e-4. Herein we suggest that these *IDH1*, *IDH2*, and *EXT2* variants interact in this patient to cause his severe multiple enchondromas phenotype with D-2HG aciduria, supporting our hypothesis that these are tumor disposition syndromes characterized by locus heterogeneity with germline (or early post-zygotic) and additional tumor variants leading to tumor formation.

## PgmNr 1218: The CHD4-related syndrome: Investigation of the clinical spectrum, genotype-phenotype correlations and molecular basis.

### Authors:

K. Weiss <sup>1</sup>; H.P. Lazar <sup>2</sup>; A. Kurolap <sup>1</sup>; M.F. Smeland <sup>3</sup>; S. Wallen <sup>4</sup>; H. Solveig <sup>4</sup>; S.A. Schrier Vergano <sup>5</sup>; H. Hove <sup>6</sup>; S.C. Quinonez <sup>7</sup>; E. Colin <sup>8</sup>; M. Jain <sup>9</sup>; D. Monteil <sup>10</sup>; E.R. Roeder <sup>11</sup>; A. van Haeringen <sup>12</sup>; M. Gambello <sup>13</sup>; A. Santani <sup>14</sup>; E. DeChene <sup>14</sup>; L. Medne <sup>14</sup>; C.M. Skraban <sup>14</sup>; T. Smol <sup>15</sup>; J. Ghoumid <sup>15</sup>; M.J. Parker <sup>16</sup>; M. Wright <sup>17</sup>; P. Turnpenny <sup>18</sup>; J. Clayton-Smith <sup>19</sup>; K. Metcalfe <sup>19</sup>; P.M. Campeau <sup>20</sup>; M. Muenke <sup>21</sup>; P. Wade <sup>2</sup>; K. Lachlan <sup>22</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, Rambam Health Care Campus, Haifa, Israel; 2) Epigenetics and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.; 3) Department of medical genetics, University Hospital of North Norway, Tromsø, Norway; 4) AP-HP, Département de Génétique, Centre de Référence Maladies Rares "Anomalies du développement et syndromes malformatifs" Hôpital de la Pitié Salpêtrière, Paris, France; 5) Division of Medical Genetics and Metabolism, Children's Hospital of The King's Daughters, Department of Pediatrics, Eastern Virginia Medical School.; 6) Centre for Rare Diseases, Rigshospitalet, Copenhagen University Hospital, Denmark; 7) Department of Pediatrics, Division of Genetics, Metabolism and Genomic Medicine, University of Michigan, Ann Arbor, MI, USA.; 8) Department of Biochemistry and Genetics, University Hospital Angers, France; 9) Department Bone and Osteogenesis Imperfecta, Kennedy Krieger Institute, Baltimore MD USA; 10) Department of Pediatrics, Naval Medical Center Portsmouth; 11) Departments of Pediatrics and Molecular and Human Genetics, Baylor College of Medicine, San Antonio, TX USA; 12) Department of Clinical Genetics, Leiden University Medical Center. Leiden, the Netherlands; 13) Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA, USA.; 14) Division of Human Genetics, Children's Hospital of Philadelphia; 15) Department of Clinical Genetics, Lille University Hospital, CHU Lille, Lille, France.; 16) Sheffield Children's Hospital NHS Foundation Trust, Western Bank, Sheffield, UK.; 17) Northern Genetics Service, Newcastle Upon Tyne Hospitals NHS Foundation Trust. Institute of Human Genetics, International Centre for Life, Newcastle upon Tyne, UK.; 18) University of Exeter Medical School. Clinical Genetics Royal Devon & Exeter Hospital, Exeter EX1 2ED, UK; 19) Institute of Evolution, Systems and Genomics, Faculty of Medical and Human Sciences, University of Manchester. Manchester Centre For Genomic Medicine, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester; 20) Department of Pediatrics, University of Montreal and CHU Sainte-Justine, Montreal, QC, Canada H3T1C5.; 21) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, 20892, USA; 22) Wessex Clinical genetics Service, University Hospital Southampton NHS Trust. Department of Human Genetics and Genomic Medicine, Southampton University, Southampton, UK.

---

**Purpose:** Sifrim-Hitz-Weiss syndrome (SIHIWES) is a recently described multi-systemic neurodevelopmental disorder caused by *de novo* variants in chromodomain helicase DNA-binding protein 4 (*CHD4*). This condition is part of an increasingly recognized group of Mendelian disorders involving chromatin remodeling abnormalities. In this study, we investigated the clinical spectrum of

the disorder, genotype-phenotype correlations and the effect of different missense variants on CHD4 function.

**Methods:** We collected detailed clinical and molecular data from 32 individuals with mostly *de novo* variants in *CHD4*, identified through next generation sequencing. We performed 3D protein modeling, ATP hydrolysis and nucleosome remodeling assays on variants from five different CHD4 domains.

**Results:** The majority of participants had global developmental delay and mild to moderate intellectual disability. However, a few participants had an IQ score in the normal range. Abnormal brain imaging was seen in 69% overall, and in 96% of the participants that underwent brain MRI. A range of findings were reported including ventriculomegaly, hydrocephalus requiring shunting, Chiari 1 and a thin corpus callosum. Two participants were diagnosed with moyamoya disease and subsequent stroke; Congenital heart defects were present in 65% and macrocephaly was seen in 40%. Additional common abnormalities included hypogonadism in males, skeletal and limb anomalies, hearing impairment and ophthalmic abnormalities.

The majority of variants were non-truncating and affected the SNF2-like region of the protein. We did not identify genotype-phenotype correlations based on the location of variants. For the 3 individuals with truncating variants, the phenotype involved fewer systems and the facial features were not similar. Alterations in ATP hydrolysis and chromatin remodeling activities were detected in missense variants from five different domains.

**Conclusion:** The CHD4-related syndrome is a multi-systemic neurodevelopmental disorder. Missense substitutions in different protein domains alter CHD4 function in a mutation-specific manner, but result in a similar phenotype in humans. Further research is required in order to better understand the disease mechanism and to determine the pathogenicity of truncating variants in this condition.



## PgmNr 1219: Distinct alterations in tricarboxylic acid cycle metabolites are associated with cancer and autism phenotypes in Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome.

### Authors:

L. Yehia<sup>1</sup>; Y. Ni<sup>2</sup>; F. Feng<sup>1</sup>; M. Seyfi<sup>1</sup>; T. Sadler<sup>1</sup>; T.W. Frazier<sup>1,3</sup>; C. Eng<sup>1,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA; 2) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA; 3) Autism Speaks, Cleveland, OH 44124, USA; 4) Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH 44195, USA; 5) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA; 6) Germline High Risk Cancer Focus Group, CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106, USA

---

While detecting germline mutations to identify heritable disease risk is utile, not all individuals with predisposition alleles will develop identical phenotypes. Thus, identifying factors for individual-level risk stratification is necessary even among those with germline mutations in an identical gene. Cowden syndrome (CS) and Bannayan-Riley-Ruvalcaba syndrome (BRRS) are autosomal dominant disorders characterized by a high risk of breast, thyroid and other cancers, and in a subset, autism spectrum disorder (ASD). In *PTEN*<sup>MUT</sup> CS/BRRS patients, up to 10% have germline *SDHx* (succinate dehydrogenase, mitochondrial complex II) variants, which modify cancer risk. *PTEN* contributes to metabolic reprogramming, a well-established role in a cancer context. Relatedly, *SDH* sits at the crossroad of the electron transport chain and tricarboxylic acid (TCA) cycle, two central bioenergetic pathways. Here, we conducted a TCA targeted metabolomics study on 511 CS/BRRS individuals with various genotypes (*PTEN* or *SDHx*, mutant or wildtype) and phenotypes (cancer or ASD), and a series of 98 population controls. We generated immortalized lymphoblastoid cell lines from peripheral blood samples (n=609) following standard methods, and cultured *in vitro* for subsequent metabolic profiling using liquid chromatography tandem mass spectrometry (LC-MS/MS). To evaluate correlations among genotypic/phenotypic features and normalized metabolite levels, we applied linear regression models controlling for age and gender. We found consistent TCA metabolite alterations in patients with various genotypes and phenotypes compared to controls, with unique correlations of individual metabolites with particular genotype-phenotype combinations. Notably, increased isocitrate ( $p=1.2 \times 10^{-3}$ ) but reduced citrate ( $p=5.0 \times 10^{-4}$ ), associated with breast cancer in *PTEN*<sup>MUT</sup>/*SDHx*<sup>WT</sup> patients. In contrast, increased lactate associated with ASD regardless of genotype ( $p=9.7 \times 10^{-3}$ ), which replicated in an independent validation series (n=171) enriched for idiopathic ASD (*PTEN*<sup>WT</sup>,  $p=5.6 \times 10^{-4}$ ). Importantly, we identified fumarate ( $p=1.9 \times 10^{-2}$ ) as a pertinent predictive biomarker to distinguish individuals who will develop ASD versus those who will develop cancer. Our data suggest TCA cycle metabolite alterations are germane to the pathobiology of *PTEN*-related CS/BRRS and to that of ASD (latter regardless of genotype), with potential implications for risk prediction for the individual and for therapeutic metabolic correction.

## PgmNr 1220: Comprehensive genetic analysis of 57 families with clinically suspected Cornelia de Lange syndrome.

### Authors:

H. Aoi <sup>1,2</sup>; T. Mizuguchi <sup>1</sup>; J. Ceroni <sup>3</sup>; V. Kim <sup>3</sup>; I. Furquim <sup>3</sup>; R. Honjo <sup>3</sup>; T. Iwaki <sup>4</sup>; F. Sekiguchi <sup>1</sup>; Y. Uchiyama <sup>1,5</sup>; Y. Azuma <sup>1</sup>; K. Hamanaka <sup>1</sup>; E. Koshimizu <sup>1</sup>; S. Miyatake <sup>1,6</sup>; S. Mitsuhashi <sup>1</sup>; A. Takata <sup>1</sup>; N. Miyake <sup>1</sup>; D. Bertola <sup>3</sup>; C. Kim <sup>3</sup>; N. Matsumoto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Yokohama City University, Yokohama, Japan; 2) Department of Obstetrics and Gynecology, Juntendo University, Tokyo, Japan; 3) Clinical Genetics Unit, Instituto da Crianca, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, Brazil; 4) Department of Pediatrics, University Hospital, Faculty of Medicine, Kagawa University, Kagawa, Japan; 5) Department of Oncology, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 6) Clinical Genetics Department, Yokohama City University Hospital, Yokohama, Japan

---

Cornelia de Lange syndrome (CdLS) is a rare multisystem disorder with specific dysmorphic features. Pathogenic genetic variants encoding cohesion complex subunits and interacting proteins (e.g., *NIPBL*, *SMC1A*, *SMC3*, *HDAC8*, and *RAD21*) are the major cause of CdLS. However, there are many clinically diagnosed cases of CdLS without pathogenic variants in these genes. To identify further genetic causes of CdLS, we performed whole exome sequencing in 57 CdLS families, systematically evaluating both single nucleotides variants (SNVs) and copy number variations (CNVs). We identified pathogenic genetic changes in 36 out of 57 (63.2 %) families, including 32 SNVs and four CNVs. Two known CdLS genes, *NIPBL* and *SMC1A*, were mutated in 23 and two cases, respectively. Among the remaining 32 individuals, four genes (*ANKRD11*, *EP300*, *KMT2A*, and *SETD5*) each harbored a pathogenic variant in a single individual. These variants are known to be involved in CdLS-like. Furthermore, pathogenic CNVs were detected in *NIPBL*, *MED13L*, and *EHMT1*, along with pathogenic SNVs in *ZMYND11*, *MED13L*, and *PHIP*. These three latter genes were involved in diseases other than CdLS and CdLS-like. Systematic clinical evaluation of all patients using a recently proposed clinical scoring system showed that *ZMYND11*, *MED13L*, and *PHIP* abnormality may cause CdLS or CdLS-like.

## PgmNr 1221: Co-occurrence of two syndromes resulting in a mixed phenotype-Prader Willi syndrome and KBG syndrome.

### Authors:

P. Gupta; H. Salhah; L. Tambini King; A. Gomez; L. Fried; K. Harwood

[View Session](#) [Add to Schedule](#)

**Affiliation:** St. Joseph's University Medical Center, Paterson, New Jersey.

---

**Introduction:** Prader-Willi syndrome (PWS) is the most common syndromic form of obesity and affects between 350,000 and 400,000 individuals worldwide. KBG syndrome has been reported in about 100 cases worldwide and is due to a mutation in the *ANKRD11* gene. We present the first case of a child with the co-occurrence of both syndromes.

**Case Presentation:** The patient was born via C- Section to a 39-year-old G2P1 female at 38 weeks gestation. He was the product of an IVF pregnancy with a frozen embryo. Prenatal ultrasounds noted single umbilical artery, breech presentation, polyhydramnios, and growth delay. Amniocentesis was declined. His birth weight was 4 pounds. He was admitted to the NICU due to an aspiration event. He continued to exhibit hypotonia and poor weight gain and underwent G-tube placement with Nissen fundoplication and muscle biopsy. He was discharged from NICU after one month. FISH for Prader Willi/Angelman syndrome and microarray were normal but a subsequent methylation analysis confirmed the diagnosis of PWS. He was evaluated in the genetics clinic around one year of age and was noted to have multiple dysmorphic features in addition to severe hypotonia, developmental delay, and failure to thrive. He has been treated with growth hormone since one year of age. Now at nine years of age, he is mostly nonverbal but occasionally uses some repetitive phrases. He does not display food-seeking behavior and has remained below the 5<sup>th</sup> percentile in weight for age. Due to many phenotypic features that seemed to be unexplained by PWS alone, a whole exome trio test was performed, which was positive for a pathogenic mutation in the *ANKRD11* gene.

**Discussion:** This is the first reported case of a patient with the simultaneous diagnosis of PWS and KBG syndrome. KBG syndrome is characterized by macrodontia of the upper central incisors, a distinct craniofacial appearance, skeletal abnormalities and neurological involvement that includes developmental delay, intellectual disability, and seizures. Our patient has prominent central incisors, distinct dysmorphic features, severe developmental delays and failure to thrive, likely resulting from KBG syndrome.

**Conclusion:** When the phenotype of a patient cannot be explained by one diagnosis alone, it is imperative to consider additional possibilities. Obtaining an accurate diagnosis not only helps in understanding the present symptoms, but it also brings closure to previously unanswered questions.

## PgmNr 1222: Review of the phenotypic spectrum associated with haploinsufficiency of *MYRF*.

### Authors:

K.E. Glinton<sup>1</sup>; L.Z. Rossetti<sup>1</sup>; B. Yuan<sup>1,2</sup>; P. Liu<sup>1,2</sup>; N. Pillai<sup>1</sup>; E. Mizerik<sup>1</sup>; P. Magoulas<sup>1</sup>; J.A. Rosenfeld<sup>1</sup>; L. Karaviti<sup>3</sup>; V.R. Sutton<sup>1</sup>; S.R. Lalani<sup>1</sup>; D.A. Scott<sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 2) Baylor Genetics, Houston, TX; 3) Division of Pediatric Endocrinology, Baylor College of Medicine, Houston, TX; 4) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX

---

The myelin regulatory factor gene (*MYRF*) encodes a transcription factor that is widely expressed. There is increasing evidence that heterozygous loss-of-function variants in *MYRF* can lead to abnormal development of the heart, genitourinary tract, diaphragm, and lungs. Here, we searched a clinical database containing the results of 12,000 exome sequencing studies. We identified three previously unreported males with putatively deleterious variants in *MYRF*: one with a point mutation predicted to affect splicing and two with frameshift variants. In all cases where parental DNA was available, these variants were found to have arisen de novo. The phenotypes identified in these subjects included a variety of congenital heart defects (hypoplastic left heart syndrome, scimitar syndrome, septal defects, and valvular anomalies), genitourinary anomalies (ambiguous genitalia, hypospadias, and cryptorchidism), congenital diaphragmatic hernia, and pulmonary hypoplasia. The phenotypes seen in our subjects overlap those described in individuals diagnosed with PAGOD syndrome [MIM# 202660], a clinically defined syndrome characterized by pulmonary artery and lung hypoplasia, agonadism, omphalocele, and diaphragmatic defects that can also be associated with hypoplastic left heart and scimitar syndrome. These cases provide additional evidence that haploinsufficiency of *MYRF* causes a genetic syndrome whose cardinal features include congenital heart defects, urogenital anomalies, congenital diaphragmatic hernia, and pulmonary hypoplasia. We also suggest that individuals with a clinical diagnosis of PAGOD syndrome could be screened for pathogenic variants in *MYRF*, and that individuals with *MYRF* deficiency who survive the neonatal period should be monitored closely for developmental delay and intellectual disability.

## PgmNr 1223: Bottom-up approach: An efficient workflow for the characterization of rare genetic diseases.

### Authors:

A. Angural <sup>1</sup>; A. Spolia <sup>1</sup>; A. Mahajan <sup>1</sup>; V. Verma <sup>1</sup>; E. Rai <sup>1</sup>; K.K. Pandita <sup>2</sup>; S. Sharma <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics Research Group, School of Biotechnology, Shri Mata Vaishno Devi University, Kakryal, Jammu and Kashmir, India - 182320; 2) Consultant, Internal Medicine, Shri Mata Vaishno Devi Narayana Superspeciality Hospital, Kakryal, Katra, Jammu and Kashmir, India - 182320

---

Rare diseases constitute one of the major global health issues that cumulatively affect nearly 10% of the world's population. These are progressively disabling and life-threatening and usually precipitate due to a single, highly penetrant genetic background at fault. Though such diseases are individually less prevalent, yet some of these have a relatively higher incidence in endogamous and consanguineous population groups such as the Ashkenazi Jews, the Finnish population, the Indian Agarwal community, and many others. A higher burden of rare diseases among the heterogeneous, highly endogamous population groups and consanguineous families of Jammu and Kashmir (J&K), located in the northern Himalayan ranges of India, is anticipated. The geographical isolation of these population groups in the difficult-to-reach mountainous terrains along with a higher incidence of endogamy and consanguinity has led to a relatively higher burden of genetic diseases, most of which have remained uncharacterized for years owing to the lack of genetic studies based on the population genetic profile and genetic disorders prevalent in J&K. We have designed a workflow namely the "Bottom-up Approach" for delineation of patho-genetics of genetic disorders prevalent in J&K. This strategic workflow is based on different study designs and is carried out with the aid of state-of-the-art biological techniques including targeted Sanger sequencing of candidate genes and high-throughput Next-Generation sequencing technologies and several *in silico* tools for the analysis of the raw sequencing data and functional/pathogenicity prediction of the potential disease-causing variants. The success of the designed "Bottom-up Approach" is boasted by the precise delineation and reporting of three rare diseases namely Progressive Pseudorheumatoid Dysplasia (PPD), Pantothenate Kinase-Associated Neurodegeneration (PKAN) and Leigh Syndrome (LS) prevalent in different regions in J&K for the first time. These cases were found to be associated with some novel and rare genetic variants that are individually restricted to highly extended, endogamous or consanguineous families. In conclusion, Bottom-up Approach is a powerful workflow that holds relevance in the characterization of less frequent monogenic Mendelian diseases and identification of potential genetic markers and could prove useful in future for precise diagnosis of such diseases and their timely therapeutic management in the affected individuals.

## PgmNr 1224: Inhibition of G-protein signaling in cardiac dysfunction of Intellectual Developmental Disorder with Cardiac Arrhythmia (IDDCA) syndrome.

### Authors:

P. De Nittis <sup>1</sup>; A. Sarre <sup>2</sup>; J. Chrast <sup>1</sup>; W.F. Simonds <sup>3</sup>; T. Pedrazzini <sup>4</sup>; A. Reymond <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Integrative Genomics, University of Lausanne, Lausanne, Vaud, Switzerland; 2) Cardiovascular Assessment Facility, University of Lausanne, Switzerland; 3) Metabolic Disease Branch/NIDDK, National Institutes of Health, Bethesda, MD, USA; 4) Experimental Cardiology Unit, Department of Cardiovascular Medicine, University of Lausanne, Switzerland

Inactivating and hypomorphic mutations in the  $\beta 5$ -subunit of guanine nucleotide-binding protein (GNB5), respectively cause severe and mild forms of IDDCA syndrome, an autosomal recessive disorder associated with cognitive disability and cardiac arrhythmia. Whereas mice *Gnb5* knock-out (KO) showed structural and functional abnormalities of the brain, little was known about possible cardiac conduction anomalies. Of note, IDDCA individuals present bradycardia and sinus arrhythmias. Our mice study shows that *Gnb5*-inhibitory signalling is essential for parasympathetic control of heart rate. We monitored i) heart electrophysiology in KO (n=11), heterozygous (HET, n=8) and wild-type (WT, n=10) mice using *in vivo* electrocardiography (ECG) telemetry, and ii) cardiac function by echocardiography. 24-hours ECG recordings unearthed a high number of arrhythmias in KO animals, including escape beats (n=130 in KO vs. n=0.7 in WT), atrioventricular block (42 vs. 0.6) and tachycardia/bradycardia episodes (25 vs. 0.1), all signs of sinus arrhythmia and conduction problems. Moreover, KO mice were smaller than WT and HET mice, and had a smaller heart, but exhibited better cardiac function, as judged by increased fractional shortening and ejection fraction. Consistent with this observation, comparison of transcriptome profiling of atria and ventricles from WT, HET and KO mice revealed overexpression of genes involved in cardiac muscle contractility (e.g. *Tnnt1* (P=0.002), and *Lrrc10* (P=0.005) in ventricles, and *Myh7* (P=0.01) in atria) along with reduced ventricular expression of genes required for development of pacemaker cells (e.g. *Tbx18* (P=0.0001)) in KO animals. Additionally, KO mice treated with carbachol, a parasympathomimetic, presented significant carbachol-induced bradycardia. In contrast, the anti-parasympathetic drug atropine had the same effect in KO, HET and WT.  $\beta$ -adrenergic activity of KO mice was unaltered, suggesting a normal sympathetic modulation of the cardiac stimulation. We hypothesize that the disease mechanism responsible for heart rate perturbations is the loss of negative regulation on the inhibitory G-protein signalling, resulting in enhanced parasympathetic activity.

Overall, our work highlights that *Gnb5* KO mice not only recapitulate IDDCA's neurologic manifestations, but also mimic the cardiac perturbations, allowing for future screening of drugs modulating the parasympathetic branch of the autonomic nervous system, in view of patients' therapy.

## PgmNr 1225: Expanding the phenotype of mosaic tetrasomy 5p: Female patient with severe scoliosis and Wolff-Parkinson-White syndrome.

### Authors:

C. Chang <sup>1</sup>; J. Chernos <sup>1,2</sup>; M.A. Thomas <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Alberta Children's Hospital, Calgary, Alberta, Canada; 2) Cytogenetics laboratory, Genetics and Genomics, Alberta Public Laboratories, Alberta Children's Hospital, Calgary, Alberta, Canada

---

**Background:** Supernumerary Isochromosomes are made up of two copies of the same arm of a chromosome leading to tetrasomy for that region. Mosaic tetrasomy of the whole 5p resulting from an isochromosome 5p, or i(5)(p10), has been described in 5 affected children to date, with features including developmental delay, seizures, hypotonia, dysmorphic features, structural brain anomalies, and streaky hyperpigmentation of the skin. Less frequent findings include cardiac anomalies, renal anomalies, musculoskeletal anomalies, and early death due to cardiorespiratory compromise. In addition, mosaic isochromosome 5p has been described in one unaffected adult with infertility, and there are also several reports of partial tetrasomy of distal 5p, with a total of 15 patients with mosaic tetrasomy 5p reported. Due to the rarity of the condition (incidence estimated < 1 in 1,000,000), there are few reports of long-term outcomes.

**Methods/results:** We describe the 6<sup>th</sup>, and oldest reported affected child to date, with a mosaic finding of tetrasomy 5p resulting from a supernumerary isochromosome 5p at a 5% level in peripheral blood. Parental karyotype demonstrated the de novo origin of this finding. We report clinical features, including developmental outcomes at 12 years of age. We compare the features seen in our patient to those of patients reported in the literature. Our patient did not have skin pigmentary findings, but had developmental delay/ID, seizures, hypotonia, cardiac anomaly, and dysmorphic features. In addition, she had severe scoliosis requiring surgery with Harrington rods, and Wolff-Parkinson-White syndrome not reported in other patients to date.

**Conclusion:** We expand and further delineate the reported phenotype of patients with mosaic tetrasomy 5p resulting from a supernumerary isochromosome 5p. The main clinical features of this condition include developmental delay, seizures, hypotonia, dysmorphic features, and additional features. Tetrasomy 5p is typically found as a mosaic anomaly, and may be missed on routine testing (e.g. by microarray or standard karyotype on blood) because there is typically a lower level of mosaicism in peripheral blood. Genetic counselling is challenging because the phenotypic spectrum and clinical severity depend in part on the level and distribution of mosaicism.



## **PgmNr 1226: Phenotype-agnostic versus phenotype-driven analysis of exome (ES) or genome sequence (GS) data: Experience in a clinical site of the undiagnosed diseases network (UDN).**

### **Authors:**

V... Shashi <sup>1</sup>; K. Schoch <sup>1</sup>; R. Spillmann <sup>1</sup>; N. Walley <sup>1</sup>; H. Cope <sup>1</sup>; K. Tan <sup>1</sup>; A. McConkie-Rosell <sup>1</sup>; Y.H. Jiang <sup>1</sup>; N. Stong <sup>2</sup>; D.B. Goldstein <sup>2</sup>; The Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Pediatrics, Division of Medical Genetics, Duke Univ Med Ctr, Durham, North Carolina.; 2) Institute for Genomic Medicine, Columbia University Medical Center, New York, NY

---

The UDN utilizes ES and GS in undiagnosed diseases. The sequencing core labs generate sequence data and analyze these. Each clinical site independently analyzes these as well. Variants from both pipelines are reconciled by the clinical site for outcome determination. This dual analysis of every patient's ES or GS is a unique feature of the UDN. We examined its utility at the Duke/Columbia clinical site.

The analytical pipeline at the Duke/Columbia site is phenotype-agnostic, but without a specific CNV caller. In contrast, the core labs integrate phenotype with sequence data and use a specific CNV caller. In 25 of 41 (60%) diagnosed cases, the core labs and the Duke/Columbia site were concordant in diagnostic variants. In five cases (12%) the core labs missed the diagnosis, including variants not reported due to poor phenotypic fit, missing a phenotypic expansion (n=2), missed mosaicism (n=1) and missed CNVs (n=2). The Duke/Columbia site did not diagnose 11 individuals (28%). These included missed CNVs (n=6), not reporting heterozygous variants in phenotypically relevant AR genes in which further targeted testing, revealed a second variant (n=2), missed variants in an alternate transcript (n=2) and not reporting an AD inherited variant (n=1).

Sub-analysis showed that the total number of reported variants (including diagnostic) across both pipelines were  $23.7 \pm 7.8$  with only  $6.1 \pm 5.0$  variants overlapping between the two. These differences were because the research variants reported by the core labs ( $15.5 \pm 6.8$ ) were in genes of interest to the phenotype and the clinical site reported research variants ( $11.1 \pm 6.8$ ) that were bioinformatically prioritized. Of interest, among undiagnosed patients, only the clinical site detected candidate genes (n=4) that are being pursued. The core labs did not report these due to limited information about the genes.

Running two distinct complementary pipelines for ES or GS analysis enhances diagnosis. A phenotype-agnostic pipeline prioritizes variants that can result in phenotypic expansion of a known disorder, as well as new candidate genes. A phenotype-driven pipeline can flag variants in phenotypically relevant genes, leading to targeted investigations and diagnoses. CNV callers for GS also increase diagnostic capabilities. The disadvantage of a dual pipeline is the increased time and effort for personnel due to the larger number of variants. However, overall the dual analyses provide an improvement in



outcomes.

## PgmNr 1227: Expanding the spectrum of novel *ARID1B* mutations in Coffin-Siris syndrome.

### Authors:

E... Tan; S. Lian; A. Lai; E.S. Tan; T.W. Ting; B. Cham; H. Wei

[View Session](#) [Add to Schedule](#)

**Affiliation:** KK Women s & Children s Hospital, Singapore, Singapore

Mutations in the *ARID1B* (AT-rich interaction domain 1B) gene is recognized as the most common cause of Coffin-Siris syndrome and intellectual disability. The gene encodes a DNA-binding non-catalytic protein that is part of the human Brahma-associated factor (BAF) complex. It is found predominantly in neural tissue and thought to be essential for early brain development. To date, more than 170 pathogenic variants have been documented in the Human Gene Mutation Database but the variants in association with CSS are mostly from patients of European ancestry.

We used next-generation sequencing on genomic DNA extracted from venous blood samples from five unrelated patients (2 Chinese, 2 Indians, 1 Malay) who had developmental and speech delay, coarse facies and thick eyebrows. Four had hypertrichosis, low frontal hairline and prominent eyelashes, while only one had hypotonia and ACC. All were full-term births from non-consanguineous parents.

Using the TruSight One targeted panel, a heterozygous variant in the *ARID1B* gene was identified in each of the five patients, with no other pathogenic candidate variants in other genes. All are truncating variants: three were nonsense substitutions (NM\_020732.3 c.3304C>T, c.4216C>T, c.2881C>T) while the other two were deletions of 1 and 8 nucleotides (c.4272delG and c.1153\_1160del) which are predicted to result in frameshifts leading to premature termination of translation from a nonsense codon downstream. Four cases had DNA samples from both parents while the fifth only had the maternal sample. All five variants were not found in the corresponding parental samples tested. Only one variant (c.2881C>T) had been reported previously in three different studies. It is a nonsense mutation and had only one parental sample. There is no existing report in the scientific literature for the remaining four variants, which were also not found in the Genome Aggregation Database, dbSNP, Human Gene Mutation Database, ClinVar, and the Exome Sequencing Project (ESP) databases.

Our finding unveils *ARID1B* variants in association with CSS among multiple ethnic groups from three major ethnic groups in Southeast Asian populations, with phenotypic features highly resembling their Caucasian and Japanese counterparts. Our finding adds to the mutation spectrum of *ARID1B* and confirms that the variants resulting in CSS1 are mostly *de novo* and truncating.

## PgmNr 1228: Global metabolomics facilitates discovery of a novel disorder of the kynurenine pathway.

### Authors:

L.C. Burrage <sup>1</sup>; C.R. Ferreira <sup>2</sup>; N. Ehmke <sup>3</sup>; J.A. Rosenfeld <sup>1</sup>; H. Dai <sup>1</sup>; B. Lee <sup>1</sup>; S. Lalani <sup>1</sup>; S.H. Elsea <sup>1,4</sup>; Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, Texas, USA; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 3) Charité – Universitätsmedizin Berlin, Institute of Medical Genetics and Human Genetics, Berlin, Germany; 4) Baylor Genetics, Houston, TX, USA

---

Disorders of the kynurenine pathway, which cause NAD (nicotinamide adenine dinucleotide) deficiency, have been associated with multiple congenital malformations reminiscent of VACTERL association. Biallelic variants in *HAAO* and *KYNU*, which encode enzymes (3-hydroxyanthranilic acid 3,4 dioxygenase and kynureninase, respectively) in the tryptophan catabolic pathway and *de novo* NAD synthesis pathway, are associated with cardiac, renal, vertebral, and limb malformations. Embryonic deficiency of NAD is the suspected mechanism underlying these congenital malformation syndromes, and supplementation with niacin during pregnancy prevents the defects in a mouse model of this disorder. We tested whether plasma and urine global metabolomic profiling (Global MAPS™) identifies defects in the kynurenine pathway using samples from two individuals with congenital cardiac defects, vertebral abnormalities, limb abnormalities and biallelic variants in *KYNU*. Marked elevations in metabolites upstream of the metabolic block were identified in urine (mean 3-hydroxykynurenine Z score = 8.03; mean xanthurenate Z score = 5.84; mean kynurenine Z-score = 2.8) and in plasma (mean Z score xanthurenate = 7.54). A reduction of metabolites downstream of the metabolic block was also confirmed in urine (mean picolinate Z score = -3.81; mean quinolinate Z score = -5.25). We have recently identified two siblings enrolled in the Undiagnosed Diseases Network with complex congenital cardiac defects, vertebral and rib anomalies, short stature, and small kidneys. Exome sequencing revealed that these siblings share biallelic variants in the gene encoding kynurenine 3-monooxygenase (*KMO*), which catalyzes the enzymatic step upstream of kynureninase. Global metabolomic profiling in plasma from one of these siblings revealed elevations in kynurenine (mean Z score = 6.35) and kynurenate (mean Z score = 8.84) and a reduction in quinolinate (mean Z score = -2.73), confirming that this deficiency in kynurenine 3-monooxygenase likely represents a novel defect in the kynurenine pathway that also causes embryonic NAD deficiency and complex congenital malformations. Overall, our studies demonstrate that global metabolomic profiling in plasma and urine may facilitate the diagnosis of kynurenine pathway defects and supports the discovery of a novel defect in this pathway. Future studies are necessary to test whether niacin rescues this congenital malformation syndrome caused by congenital NAD deficiency.

## PgmNr 1229: *De novo* missense variant in the kinetochore-associated *NUF2* gene in a male patient with intra-uterine growth retardation and microcephaly.

### Authors:

D. Uehara <sup>1</sup>; H. Mitsubuchi <sup>2</sup>; J. Inazawa <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 2) Department of Pediatrics, Graduate School of Medical Sciences, Kumamoto University, Kumamoto, Japan; 3) Bioresource Research Center, Tokyo Medical and Dental University, Tokyo, Japan

---

Mutations in proteins involved with cell division and chromosome segregation, such as microtubule-regulating, centrosomal and kinetochore-associated proteins, have been reported to be associated with microcephaly and/or intra-uterine growth retardation (IUGR). In particular, the kinetochore plays a critical role for mitosis and cell division by mediating interactions between chromosomal DNA and spindle microtubule to direct proper chromosome alignment and segregation. To date, a few genes encoding for proteins of the kinetochore complex have been identified as causes of syndromes that included microcephaly and/or IUGR, namely *CENPE*, *CASC5*, *CHAMP1* and *BUB1B*. Here we report a male patient detected with a rare *de novo* *NUF2* missense variant after trio whole-exome sequencing analysis. The patient was born with IUGR and microcephaly, with additional features including low-set ears, micrognathia and atrial septal defect. *NUF2*, located at 1q23.3, encodes a component of the outer kinetochore-associated NDC80 complex, a highly conserved complex crucial for proper microtubule binding and spindle assemble checkpoint. The mutated residue is located at the N-terminal domain superfamily of *NUF2*, which is known to interact with the N-terminus portion of the NDC80 protein. Analysis with a patient-derived lymphoblastoid cell line showed that *NUF2* is decreased at protein levels. Moreover, NDC80 is downregulated in lysates from patient-derived cells. The variant detected in the proband is predicted *in silico* to cause loss of hydrophobic interactions in the core of the protein, thereby the stability of the N-terminal interaction of *NUF2*-NDC80 might be impaired. Functional *in vitro* assays are being carried out in order to establish a causal role of the mutation in the patient's phenotype. *NUF2* might be a candidate for IUGR/microcephaly, and if so, it would be the first component of the NDC80 complex to be implicated with such features.

## PgmNr 1230: Genome sequencing analysis of a family with a child displaying severe abdominal distention and recurrent hypoglycemia.

**Authors:**

J.L. Liu; G.L. Ding; M.F. Wangler; Z. Sun

[View Session](#) [Add to Schedule](#)

**Affiliation:** Baylor College of Medicine, Houston, TX.

---

We report a boy who developed severe progressive abdominal distention due to a dramatic adipose mass from the age of 7 months. He developed recurrent hypoglycemia that led to seizures at the age of 4 years. Imaging studies supported tumor development from the abdominal white adipose tissue. We performed whole-genome sequencing using blood DNA from the child and his parents. The child inherited a single-nucleotide deletion NM\_000314.6:c.849delA (p.Glu284Argfs) in the tumor suppressor gene *PTEN* (phosphatase and tensin homolog) from his father. The paternal family members have a history of cancers. It is conceivable that *PTEN* loss-of-function induced the adipose tumor growth and hypoglycemia, although the proband did not meet the usual diagnosis criteria of Cowden syndrome or Bannayan-Riley-Ruvalcaba syndrome that are characterized by germline mutations of *PTEN*. This case underlines the variability of phenotypes associated with *PTEN* germline mutations and provides useful information for diagnosis and genetic counseling of *PTEN*-related diseases for pediatric patients.

# PgmNr 1231: Knockdown of *pgap3* in zebrafish uncovers an early neural tube phenotype associated with hyperphosphatasia and mental retardation syndrome in humans.

## Authors:

S. Da'as <sup>1</sup>; W. Aamer <sup>1</sup>; W. Hasan <sup>1</sup>; K. Zamel <sup>1</sup>; M. Stotland <sup>1</sup>; K. Fakhro <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Sidra Medicine, Doha, Qatar; 2) Hamad Bin Khalifa University, Doha, Qatar; 3) Weill Cornell Medical College, Doha, Qatar

---

Recessive mutations in Post-GPI attachment to proteins 3 (*PGAP3*), encoding glycosylphosphatidylinositol (GPI)-specific phospholipase protein cause the rare neurologic disorder hyperphosphatasia with mental retardation syndrome 4 (HPMRS4). HPMRS4 is characterized by severely delayed psychomotor development, seizures, hyptonia and dysmorphic facial features. Here, we report a novel homozygous *PGAP3* nonsense mutation, c.265C>T - p. Gln89\*, in a 3-year-old boy, presenting with global developmental delay, speech impairment, seizures, hypotonia, dysmorphic features, cleft palate, low sets of ears, partial agenesis of corpus callosum and elevated alkaline phosphatase serum levels. Given that complete loss of GPI anchors in mouse cause embryonic lethality, we examined the functional consequences of transient knockdown of *pgap3* by morpholino in developing zebrafish embryos. Morphants displayed a HPMRS4-like features of developmental delay, brain abnormalities, dysmorphic head features, hyptonia and seizure. Notably, we observed defective neural tube formation during early stages of development, a phenotype never documented before in humans, which may play a yet uncharacterized role of *pgap3* pathogenesis at early stages of nervous system development. Significantly, *pgap3* loss disrupted midbrain and hindbrain formation by creating a separation of the left and right tectal ventricles, and defects in cerebellar corpus and caudal hindbrain formation. Finally, assessment of neuromuscular response revealed epileptic-like movements during the initial phases of development, followed by seizure-like symptoms, loss of touch response and hypotonia, all mimicking human patients' clinical phenotypes. In conclusion, we investigated *pgap3* knockdown genetic model in a vertebrate species and reveal new and important roles of this gene that may have during the nervous system development and wiring of the brain.

**Keywords:** Hyperphosphatasia mental retardation syndrome 4 (HPMRS4), Post-GPI attachment to proteins 3 (PGAP3), zebrafish model, neural tube defect, corpus callosum.

## PgmNr 1232: Characterization of phenotype - genotype of a cohort of patients with Fanconi anemia in Colombia.

### Authors:

A. Paredes Brijaldo <sup>1</sup>; F. Suarez Obando <sup>1,2</sup>; A. Rojas <sup>1</sup>; I. De los Reyes <sup>2,3</sup>; E. Rueda <sup>4</sup>; D. Medina <sup>5</sup>; J. Bernal <sup>1</sup>; J. Prieto <sup>1</sup>; M. Peiró <sup>6</sup>; R. Pieró <sup>6</sup>; M. Bogliolo <sup>6</sup>; J. Benítez <sup>7</sup>; J. Surrallés <sup>6</sup>; O. Moreno Niño <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine Department, Pontificia Universidad Javeriana, Bogotá D.C., Colombia; 2) San Ignacio University Hospital; 3) Central Military Hospital; 4) University Hospital of Santander; 5) Valle del Lili Foundation; 6) Department of Genetics and Microbiology, Autonomous University of Barcelona; 7) CNIO: National Center for Oncological Research, Madrid Spain.

---

**Introduction:** Fanconi anemia is a hereditary disease that presents with congenital malformations and affect different system; has bone marrow failure and predisposition to develop malignant tumors, which affect the patient's prognosis. The diagnosis is based on clinical suspicion and confirmation in genetic studies, where the chromosomal fragility test and sequencing of genes through next-generation sequencing (NGS) are preferred. Clinical expressivity, genetic heterogeneity and limited information in Colombia are factors that make it difficult to diagnose.

**Objective:** Characterization of phenotype-Genotype of a cohort of individuals diagnosed with Fanconi anemia by cytogenetic test and molecular study.

**Methodology:** Patients were recruited from several hospitals in the country and signed a consent to participate in the study. They underwent the cytogenetic test of hypersensitivity to the crosslinking agent (diepoxybutane). In all positive cases, MLPA was performed for the FANCA gene and/or analysis of FANC genes by Next-generation sequencing. We took the clinical and paraclinical data from the medical history, and patients were cited to complete physical examination.

**Results:** 191 individuals were analyzed by DEB sensitivity test; 42 had a positive sensitivity test. 19 patients were positive for a pathogenic variant in FANCA gene and 1 in FANCD2. In the cohort 24 men (57%) and 18 women (43%) were found, with an average age of 9.3 years, the age range was 10 months to 57 years. 26% of the patients were children of consanguineous parents. The most common clinical feature was hematological involvement, followed by alterations in the skeletal system, mainly in the thumbs; in the skin with hyper and hypopigmented lesions and anomalies in the facial phenotype. The age of onset of the hematological involvement was at 6.8 years, started with a commitment of a single cell line and progression to bone marrow failure.

**Conclusion:** The cohort studied with Fanconi anemia presents similar characteristics to those reported in the literature worldwide. We didn't have a phenotypic characterization of our population. Cytogenetic and molecular studies allow to confirm diagnoses of diseases with shared symptoms, facilitated the development of knowledge of the phenotype-genotype relationship, as well as that of physiopathology, which drives new research for the development of new drugs, which impacts in morbidity and mortality.

## PgmNr 1233: Linked-read whole genome sequencing identifies biallelic mutations in *DONSON* as a novel cause of Meier-Gorlin syndrome.

### Authors:

L.S. Bicknell <sup>1</sup>; K.M. Knapp <sup>1</sup>; R. Sullivan <sup>1</sup>; J. Murray <sup>2</sup>; G. Gimenez <sup>1</sup>; P. Arn <sup>3</sup>; P. D'Souza <sup>4</sup>; A. Gezdirici <sup>5</sup>; W. Wilson <sup>6</sup>; A.P. Jackson <sup>2</sup>; C.R. Ferreira <sup>7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology, University of Otago, Dunedin, Otago, New Zealand; 2) MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom.; 3) The Nemours Children's Clinic, Jacksonville, Florida, USA.; 4) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland.; 5) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, 34303, Istanbul, Turkey.; 6) Department of Pediatrics, University of Virginia School of Medicine, Charlottesville, Virginia, USA.; 7) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA.

---

Linked-read whole genome sequencing presents a new opportunity for cost-efficient singleton sequencing in place of traditional trio-based designs, while generating informative phased variants – effective for recessive disorders when parental DNA is unavailable. We have applied linked-read WGS to identify novel causes of Meier-Gorlin syndrome (MGORS), a condition recognised by short stature, microtia and patella hypo/aplasia. There are eight genes associated with MGORS to date, all encode essential components involved in establishing and initiating DNA replication. Our successful phasing of linked-read data led to the identification of biallelic rare variants in four individuals (24% of cohort) in *DONSON*, a recently established DNA replication fork surveillance factor. The variants include five novel missense variants and one deep intronic variant. All were demonstrated to be deleterious to function; the missense variants all disrupted the nuclear localisation of *DONSON*, while the intronic variant created a novel splice site which generated an out-of-frame transcript with no residual canonical transcript produced. Variants in *DONSON* have previously been associated with extreme microcephaly, short stature and limb anomalies (MISSLA) and perinatal lethal microcephaly-micromelia syndrome (MIMIS). Our novel genetic findings extend the complicated spectrum of phenotypes associated with *DONSON* variants and promote novel hypotheses for the role of *DONSON* in DNA replication. While our findings reiterate that MGORS is a disorder of DNA replication, the pathophysiology is obviously complex. This successful identification of a novel disease gene for MGORS highlights the utility of linked-read WGS as a successful technology to be considered in the genetic studies of recessive conditions.



## PgmNr 1234: 1p36 deletions: One cytogenetic band, multiple genetic syndromes and disease genes.

### Authors:

D.A. Scott <sup>1,2,3</sup>; T.M. Scott <sup>2,4</sup>; J.A. Rosenfeld <sup>1</sup>; S.R. Lalani <sup>1,2</sup>; B.J. Kim <sup>1</sup>; A. Hernandez-Garcia <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, TX; 2) Texas Children's Hospital, Houston, TX; 3) Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX; 4) Brigham Young University, Provo, UT

---

Deletions of chromosome 1p36 are the most common terminal deletions in humans and affect approximately 1 in 5000 newborns. 1p36 deletion syndrome is characterized by intellectual disability, a spectrum of functional and structural birth defects, and characteristic dysmorphic features. Distal and proximal critical regions for 1p36 deletion syndrome have been defined. Individuals with deletions that do not contain these critical regions do not have the characteristic features of 1p36 deletion syndrome and should be considered to have distinct genetic syndromes. The heterogeneous nature of 1p36 deletions, and an incomplete understanding of the role of individual 1p36 genes, makes it difficult for physicians to provide accurate prognostic information to families, and to create personalized medical plans for affected individuals. To overcome these obstacles, we have generated deletion/phenotype maps using clinical and molecular data from over 200 individuals with isolated 1p36 deletions. Using these maps, we have delineated non-overlapping 1p36 critical regions for seizures, sensorineural hearing loss, cleft lip/palate, congenital heart defects, cardiomyopathy, and renal anomalies. Many of these critical regions encompass genes that have been previously associated with these phenotypes in humans and/or mice. Using this information, physicians can readily identify individuals who are at risk for these phenotypes. We have also annotated the locations of 1p36 genes that have high loss-of-function intolerance scores (pLI > 0.8). Many of these genes have yet to be associated with a particular human phenotype. However, data from mouse models, and the clinical phenotypes of individuals with small deletions and/or loss-of-function variants suggest that haploinsufficiency of a subset of these dosage-sensitive genes may contribute to important 1p36 deletion phenotypes including intellectual disability (*SPEN*, *TP73*, *DNAJC11*, *CLSTN1*, *UBE4B*, *UBR4*, *WASF2*), seizures (*KCNAB2*), orofacial clefting (*SKI*), congenital heart defects (*CASZ1*, *SPEN*, *UBR4*, *ELOA*, *WASF2*), cardiomyopathy (*UBE4B*), and obesity (*WDFC1*). We conclude that many 1p36 deletions are associated with unique genetic syndromes whose phenotypic patterns are distinct from those described by 1p36 deletion syndrome. The distinctive phenotypic patterns of 1p36-related syndromes are caused by haploinsufficiency of dosage sensitive genes that can be identified using a combination of mouse models and human data.

## PgmNr 1235: Identification of a *TP63* pathogenic variant (c.797G>A) in a Mexican family with ADULT syndrome in the father and EEC syndrome in his son.

### Authors:

I.M. Rios-Flores <sup>1</sup>; J.C. Zenteno-Ruiz <sup>2</sup>; G. Rivas-Soto <sup>1</sup>; C.I. Aranda-Sanchez <sup>1</sup>; C. Peña-Padilla <sup>1</sup>; L. Bobadilla-Morales <sup>1</sup>; J.R. Corona-Rivera <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

- 1) Servicio de Genética, Hospital Civil de Guadalajara Juan I. Menchaca, Guadalajara, Jalisco, México;
- 2) Instituto de Oftalmología Conde de Valencia, Ciudad de México, México

---

*TP63* gene pathogenic variants (MIM 603273) causes the syndromes acro-dermato-ungual-lacrima-tooth (ADULT, MIM 103285), Rapp-Hodgkin (MIM 129400), ectrodactyly, ectodermal dysplasia, and cleft lip/palate (CLP) 3 (EEC3, MIM 604292), Hay-Wells (MIM 106260), limb-mammary (MIM 603543), orofacial cleft 8 (MIM 618149), and split-hand/foot malformation 4 (MIM 605289). Overlapping phenotypic features between these syndromes hamper the genotype-phenotype correlation in some families. Here, we report a Mexican family with ADULT syndrome in the father and EEC syndrome in his son due to a *TP63* mutation (c.797G>A). The father had only subtle features of ADULT syndrome, which included dental anomalies (agenesis of the right upper permanent lateral incisor, and a conical-shaped left upper lateral incisor), surgically treated nasolacrimal duct obstruction (NLDO) during infancy, and repetitive episodes of urinary retention until adolescence, with otherwise normal physical exam. His son showed tetraectrodactyly, chronic conjunctivitis, bilateral NLDO, lacrimal fistulae, and mild ectodermal signs (thin and sparse hair, brittle nails), indicative of EEC syndrome without CLP. In both, transpiration was referred as normal. By direct sequencing, the father and his son showed the same *TP63* pathogenic variant (c.797G>A). This family showed striking similarities with those included in the molecularly unconfirmed entry of “NLDO and cleft hands syndrome” in the London Medical Databases. The mutation found in our family has been previously reported in three families with EEC syndrome, and in another with ADULT syndrome. A common feature reported to this mutation is the micturition problems beginning in infancy and improving during adolescence, as in the father of our family. This report confirms that the c.797G>A *TP63* mutation can produce variable and subtle intrafamilial manifestations of the ADULT (lacrimal, tooth anomalies and urinary retention) and EEC3 (ectrodactyly and ectodermal manifestations) syndromes, and probably, also explains the proposed entity previously denominated as “NLDO and cleft hand syndrome”.

## PgmNr 1236: Functional characterization of lymphatic anomaly-associated, ERK-activating mutations in primary lymphatic endothelial cells.

### Authors:

M.E. March <sup>1</sup>; D. Li <sup>1</sup>; L.S. Matsuoka <sup>1</sup>; R. Pellegrino da Silva <sup>1</sup>; J. Garifallou <sup>1</sup>; C. Vaccaro <sup>1</sup>; J. Smiler <sup>1</sup>; C. Kao <sup>1</sup>; H. Hakonarson <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Central conducting lymphatic anomaly (CCLA) is a rare, congenital, and progressive disorder in development of lymphatic vessels and architecture. The standard therapy for lymphatic anomaly is rapamycin (Sirolimus). We recently published a report of a subject unresponsive to rapamycin, in whom we identified an activating mutation in *ARAF*. This mutation strongly activated ERK signaling when retrovirally transduced into lymphatic endothelial cells, causing morphological changes that were reversible through treatment with MEK inhibitors. The patient showed dramatic improvement upon treatment with a MEK inhibitor. As a follow up to this report, we are continuing analysis of mutations discovered in exomes of CCLA patients that fail to respond to rapamycin. Several mutations have been found in genes that influence the ERK signaling pathway. Preliminary results from analysis of mutations in the *KRAS*, *BRAF*, and *RAF1* genes indicate activation of ERK signaling, changes in cell morphology, and reversibility with inhibitors of the ERK pathway, consistent with observations with the *ARAF* mutation. We have also observed a lack of activation of signaling pathways downstream of mTOR, possibly providing a molecular explanation for the failure to respond to rapamycin. Our results suggest the existence of a category of CCLA patients that would benefit from treatment with inhibitors of ERK signaling.

## **PgmNr 1237: Bifid nose with or without anorectal and renal anomalies (BNAR) syndrome: Congenital heart disease belongs to the clinical spectrum.**

### **Authors:**

E. Brischoux-Boucher <sup>1</sup>; F.S. Alkuraya <sup>2</sup>; C. Gronier <sup>3</sup>; F. Nobili <sup>4</sup>; E. Marcoux <sup>5</sup>; P. Saugier-veber <sup>6</sup>; T. Schober <sup>7</sup>; L. Van Maldergem <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Centre de génétique humaine, CHU Saint Jacques, Besançon, Besancon, France; 2) Saudi Diagnostic Laboratories, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia; 3) Unité de cardiologie congénitale, University Hospital, Strasbourg, France; 4) Service de néphrologie pédiatrique, CHU, Université de Franche-Comté, Besançon, France; 5) Service de pédiatrie, Hôpital Nord Franche-Comté, Trévenans, France; 6) Department of Genetics , University Hospital, Rouen, France; 7) Department of Pediatrics, Dr. von Hauner Children's Hospital, Munich, Germany

---

*FREM1*-related autosomal recessive disorders include Manitoba oculotrichoanal (MOTA) syndrome, bifid nose with or without anorectal and renal anomalies (BNAR) syndrome and isolated congenital anomalies of kidney and urinary tract (CAKUT). It is a very rare group of disorders with only a few cases reported so far. To date, the authors are aware of 10 published individuals presenting as BNAR syndrome (Al-Gazali, 2002; Alazami, 2009). All of them belong to three consanguineous families of Egyptian, Afghani, and Pakistani ancestry. We report here 2 additional cases in siblings born to Turkish consanguineous parents. Both have the characteristic facial appearance with a bifid nose, unilateral renal agenesis in one of them, and no anorectal anomalies. Interestingly, one of them, a boy has a hitherto unreported congenital heart disease, namely Ebstein malformation. A homozygous *FREM1* exons 18-30 deletion has been identified in brother and sister. By virtue of so-called congruence, two additional autosomal recessive conditions segregate independently in this endogamic family: a nephrotic syndrome due to biallelic *ADCK4* mutation in four first cousins and autoimmune polyendocrinopathy syndrome type I (APECED) due to biallelic *AIRE* variant in two other cousins. Apart from expanding the clinical spectrum of BNAR syndrome, the current family report illustrates the burden of autosomal recessive conditions in endogamic populations.

## PgmNr 1238: A genetic two-hit mechanism drives vascular malformation in hereditary hemorrhagic telangiectasia.

### Authors:

D.A. Snellings<sup>1</sup>; C.J. Gallione<sup>1</sup>; D.S. Clark<sup>2,3</sup>; T. Ribar<sup>4</sup>; A. Fletcher<sup>4</sup>; X. Chen<sup>4</sup>; S. McCall<sup>4</sup>; N.T. Vozoris<sup>2,3</sup>; M.E. Faughnan<sup>2,3</sup>; D.A. Marchuk<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Genetics and Microbiology, Duke University, Durham, NC, USA; 2) Toronto HHT Centre, St. Michael's Hospital and Li Ka Shing Knowledge Institute, Toronto, Canada; 3) Division of Respiriology, Department of Medicine, University of Toronto, Toronto, Canada; 4) BioRepository & Precision Pathology Center, Duke University, Durham, NC, USA

---

Hereditary Hemorrhagic Telangiectasia (HHT) is a Mendelian disease characterized by vascular malformations including visceral arteriovenous malformations and mucosal telangiectasia. HHT is caused by loss-of-function (LOF) mutations in one of 3 genes; *ENG*, *ACVRL1* or *SMAD4* and is inherited as an autosomal dominant condition. Intriguingly, the constitutional mutation causing HHT is present throughout the body, yet the multiple vascular malformations in HHT patients occur focally, rather than manifesting as a systemic vascular defect. This disconnect between genotype and phenotype suggests that a local event is necessary for the development of vascular malformations. We investigated the hypothesis that one trigger for HHT-associated vascular malformations is somatic mutation resulting in biallelic loss of *ENG* or *ACVRL1*. We performed targeted deep sequencing on 20 punch biopsies of telangiectasia, followed by amplicon resequencing, to identify and validate LOF somatic mutations. Somatic mutations were identified in 9/20 telangiectasia with allele frequency ranging from 0.5 to 9%; 5 in *ENG* and 4 in *ACVRL1*. The mutations consisted of 5 indels causing frameshifts, 3 in-frame deletions, and 1 variant predicted to affect splicing; all consistent with LOF. Notably, across our capture panel of 16 vascular malformation-related genes, somatic mutations were exclusively present in the same gene as the constitutional mutation. We established phase for 7 of 9 samples using long-read sequencing, which confirm that the germline and somatic mutations in all 7 samples exist in *trans* configuration; consistent with a genetic two-hit mechanism. Additionally, multiple telangiectasia removed from singular patients contain unique somatic mutations; providing evidence that telangiectasia do not arise from a metastasis-like mechanism but rather from independent mutation events. Although somatic mutations were only found in 9/20 samples, the technical limitations of targeted short-read sequencing preclude the discovery of large indels, mitotic recombination, or epigenetic alterations; any one of which would generate loss of heterozygosity. These common molecular events may seed vascular malformation in a significant portion of telangiectasia. These combined data suggest that biallelic loss of *ENG* or *ACVRL1* may be a required event in the development of telangiectasia, and that rather than haploinsufficiency, vascular malformations in HHT is caused by a Knudsonian two-hit mechanism.

## PgmNr 1239: An integrated genomic medicine program for Mendelian disease gene discovery in Qatar.

### Authors:

W. Aamer<sup>1</sup>; A. Al-Kurbi<sup>1</sup>; A. Al-Maraghi<sup>1</sup>; N. Elbashir<sup>1</sup>; K. Al-Malki<sup>1</sup>; E. Aliyev<sup>1</sup>; N. Syed<sup>2</sup>; S. Hussein<sup>1</sup>; M. Gallab<sup>1</sup>; S.S. Padmajeya<sup>1</sup>; S. Da'as<sup>3</sup>; A. Akil<sup>1</sup>; M. Stotland<sup>4</sup>; G. Glass<sup>4</sup>; K. Zamel<sup>5</sup>; N. Robertson<sup>6</sup>; K. Hussain<sup>7</sup>; M. Fawzi<sup>5</sup>; I. Janahi<sup>9</sup>; R. Benini<sup>5</sup>; K. Omer<sup>5</sup>; B. Al-Adba<sup>10</sup>; M. Kamal<sup>11</sup>; D. Thomas<sup>12</sup>; M.F. Janjua<sup>13</sup>; N. Alateeqi<sup>13</sup>; A. Imam<sup>14</sup>; Z. Hijazi<sup>8</sup>; M. Kambouris<sup>15</sup>; K. Fakhro<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, research branch, Sidra Medicine, Doha, Qatar; 2) Biomedical informatics section, research branch, Sidra Medicine, Doha, Qatar; 3) Zebrafish core facility, research branch, Sidra Medicine, Doha, Qatar; 4) Plastic/Craniofacial surgery section, Sidra Medicine, Doha, Qatar; 5) Neurology unit, Sidra Medicine, Doha, Qatar; 6) Neonatology unit, Sidra Medicine, Doha, Qatar; 7) Endocrinology section, Sidra Medicine, Doha, Qatar; 8) Pediatric Medicine division, Sidra Medicine, Doha, Qatar; 9) Pediatric Pulmonology division, Sidra Medicine, Doha, Qatar; 10) Rheumatology unit, Sidra Medicine, Doha, Qatar; 11) Adolescent medicine unit, Sidra Medicine, Doha, Qatar; 12) Dermatology unit, Sidra Medicine, Doha, Qatar; 13) Developmental pediatric section, Sidra Medicine, Doha, Qatar; 14) Nephrology unit, Sidra Medicine, Doha, Qatar; 15) Pathology genetics division, Sidra Medicine, Doha, Qatar

---

The advent of next generation sequencing has a tremendous impact on the study of rare mendelian diseases, which a large proportion has no defined genetic cause. Furthermore, opportunities emerge in populations that experienced long history of genetic and cultural isolation, where the successive accumulation of recessive traits provides favorable ground for studying rare monogenic disorders. As many of the Middle-Eastern populations fit into these criteria, we have established an integrated clinical and research program that focuses on recruiting families of cases with undiagnosed phenotypes or uncommon clinical presentations, including structural deformities and developmental disorders. The program utilizes the power of whole genome sequencing technology and a computational pipeline with the capacity of identifying, both, single nucleotide variants (SNVs) and copy number/structural variants (CNVs/SVs). The potential variants then undergo a prioritization scheme to narrow down the list of candidate genes, which is further prioritized, either by evidence from medical literature, or experimentally validated through functional studies. This collaborative effort of clinicians, geneticists, bioinformaticians, and experimental biologists, so far, has recruited fifty-five families (264 subjects), forty-five (~80%) of which are already sequenced and analyzed. Among those analyzed, a known causative gene was identified in eight families (18%) and in another two (4%), novel genes are being verified through functional studies in zebrafish. The translational impact and the multidisciplinary nature of the program offers ample opportunity for collaboration, education of junior clinicians and biologists, and ultimately it establishes Sidra Medicine as a regional hub of genomic medicine.

## PgmNr 1240: Nucleolar alterations and premature senescence in *POLR3A*-mutated Wiedemann-Rautenstrauch syndrome fibroblasts.

### Authors:

G. Arboleda <sup>1,2</sup>; E. Valencia-Rincón <sup>2</sup>; C.T. Báez-Becerra <sup>2</sup>; N. Ramírez <sup>2</sup>; C. Guevara <sup>2</sup>; A. Sandoval-Hernandez <sup>2,3</sup>; C.E. Arboleda-Bustos <sup>2</sup>; L. Oliveros-Cisneros <sup>4</sup>; G. Gutierrez-Ospina <sup>4</sup>; H. Arboleda <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Instituto de Genética y Departamento de Patología, Facultad de Medicina, Universidad Nacional de Colombia, Bogota, Colombia; 2) Grupo de Neurociencias, Instituto de Genética, Universidad Nacional de Colombia, Bogotá, Colombia; 3) Departamento de Química, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia; 4) Departamento de Biología Celular y Fisiología, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Mexico City, México

---

**Introduction:** Recently mutations in *POLR3A* have been described as the cause of a rare neonatal progeria known as Wiedemann-Rautenstrauch syndrome (WRS). *POLR3A* is one of the two catalytic subunits of RNA polymerase III, which have important roles in the regulation of transcription of small RNAs, including transference RNA (tRNA), the 5S subunit of ribosomal RNA (5S rRNA), U6 small nuclear RNA (U6 snRNA), between others. However, the contribution of *POLR3A* mutations to cellular senescence and aging remains largely unknown. **Aim:** to describe diverse cellular features of fibroblast derived from one WRS patients and compared it with control fibroblasts. **Materials and methods:** Cultures of primary fibroblasts from one WRS patient [monoallelic *POLR3A* variant c.3772\_3773delCT (p.Leu1258Glyfs\*12)] (Paolacci et al., 2018) and one control were grown. Senescence was analysed by B-galactosidase and P16 expression. Morphology and volumetric analysis of nucleus (lamin A and hoescht staining) and nucleolus (fibrillarin) were analysed by immunostaining using confocal microscopy. Expression of proteins (*POLR3A*, Lamin A, P53, H2A) was evaluated by Western blot and inmunohistochemistry. **Results:** WRS fibroblasts become senescent (increased expression of beta/galactosidase, P16 and P53) in earlier culture passages as compared to control cells, and not associated with changes in nuclear morphology (nuclear blebs and progerin expression) as observed in HGPS cells. However, WRS nuclei are significantly larger. Mutation in *POLR3A* in the present case is associated to deletion of 5 amino acids, which causes changes in the expression of the protein. These changes were associated to an increase number of nucleoli and P53 expression. **Conclusions:** Mutations in *POLR3A* associated to WRS cause an early senescent phenotype in fibroblast associated to alterations in the number and structure/area of nucleolus. The present observations add to our understanding of the differences between HGPS and WRS, and opens new alternatives to study cell senesce and human aging.



## PgmNr 1241: Using suppressor tRNA to rescue protein production in rare genetic diseases caused by nonsense mutations.

### Authors:

K. Plona <sup>1</sup>; T. Sweet <sup>2</sup>; J. Collier <sup>2</sup>; M. Drumm <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio.; 2) Center for RNA Science and Therapeutics, Case Western Reserve University, Cleveland, Ohio.

---

Genetic diseases can be caused by a variety of mutation types including missense, nonsense, frameshift, and splice site alterations. Of these, nonsense mutations account for about 11% of all inherited diseases and are particularly hard to treat due to absence of protein for the affected gene allele. We have tested a nonsense suppressor tRNA for its ability to rescue protein production for a single base pair substitution in the *CLDN1* gene [MIM 603718] which is reported to cause the ultra-rare Mendelian disease neonatal ichthyosis-sclerosing cholangitis syndrome (NISCH [MIM 607626]), and are evaluating its potential in other rare diseases.

Nonsense suppressor tRNA's are modified transfer RNA's that can incorporate an amino acid at a stop codon. Our construct causes an amino acid residue to be inserted at UGA stop codons, suppressing termination and allowing full-length protein to be produced. To test *CLDN1* protein production, transgenes encoding a fusion protein of *CLDN1* with mCherry were expressed in liver (HepG2) and intestinal (Caco2) cell models. Two versions, wild-type *CLDN1* and *CLDN1* harboring a premature stop codon (c.192 C>A, p.C64X), were compared along with negative controls. The transgenes were expressed with and without suppressor tRNA treatment and assessed for mCherry fluorescence and cellular localization using a high content imaging system.

With nonsense suppressor tRNA treatment, the p.C64X mutant *CLDN1*-mcherry expression was rescued to 42.73% of wild type levels. Further analysis will be done to determine if the newly produced protein is functional as it now contains a different amino acid at position 64. These results indicate that suppressor tRNA therapy has potential to increase protein production in diseases caused by nonsense mutations. This work is currently being expanded to include other mutations and diseases including cystic fibrosis (CF [MIM 219700]) and glycogen storage disease type 1a (GSD1a [MIM 232200]). Additional work is being done to develop suppressor tRNA constructs that target other nonsense mutations and deliver different amino acids so that they can be applied to a broader range of nonsense mutations.



## PgmNr 1242: A homozygous TANGO2 deletion identified in a Brazilian patient.

### Authors:

Y. Lei <sup>1</sup>; T.O. Silva <sup>2</sup>; F.O. Poswar <sup>2,3</sup>; C.F.M. Souza <sup>2</sup>; I.V.D. Schwartz <sup>2,3,4</sup>; R.H. Finnell <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Cellular Biology, Baylor College of Medicine, Bellaire, Texas.; 2) Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; 3) Post-Graduation Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 4) Laboratory of Basic Research and Advanced Investigations in Neurosciences (BRAIN), Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil

---

The etiology for a significant fraction of recurrent rhabdomyolysis is still unclear, with a few known genetic causes. Using whole exome sequencing (WES), we identified a homozygous TANGO2 gene deletion in a Brazilian 8 years old girl. She presented recurrent episodes of flaccid tetraparesis and rhabdomyolysis with extreme hyperCKemia (highest known level 147,780 U/L) since 2 years of age, worsening with aging. She also presented global developmental delay, moderate intellectual disability and seizures. No prenatal and birth issues. She was born from a consanguineous couple from southern Brazil, with German ancestry. At age 7, she was evaluated by a medical genetics team. Initial workup was directed to the major differential diagnosis of episodic rhabdomyolysis (Fatty acid beta-oxidation disorders, Glycogen storage disorders), but the extreme hyperCKemia led to TANGO2-related disorder suspicion. Hypothyroidism and severe cardiac dysfunction due to arrhythmias, especially ventricular tachycardia, were also observed, establishing the clinical diagnosis, confirmed then by WES. PCR and SNP array data confirmed the previously reported exons 3-9 deletion in homozygosis. TANGO2-related disorder is a recently reported autosomal recessive condition, which phenotypic spectrum has been expanded since the first reports in 2016, by two independent groups. Several cases have been reported since then, most relying on next-generation sequencing (NGS) strategies. The approximately 34-kb deletion was observed in many cases, and the authors found a minor allele frequency (MAF) of 0.0026-0.13% in European populations and 0.26% in Hispanic/Latino population from control subjects databases. MAF in the Brazilian population is not available, but due to the high ethnic diversity, it is probably different from those previously stated. TANGO2 function and TANGO2-related disorder pathogenesis are not fully understood, although the TANGO family is predicted to contribute in secretory protein loading from Golgi to the endoplasmic reticulum; in mouse tissue culture cells, however, TANGO2 protein co-localizes with a mitochondrial protein. Probably the complex molecular genetics leads to the heterogeneous phenotype. To estimate the common deletion carriers frequency, a 1,000 healthy individuals from southern Brazil screening is ongoing.

## PgmNr 1243: A biallelic *TMEM251* variant in patients with severe skeletal dysplasia, coarse facial features, and early lethality.

### Authors:

N. Ain <sup>1,2</sup>; N. Muhammad <sup>1</sup>; M. Baroncelli <sup>3</sup>; M. Iqbal <sup>1</sup>; I. Bukhari <sup>1</sup>; O. Nilsson <sup>3,4</sup>; O. Makitie <sup>2,5</sup>; S. Naz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Biological Sciences., University of the Punjab, Lahore, Pakistan, Lahore, Punjab, Pakistan;  
2) Department of Molecular Medicine and Surgery and Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 3) Pediatric Endocrinology Unit and Center for Molecular Medicine, Department of Women's and Children's Health, Karolinska Institutet and University Hospital, Stockholm, Sweden; 4) School of Medical Sciences, Örebro University and Örebro University Hospital, Örebro, Sweden; 5) Children's Hospital, University and Helsinki University Hospital, Helsinki, Finland

---

Skeletal dysplasias are a clinically and genetically heterogeneous group of disorders ranging in severity from mild to lethal forms. Several metabolic disorders, such as lysosomal storage disorders and congenital disorders of glycosylation, include skeletal dysplasia as part of their clinical manifestations. We investigated a consanguineous family with five individuals affected by a recessively inherited progressive disease involving skeletal dysplasia similar to dysostosis multiplex and severe short stature (height SD score <-8.5). Two oldest affected individuals died during the course of this study before the age of 30 years while the others were less than twenty years old. The disorder was reminiscent of mucopolysaccharidosis and mucopolysaccharidosis, with affected individuals manifesting increasing coarsening of facial features, protruding abdomen, and progressive worsening of skeletal dysplasia with age. The patients had gradual loss of mobility, with the oldest being bed ridden. Using whole-exome sequencing we identified a biallelic missense variant affecting both isoforms of *TMEM251* which co-segregated with the disorder in the family. Variants in all other known genes for mucopolysaccharidosis and mucopolysaccharidosis were excluded by the sequencing data. *TMEM251* is an evolutionary conserved (0.69 Z score and 21.351 GDI score), ubiquitously expressed gene. The identified variant was predicted to be damaging by all used software, affected an amino acid which is conserved in diverse vertebrate species, and was absent from public databases and DNA of 190 ethnically matched controls. We determined that *TMEM251* is a transmembrane protein which localizes to the Golgi complex and plasma membranes in human osteosarcoma cell lines. However, mutant *TMEM251* protein with the identified variant was targeted less efficiently to the membranes and the localization was punctate as compared to the wild-type protein. In addition, *Tmem251* knockdown with small interfering RNA induced dedifferentiation of rat epiphyseal chondrocytes without significantly affecting apoptosis of these cells. Our findings implicate *TMEM251* in genetics of a new disorder and suggest a possible role of the encoded protein in differentiation of chondrocytes. Funded by Swedish Research Council, Sweden, Koshish Foundation USA & Higher Education Commission, Pakistan.

## PgmNr 1244: Detailed clinical description of hyaline fibromatosis syndrome patients in Japan.

### Authors:

H. Futagawa <sup>1</sup>; S. Ito <sup>2</sup>; M. Honda <sup>3</sup>; H. Yoshihashi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical genetics, Tokyo Metropolitan Children's Medical Center, Tokyo; 2) Nursing, Tokyo Metropolitan Children's Medical Center, Tokyo; 3) Nephrology, Tokyo Metropolitan Children's Medical Center, Tokyo

---

**Introduction** Hyaline Fibromatosis Syndrome (HFS; MIM#228600) is a rare autosomal recessive congenital disorder of the connective tissue caused by (the loss-of-function mutations) mutation in the anthrax toxin receptor-2 (*ANTXR2*) gene (located on chromosome 4q21). It is characterized by accumulation of the hyalinized fibrous tissue with cutaneous, mucosal, osteoarticular and systemic involvement including painful multiple joint stiffness and gingival hyperplasia. Although more than 70 cases have been reported worldwide, reports from Japan are scarce. Here, we report four cases of HFS presented to our hospital.

**Cases** The four Japanese subjects aged between eight months and eight years are presented here. Three cases were genetically diagnosed. Their birth weight ranged from 2546g to 3228g. All cases had failure to thrive (present body weight is between -3.1SD and -2.3SD), arthrogryposis multiplex congenita (multiple joint stiffness), skin nodules, gingival hyperplasia, trismus and bone lesions of X-ray as previously reported. In three cases, chronic pain due to physical stimulation, susceptibility to infection and curly nails were also observed. Interestingly, most of our cases (3/4) had pruritic skin nodules although they could not scratch because of arthrogryposis. Half of the subjects had chronic diarrhea. A single case diagnosed as protein losing gastroenteropathy died at three years of age due to multiple organ failure. None had intellectual disability or verbal communication problem.

**Discussion** In all cases, progressive arthrogryposis preceded the appearance of skin nodules and gingival hyperplasia and thus, compatible with HFS. The pruritis of skin nodules which the present cases suffered has not been previously reported and should be noted during the management. As presented, there should be more emphasis on deteriorating the living quality of subjects with HFS due to the pruritis.

**Conclusion** Recognizing the gradually progressive physical symptoms are essential in total health management of HFS.

## PgmNr 1245: Familial cases of DeSanto Shinawi syndrome with previously unreported nonsense and missense variants in *WAC* .

### Authors:

M. Shinawi <sup>1</sup>; J. Fleischer <sup>2</sup>; H. Andersson <sup>3</sup>; L. White <sup>1</sup>; S. Hughes <sup>4</sup>; T. Hsu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dep Pediatrics, Div Genetics & Genomic Medicine, Washington University School of Medicine, St. Louis, MO.; 2) Dep of Pediatrics, SIU Medicine, Springfield, IL.; 3) Hayward Genetics Clinic, Tulane University Medical School, New Orleans, LA.; 4) Pediatric Genetics, Children's Mercy Hospital, Kansas City, MO.

---

**Background:** DeSanto Shinawi (DESSH) syndrome is an autosomal dominant disorder caused by loss of function pathogenic variants in *WAC*. It is characterized by variable degree of developmental delay/intellectual disability (DD/ID), behavioral abnormalities, hypotonia, and recognizable facial features. Up until now, 20 patients were reported with *de novo* variants (10 nonsense, 7 frameshift, 1 splicing, 1 deletion) with 1 familial recurrence due to a presumed germline mosaicism.

**Methods:** Clinical and molecular characterization of two families with predicated pathogenic variants in *WAC* detected via exome sequencing. Clinical data were obtained by retrospective chart review and direct patient interaction with providers.

**Results:** The proband in Family I is a 6-year-old female who presented with DD, dysmorphic features, microcephaly and poor weight gain and later diagnosed with autism and ADHD. Exome sequencing revealed a maternally inherited nonsense variant in *WAC*: c.25C>T;p.Gln9\*. The proband's brother and maternal grandmother were carriers for same variant. All affected family members exhibited DD/ID, ADHD, autism or autistic traits and a few dysmorphic features. The proband in Family II is an 8-year old female who presented with ID/DD, ADHD, and autistic features. Her 29-year-old mother exhibited ID, ADHD, bipolar disorder and depression. The proband and her mother had eye refractive errors and shared a few dysmorphic facial features. Exome sequencing found a maternally inherited heterozygous missense variant in *WAC*: c.1738G>A;p.Glu580Lys. Glu580 is evolutionary conserved and the variant is absent from population databases and predicted to be deleterious based on *in silico* tools (CADD score of 41).

**Conclusion:** To our knowledge, these are the first reported cases of familial DESH syndrome and the first reported family with a missense variant in *WAC*. The patients exhibited similar neurobehavioral profile to previously reported case but facial features and lack of hypotonia were different. The implications of our results: 1) pathogenic variants in *WAC* can be inherited and therefore testing parents is recommended for genetic counseling purposes; 2) reduced penetrance for certain phenotypes and variable expressivity should be thoroughly discussed when families are counseled; 3) patients with DESH syndrome can be fertile and conceive children; 4) adult patients with DESH syndrome exhibit no signs or symptoms of developmental regression or neurodegeneration.

## PgmNr 1246: The challenge of recognizing phenotype-genotype correlations in prenatal exome sequencing.

### Authors:

T. Gao <sup>1</sup>; W. Guo <sup>2</sup>; X. Zhu <sup>2</sup>; J. Shen <sup>1</sup>; W. He <sup>1</sup>; T. Chiang <sup>1</sup>; X. Wang <sup>1</sup>; Y. Yang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) AiLife Diagnostics, Pearland, TEXAS; 2) Peking University Third Hospital, Beijing, China.

---

Whole exome sequencing (WES) has been used recently to identify genetic etiology of fetal structural abnormalities which other diagnostic methods failed to uncover. We report WES analyses of two fetuses in a family in which the mother had three adverse pregnancies. The first fetus was stillbirth at 24 weeks of gestation. The second had abnormal posture of both hands and feet, skin edema, polyhydramnios, and micrognathia at 27 weeks. The third fetus was similarly affected with fetal scalp edema, asymmetrical lower limbs, large hands and feet, and micrognathia at 22 weeks. WES analysis of the POC sample from the third fetus and the parents identified compound heterozygous variants c.512G>T (p.R171L) and c.511C>T (p.R171W) in the ALG3 gene. Sanger studies showed that the second fetus is compound heterozygous for the ALG3 variants too. ALG3 is associated with congenital glycation disorder (CDG-Id), an autosomal recessive disorder caused by glycoprotein biosynthesis defect resulting in multiple affected systems particularly nervous system. The two rare variants have not been previously reported in patients. Another variant affecting the same codon, c.512G>A (p.R171Q), was reported as disease causing in multiple patients; in addition, multiple in silico prediction programs indicate that the two variants in this study are possibly damaging. So far only about 10 patients have been reported to have CDG-Id in the literature. No prenatal phenotype of CDG-Id is available in the OMIM database, and the postnatal phenotype summarized in OMIM does not seem to overlap well with those of fetuses in this study. Further literature search identified two sisters affected prenatally with CDG-Id. The phenotypic features overlap well with those in our fetuses, including prenatal development delay, micrognathia, multiple skeletal abnormalities, skin edema, cerebellar hypoplasia and partial corpus callosum agenesis and pontocerebellar hypoplasia. Based on collective data, we determined that the two variants may be related to the prenatal phenotype. This case highlights the challenge of phenotype correlation in prenatal WES analysis due to lack of information of prenatal phenotype for genetic disorders. The current disease databases usually only include postnatal phenotypes. It would be beneficial to integrate both prenatal and postal phenotypes in such databases, which will greatly facilitate fetal WES analysis.

## PgmNr 1247: Glutamic acid residue in basic domain of *BHLHA9* regulates transcription efficiency.

### Authors:

A. Kar <sup>1,2</sup>; S. Phadke <sup>3</sup>; A. Dalal <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Diagnostics Division, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, Telangana State, India; 2) Graduate Studies, Manipal Academy of Higher Education, Manipal, India; 3) Department of Medical Genetics, Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India

---

Embryonic limb development is a highly co-ordinate and complex process involving growth and apoptosis in synchronization. Regulation of multiple genes through transcription and dosage dependent signals are involved in regulation of molecular and cellular activities during limb development. We reported c.220\_221delinsTT variant in *BHLHA9* responsible for Complex Camptosynpolydactyly (MIM: 607639) a complex type of hand (limb) malformation characterized by grossly malformed hands, digits arising from dorsum of hand, polydactyly and soft tissue syndactyly of toes. This is an only report on such hand malformation till date of study. *BHLHA9* located on chr17p13.3 encodes for bHLH (basic helix loop helix) transcription factor and previously associated with Split Hand and Foot Malformation with Long bone Deficiency 3 (SHFLD3; MIM : 612576) in cases of duplication and Mesoaxial Synostotic Syndactyly with Phalangeal reduction (MSSD; MIM: 609432) for missense variations. Phenotypically MSSD and Complex Camptosynpolydactyly are remarkable different from each other, even though our variants p.Glu74Leu and MSSD variants p.Asn71Asp, p.Arg73Pro, p.Arg75Leu are neighboring location in basic domain of *BHLHA9*. We report on importance of Glutamic acid residue in basic domain of *BHLHA9* in fine tuning of transcription activity. Glutamic acid at 74<sup>th</sup> position in *BHLHA9* is highly conserved across species and other bhlh transcription factors, is the only acidic amino acid present in basic DNA binding domain. *BHLHA9* known to form heterodimer with other transcription factors such as TCF3, TCF4 etc and regulates gene expression by binding to E-box sequences (diverse group of enhancer sequences which are binding site for bHLH transcription factors). Using luciferase assay we demonstrated that MSSD variants show reduced luciferase activity compared to wild type *BHLHA9* and Glu74Leu variant of Complex Camptosynpolydactyly shows elevated activity than wild type *BHLHA9*. Our study highlights the unique property of Glutamic acid residue in basic DNA binding domain in regulation of transcription. We propose that MSSD is caused by loss of function variants and Complex Camptosynpolydactyly is caused due to gain of function variant.

## PgmNr 1248: Clinical exome sequencing data reveals high diagnostic rates and new susceptibility genes for congenital diaphragmatic hernia plus (CDH+).

### Authors:

T.M. Scott <sup>1,2</sup>; I. Campbell <sup>3</sup>; S.R. Lalani <sup>2,4</sup>; C.A. Shaw <sup>4</sup>; J.A. Rosenfeld <sup>4</sup>; D.A. Scott <sup>2,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Brigham Young University, Provo, UT; 2) Texas Children's Hospital, Houston, TX; 3) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX

---

Congenital diaphragmatic hernia (CDH) is a life-threatening birth defect with an incidence of approximately 1 in 4000 newborns. In approximately 40-60% of cases, CDH occurs in conjunction with other non-hernia-related anomalies (CDH+). Although array-based copy number variant analyses are routinely ordered on individuals with CDH+, exome sequencing has not been universally adopted as a diagnostic test in this disorder. Here, we analyze a clinical database of approximately 12,000 cases to determine the diagnostic efficacy of exome sequencing in individuals with CDH+ and to identify new CDH susceptibility genes. In this database, exome sequencing provided a molecular diagnosis in 45 out of 82 CDH+ cases, yielding a diagnostic rate of 47%. When CDH co-occurs with cardiovascular malformations (CVMs), the mortality rate increases from 30 to 60%. In the subset of individuals with this life-threatening combination of defects, the diagnostic rate was also 47% (21/45). We identified multiple individuals with positive exome results who carried putatively pathogenic variants in *KMT2D* (n = 4 Kabuki syndrome 1), *EP300* (n = 2; Rubinstein-Taybi syndrome 2), and *ALG12* (n = 2; congenital disorder of glycosylation 1G). This suggests that these genes/genetic disorders are likely to be associated with an increased risk of developing CDH. We also identified individuals with putatively pathogenic variants in *ANKRD11*, *BRCA2*, *FOXC2*, *FOXP1*, *MED12*, *MCPH1*, *RASA1*, *SMARCA4*, *SMARCC2*, and *TCF12*, all of which have been previously associated with genetic syndromes in humans. Published case reports and/or CHD-specific pathogenicity scores generated using a validated machine learning algorithm provided additional evidence in support of their association with CDH. We conclude that exome sequencing can be used to identify a molecular diagnosis in a high percentage of individuals with CDH+. We have also identified several new genes/genetic disorders that may be associated with increased risk of developing CDH.



## PgmNr 1249: Case report: Compound heterozygous loss of function variants in *MYL9* in a case of megacystis-microcolon-intestinal hypoperistalsis syndrome (MMIHS).

### Authors:

J. Kandler <sup>1</sup>; E. Sklirou <sup>2</sup>; A. Woerner <sup>2</sup>; L. Walsh <sup>2</sup>; E. Cox <sup>1</sup>; Y. Xue <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Fulgent Genetics, Atlanta, GA; 2) Division of Medical Genetics, Department of Pediatrics, Children's Hospital of Pittsburgh, University of Pittsburgh Medical Center, Pittsburgh, PA

---

Megacystis-microcolon-intestinal hypoperistalsis syndrome (MMIHS), also referred to as “visceral myopathy”, is a severe early onset disorder characterized by impaired muscle contractility in the bladder and intestines. The most well-studied gene associated with cases of MMIHS is *ACTG2*. To date, four other genes have also been linked to MMIHS among children born to consanguineous parents, including *LMOD1*, *MYH11*, *MYLK*, and *MYL9*. Here we present the case of a two-year-old girl with a prenatal diagnosis of bilateral hydronephrosis at 20 weeks and congenital mydriasis (both of which have been previously observed among cases of MMIHS) and a suspected clinical diagnosis of MMIHS based on her presentation with megacystis, lack of urinary bladder peristalsis, and intestinal pseudo-obstruction. After initial sequencing and deletion/duplication analysis of the *ACTG2* was negative, further sequencing and deletion/duplication testing was performed on the *LMOD1*, *MYH11*, *MYLK*, and *MYL9* genes. Our laboratory identified two heterozygous loss of function variants in *MYL9*, an exon 4 deletion and a 9 base pair deletion that removes the canonical splicing donor site at exon 2 (NM\_006097.4:c.184+2\_184+10del). Subsequent parental testing confirmed these variants to be *in trans* in our patient. *MYL9* is located on chromosome 20 and encodes a regulatory myosin light chain protein that is necessary for normal cellular contractility. Of note, the patient's mother (who carries the exon 4 deletion) has a history of chronic constipation. Further, one of the patient's siblings was lost at 20 weeks gestation. The sibling's autopsy results showed a urethral valve anomaly and enlarged bladder, potentially consistent with early signs of MMIHS (the sibling's sample was not available for molecular testing). To our knowledge, only 1 other case of MMIHS has been linked to *MYL9* variants: a homozygous deletion encompassing exon 4 in *MYL9* which was discovered during whole exome sequencing, then confirmed by chromosomal microarray analysis and Sanger sequencing. Furthermore, we were unable to find any previous clinical reports of the splicing variant in our patient. Although this report of a second proband with genetically-confirmed *MYL9*-associated MMIHS does not yet move the gene-disease association beyond “limited” per ClinGen guidelines, we suggest *MYL9* be considered for inclusion on genetic testing panels for MMIHS and perhaps other myopathies.



## PgmNr 1250: Bi-allelic pathogenic variants in *COX6A2* cause a striated muscle-specific cytochrome oxidase deficiency.

### Authors:

A. Iida<sup>1</sup>; M. Inoue<sup>1</sup>; S. Uchino<sup>2</sup>; S. Noguchi<sup>1</sup>; S. Hayashi<sup>1</sup>; T. Takahashi<sup>3</sup>; K. Fujii<sup>4</sup>; H. Komaki<sup>1</sup>; E. Takeshita<sup>1</sup>; I. Nonaka<sup>1</sup>; Y. Okada<sup>5</sup>; T. Yoshizawa<sup>6</sup>; L. Van Lommel<sup>7</sup>; F. Schuit<sup>7</sup>; Y. Goto<sup>1</sup>; M. Mimaki<sup>2</sup>; I. Nishino<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan; 2) Teikyo University, Tokyo, Japan; 3) Saiseikai Utsunomiya Hospital, Tochigi, Japan; 4) Chiba University, Chiba, Japan; 5) Osaka University, Osaka, Japan; 6) Ritsumeikan University, Shiga, Japan; 7) KU Leuven, Leuven, Belgium,

---

Cytochrome oxidase (COX) deficiency is genetically and phenotypically heterogeneous disorder characterized by mitochondrial respiratory chain defect. Two unrelated Japanese individuals were suspected to have congenital myopathy but pathologically diagnosed with COX deficiency. Both patients presented limb and facial muscle weakness and hypotonia and one also had cardiomyopathy. Neither patient exhibited involvement from other organs. To isolate a causative gene for COX deficiency which presents with only striated muscle-specific symptoms, we performed an exome sequencing in the patients and identified bi-allelic missense variants in *COX6A2* including a homozygous c.117C>A (p.Ser39Arg), and a compound heterozygous c.117C>A (p.Ser39Arg) and c.127T>C (p.Cys43Arg). *COX6A2* is exclusively expressed in the skeletal muscle and heart. We found the specific reduction in complex IV activities among the enzymes in mitochondrial respiratory chain, complex IV activity in skeletal muscles from both individuals. Assembly of complex IV and its supercomplex formation were impaired in muscles. In muscles of *Cox6a2* knockout mice, both complex IV activities and assembly of complex IV were also diminished. In the patient with *COX6A2* pathogenic variants, the symptoms are theoretically restricted to be seen only in striated muscles due to its expression pattern. Therefore, we proposed that our patients had distinctive characteristics from conventional COX deficiency.

In conclusion, this study indicates that bi-allelic variants in *COX6A2*, as a novel causative gene for COX deficiency, cause a muscle-specific form of COX deficiency.

## PgmNr 1251: Biallelic novel variants in *SPEG* causing arthrogryposis and neuromuscular phenotype.

### Authors:

J. Fatih <sup>1</sup>; J. Punetha <sup>1</sup>; A. Gezdirici <sup>2</sup>; M. Tarnopolsky <sup>3</sup>; B. Tuysuz <sup>4</sup>; Y. Bayram <sup>5</sup>; Z.C. Akdemir <sup>1</sup>; L. Brady <sup>3</sup>; S.N. Jhangiani <sup>6</sup>; D.M. Muzny <sup>6</sup>; D. Pehlivan <sup>1,7</sup>; R.A. Gibbs <sup>6</sup>; J.E. Posey <sup>1</sup>; J.R. Lupski <sup>1,6,8,9</sup>; Baylor-Hopkins Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 2) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey; 3) Department of Pediatrics, McMaster University Medical Center, Hamilton, Ontario, Canada; 4) Department of Pediatric Genetics, Istanbul University-Cerrahpasa Medical Faculty, Istanbul, Turkey; 5) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA; 7) Section of Neurology, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA; 8) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA; 9) Texas Children's Hospital, Houston, Texas, USA

*SPEG* encodes striated muscle preferentially expressed protein kinase (*SPEG*), which plays an important role in myocyte cytoskeletal development. Biallelic rare *SPEG* missense, nonsense and frameshift variants have been previously reported in 9 individuals in association with centronuclear myopathies (CNM) and dilated cardiomyopathy (DCM). CNMs are a type of congenital myopathy characterized by myopathy, weakness, cardiomyopathy, and ophthalmoplegia in individuals with biallelic *SPEG* variants.

In our study, trio exome sequencing (ES) was performed on three families with probands clinically diagnosed with neuromuscular disorders. The first individual (BAB8532) is a 3 month old girl whose exome showed variants in three genes; a homozygous missense change (c.6971T>A:p.Ile2324Asn) in *SPEG*, a compound heterozygous variant in *MYOM3*, and a *de novo* potentially deleterious heterozygous variant in *CIT*. Phenotypically, she presented with muscle weakness and arthrogryposis. She also had DD/ID and ophthalmoplegia. The second individual (BAB9744) is a 22 year old male with compound heterozygous variants (c.[3941-5G>T];[8026+5G>A]) in *SPEG*. BAB9744 was diagnosed with neuromuscular weakness and DCM. The third individual (BAB10708) is a 4 year old male with muscle weakness and arthrogryposis. Exome sequencing results showed *de novo* variants (c.9575C>A:p.[Thr3192Asn]) in *SPEG* and *TPM2*, and additionally compound heterozygous variants in *ABCA7*; of note, no second variant in *SPEG* was identified in this case, suggesting that *SPEG* may be associated with both dominant and recessively inherited conditions. Though the second variant was not identified it could be due to the deep intronic state of the variant not covered by ES.

Here, we provide 3 additional families with deleterious *SPEG* variants which support *SPEG* as a neuromuscular disorder gene and provide evidence to suggest that *SPEG* may be associated with both dominant and recessive segregation of disease. Variants in *MYOM3* and *CIT* in BAB8532, and *TPM2* and *ABCA7* in BAB10708 contribute to the additional phenotypes presented in each individual. The major presenting features in these cases include muscle weakness (3/3), ophthalmoplegia (1/3), arthrogryposis (2/3) and cardiomyopathy (1/3). The variability in clinical presentation and observed

mode of inheritance illustrate the potential contribution of development of a *SPEG* allelic series to fully elucidate genotype-phenotype correlations.

## PgmNr 1252: VCP inhibitor as a potential therapy for VCP myopathy and ALS.

### Authors:

V. Kimonis; L. Weiss; C. Cheng; L. Ta; G. Bonesteele; T. Ton

[View Session](#) [Add to Schedule](#)

**Affiliation:** Div Gen, Dept Pediatrics, Univ CA Irvine, Orange, California.

**Background:** Valosin-containing protein (VCP) disease mutants cause inclusion body myopathy with early-onset Paget disease and frontotemporal dementia (IBMPFD) through a gain of function activity of the VCP gene. Studies in drosophila models as well as in patient fibroblasts indicate that VCP inhibitors improve the disease pathology. CB-5083 is a reversible and competitive specific inhibitor of AAA ATPase p97/VCP and has a documented safety, and is well tolerated in human studies for cancer treatment. Thus, we hypothesize that normalizing the gain of VCP function in patient myoblasts and VCP mutant mice will correct the disease phenotype.

**Objectives:** Correction of muscle and spinal cord pathology in vivo in the homozygous VCP<sup>R155H/R155H</sup> disease mouse model with VCP inhibitor CB-5083.

**Approach:** The VCP<sup>R155H/+</sup> knock in mouse model developed by our lab displays pathological features of human VCP-associated disease however is a slow model of the disease. The homozygous VCP<sup>R155H/R155H</sup> mouse (2-month old, both sexes) were divided into 2 randomly assigned groups of 12 mice each for daily gavage treatment with selected doses of CB-5083 (½ of the maximum tolerated dose) and vehicle group, for 2-4 months. Findings were compared with wild type mice.

**Results:** Homozygous VCP mice tolerated the CB-5083 treatment well with normal behavior and body weights. They also showed improvement of muscle strength when compared with vehicle group and wild type littermates.

**Conclusions:** VCP inclusion body myopathy with early-onset Paget disease and/or frontotemporal dementia (IBMPFD) is caused by a gain of function mutation of the VCP gene. Preliminary studies show that Homozygous VCP mice tolerated the CB-5083 treatment well with improvement of muscle strength when compared with littermates. If successful in patients with VCP disease, ALS and other sporadic IBM could be treated using this strategy.

## PgmNr 1253: Reaching the proper diagnosis of patients with muscular dystrophy through exome sequencing.

### Authors:

B. Azab <sup>1,3</sup>; S. Swedan <sup>2</sup>; Z. Dardas <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pathology, Microbiology and Forensic Medicine Department, The University of Jordan, Amman, Jordan; 2) Department of Medical Laboratory Sciences, Faculty of Applied Medical Sciences, Jordan University of Science and Technology, Irbid, Jordan; 3) Department of Human and Molecular Genetics, School of Medicine, Virginia Commonwealth University, Richmond, VA

---

**Background:** Muscular dystrophies (MDs) are a heterogeneous group of inherited disorders that are characterized by progressive skeletal muscle weakness and dystrophic changes on muscle biopsy. The broad genetic and clinical heterogeneity of MDs make the accurate diagnosis difficult via conventional genetic and clinical approaches. Exome sequencing (ES) is an efficient, time-saving, and cost-effective tool, enabling disease-causing variant (DCV) detection in affected individuals.

**Aim of the study:** we aimed to investigate the use of the ES for the accurate designation of MD types for 24 patients from 8 unrelated consanguineous Jordanian individuals.

**Methods:** Thorough clinical examination and ES were performed to identify the DCVs in affected individuals from each family, followed by segregation analysis if candidate variants in affected individuals and unaffected family members by Sanger sequencing.

**Results:** We achieved 8 remarkable diagnostics success rate of 75% (6 out of 8 families), identified novel pathogenic variants in DYSF gene, conformed the pathogenicity of recurrent variants and added new phenotypes in patients with these variants.

**Conclusion:** To our knowledge, this is the first study applying an ES-based comprehensive molecular diagnosis to Jordanian patients with MDs. Our findings confirmed that ES as a powerful approach for the diagnosis of MD patients. This efficient method of molecular diagnosis is crucial for guiding patient clinical care, genetic counseling and most importantly, paving the way for gene therapy in the future.

## PgmNr 1254: Compound heterozygous *TRAPPC11* mutations in a Chinese patient with limb girdle muscle dystrophy-18.

### Authors:

C. Bharatendu <sup>1</sup>; M.L. Tung <sup>2</sup>; B. Tan <sup>1</sup>; S. Tomar <sup>3</sup>; R. Sethi <sup>3</sup>; G. Tan <sup>3</sup>; J. Chai <sup>4</sup>; S.K.H. Tay <sup>3</sup>; P.S. Lai <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Neurology, Department of Medicine, National University Health System, Singapore 119228; 2) Haematology-Oncology, National University Cancer Institute of Singapore, National University Health System, Singapore 119074; 3) Department of Paediatrics, National University of Singapore, Singapore 119228; 4) Department of Neurology, National Neuroscience Institute, Singapore 308433

---

Limb-girdle muscular dystrophy (LGMD) is a heterogeneous group of neuromuscular disease characterized primarily by progressive proximal myopathy. Due to clinical and genetic heterogeneity, it is often challenging to diagnose subtypes due to phenotype overlap. Recently, *TRAPPC11* (Transport protein particle complex 11) gene mutations were reported to cause autosomal recessive limb-girdle muscular dystrophy-18 (LGMD2S). To date, there have been only 19 patients reported in the literature, and the molecular basis for the wide phenotypic heterogeneity remains poorly understood. In this case report, we describe the role of whole exome sequencing (WES) in identifying pathogenic mutations in *TRAPPC11* gene to resolve diagnosis for a Chinese patient who presented early onset symptoms of global developmental delay at 1 year of age, with gross motor delay being most prominent. Subsequently over time, he showed symptoms of proximal and progressive muscle weakness, early onset scoliosis as well as early cataracts. Serum creatine kinase level was elevated (10800 IU/L). There was no hepatic steatosis. Muscle biopsy revealed mild loss of dystrophin and patchy loss of  $\alpha$ -dystroglycan. Due to clinical overlap with dystrophinopathies, screening for abnormalities in *DMD* gene was performed at time of presentation but was found negative. Trios whole exome sequencing revealed two compound heterozygous variants in *TRAPPC11* gene in proband inherited in trans from parents. Both variants have been previously observed, one was reported as a pathogenic missense mutation in exon 26 c.2938G>A;(p.Gly980Arg), while the other was a nonsense mutation in exon 2 c.142C>T (p. Arg48Ter) with unspecified pathogenicity. While results of WES should be interpreted cautiously taking into account the complete clinical picture inclusive of biochemical and pathological information, it has proved useful in resolving many diagnostic odysseys for rare disorders. The genotype-phenotype correlate for our patient is consistent with LGMDR18 as further advanced symptoms of scoliosis and early cataracts are now observed in patient after the molecular diagnosis. This case further expands the ever-growing spectrum of limb-girdle muscular dystrophy associated with *TRAPPC11* mutations.

## PgmNr 1255: Congenital myopathy: Two candidate genes in a Malian family.

### Authors:

L. Cissé <sup>1</sup>; I. Pamanta <sup>2</sup>; S. Diarra <sup>1,3</sup>; S. Touré <sup>2</sup>; S. Diallo <sup>2</sup>; S. Diallo <sup>4</sup>; F. Kané <sup>1</sup>; T. Coulibaly <sup>1</sup>; S.H. Diallo <sup>4</sup>; H. Ba <sup>5</sup>; C.O. Guinto <sup>1</sup>; K. Fischbeck <sup>3</sup>; G. Landouré <sup>1,3</sup>; I.A. Cissé <sup>2</sup>; The H3Africa Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurology, Teaching Hospital of Point G, Bamako, Mali; 2) Service de Rhumatologie, CHU du Point "G", Bamako, Mali; 3) Neurogenetics Branch, NINDS, NIH, Bethesda, MD; 4) Service de Neurologie, CHU de Gabriel Touré, Bamako, Mali; 5) Service de Cardiologie, CHU de Gabriel Touré, Bamako, Mali

**Introduction:** Muscular dystrophies are heterogeneous disorders that are characterized by proximal weakness and progressive disability. Several entities have been reported, but cases with molecular diagnosis are scarce in sub-Saharan Africa.

**Objectives:** To characterize a family with congenital muscular disease and to identify the genetic defect causing the disease.

**Methodology:** Participants were enrolled after giving consent. Neurological examination was performed, and patients were seen by a cardiologist to determine cardiac involvement. DNA was extracted for genetic analysis. Whole exome sequencing using one patient, the parents and an unaffected sibling was performed. Putative variants were checked in SNP databases.

**Results:** This is a consanguineous family in which three affected siblings presented with congenital myopathy. Symptoms started with motor delay with walking at age 2. While the oldest patient never walked, the other two presented with a waddling gait that deteriorated progressively. On examination there was no cognitive impairment but they had generalized muscle atrophy and weakness more prominent in proximal limbs and skeletal deformities including scoliosis and thoracic kyphosis. The patients had hypotonic muscles with absent reflexes, joint hyperlaxity and ankyloses at the elbows and knees. In addition, they had dental crowding but no cardiac involvement. CK levels were normal. EMG showed myopathic features. Whole exome sequencing identified two homozygous variants in different genes. The first is in *C21ORF2*, a gene associated with autosomal recessive axial spondylometaphyseal dysplasia. The variant was not found in 178,834 alleles, including 15,500 African in gnomAD. The second variant is in the *MYOF* gene, and was not seen in SNP databases. *MYOF* is not associated with a disease but previous reports suggested it as candidate gene for myopathy. The first variant segregated with the disease in the family; sequencing for the second variant in the family is not yet complete.

**Conclusion:** Our study has identified two candidate genes for congenital myopathy. The unusual symptoms presented by the patients could indicate a novel entity or a variant of the disease previously described with mutations in *C21Orf2*. Further studies may shed light into these hypotheses.

## PgmNr 1256: A case of tuberous sclerosis complex with pathological mutations in intron of TSC2 gene.

### Authors:

G. Furukawa <sup>1,2</sup>; N. Ishihara <sup>2</sup>; M. Kato <sup>3</sup>; T. Kato <sup>3</sup>; H. Kurahashi <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Toyokawa City Hospital, Aichi, Japan; 2) Department of Pediatrics, Fujita Health University School of Medicine, Aichi, Japan; 3) Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Aichi, Japan

---

**Introduction:** Tuberous Sclerosis Complex (TSC) is a multi-system disorder, typically involving cutaneous manifestation, various tumors and neurological symptoms, such as epilepsy, cognitive deficit and autism. TSC is inherited in an autosomal dominant pattern and is caused by mutations in either TSC1 or TSC2 genes. We have previously reported that more than 70% of the patients in our cohort have nonsense or frameshift mutations in TSC2 gene. Here we report a case of intronic mutation in TSC2 gene with clinical features of definite TSC.

**Subject:** Subject is a 1-year-old boy who was born to non-consanguineous parents. His father has nonspecific cutaneous manifestation on his back, but cannot even diagnose with possible TSC. Subject had symptoms of cardiac rhabdomyoma, hypomelanotic macules, subependymal nodules and cortical tubers at birth, and was diagnosed as definite TSC. He developed focal epilepsy at 2 months of age, became West syndrome at 4 months of age and vigabatrin treatment was administrated. Subject and his parents were underwent for genetic testing after standard genetic counseling.

**Methods:** Genomic DNA was extracted from white blood cell, and made ready to use for next generation sequence library preparation. Trio-based target exome sequencing was performed using TruSight One sequencing panel (Illumina, Inc. San Diego, CA, USA). Annotation and filtering of variants were performed using Variant Studio software (Illumina). Putative candidate variants were validated using Sanger sequencing.

**Result:** We found heterozygous 8 bases deletion in 20th intron of TSC2 gene (c.2220+2\_2220+9delTACCATGG) in the subject, and validated with Sanger sequencing. According to the peak flow of the Sanger sequencing data, we found a same mutation with 40% mosaic ratio in his father. We predicted it splicing abnormalities due to loss of function of splicing donor site using Human Splicing Finder (<http://umd.be/HSF3/>).

**Discussion:** As our case has an intronic deletion, the effect of splicing due to the mutation is expected to be different for each organ. Both TSC1 and TSC2 are tumor suppressor genes that function according to the two-hit theory, and are autosomal dominant in patients' organs. They become autosomal recessive in cell units, therefore clinical features are considered to be diverse. It is a valuable case leading to the elucidation of the mechanism of TSC and the abnormality of splicing, hence it is necessary to advance to the analysis of cell units.



## PgmNr 1257: Clinical and neuropathological features of four familial and two sporadic cases of Neuronal Intranuclear Inclusion Disease (NIID).

### Authors:

H. Furuya <sup>1</sup>; Y. Miyamoto <sup>1</sup>; T. Furushima <sup>1</sup>; Y. Morita <sup>1</sup>; K. Furuta <sup>1</sup>; Y. Osaki <sup>1</sup>; H. Arahata <sup>2</sup>; N. Sasagasako <sup>2</sup>; T. Iwaki <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology, Kochi Medical School, Kochi University, Kochi, Japan; 2) Department of Neurology, Neuro-Muscular center, NHO Omuta hospital, Fukuoka, Japan; 3) Department of Neuropathology, Neurological Institute, Kyushu University, Fukuoka, Fukuoka, Japan

---

**Background :** Neuronal intranuclear inclusion disease (NIID)(MIM 603472) is a neurodegenerative disorder characterized by eosinophilic intranuclear inclusions in neuronal cells. Such inclusions are also found in non-neuronal cells. The clinical features and pathological findings in patients with NIID are highly varied.

**Objective:** Here, we present seven NIID cases. Two of them, autopsy was performed.

**Cases :** 5 cases are adult onset NIID (aNIID) and 2 cases are a juvenile onset (jNIID). Case 1 is diagnosed as familial, and case 2, 4 is sporadic adult onset Parkinson disease with dementia. Case 2 and 5 is diagnosed as sporadic mild Parkinsonism with dementia and severe psychosis. Case 6 and 7 is a sister of juvenile onset familial cases and diagnosed as oculopharyngeal muscular dystrophy with mental deterioration by muscle biopsy. Autopsy was performed after sudden death at age 42 and 43.

**Results :** In MRI imaging, diffuse cerebral atrophy was observed in all cases. High intensity of the cortico-medullary junction in diffusion-weighted imaging (DWI), which is characteristic findings in NIID, was seen in aNIID but not seen in two jNIID. In neurological examination, all 5 aNIID showed laterality or severely increased Pulmomenal Reflex (PMR). Ante-mortem diagnosis was made by identification of intranuclear inclusions with skin biopsy in five aNIID. Post-mortem diagnosis was made in two jNIID. In this case, acidophilic intranuclear inclusion body with positive p62 immune staining was observed, not only in neuron and glia cell, but all other organ tissues.

**Conclusion :** Intranuclear accumulation of abnormal proteins and/or dysfunction of protein degradation might underlie in the pathogenesis of NIID. However, NIID is considered as a heterogeneous disease entity and shows various clinical features. Especially, there are different pattern of neuroimaging examination between aNIID and jNIIS. Not only neurologist, but also psychiatrist and pediatrician should pay attention to this heterogeneous disorder.

## PgmNr 1258: VWA2: Novel autosomal recessive candidate gene in Alzheimer's disease (AD).

### Authors:

J. Hoogmartens<sup>1,2,3</sup>; S. Engelborghs<sup>2,3,4</sup>; R. Vandenberghe<sup>5,6</sup>; K. Peeters<sup>1,2,3</sup>; P.P. De Deyn<sup>2,3,4</sup>; R. Cacace<sup>1,2,3</sup>; C. Van Broeckhoven<sup>1,2,3</sup>; On behalf of the BELNEU consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurodegenerative Brain Diseases Group, Center for Molecular Neurology, VIB, Antwerp, Belgium; 2) Institute Born-Bunge, Antwerp, Belgium; 3) Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium; 4) Memory Clinic, Department of Neurology, Hospital Network Antwerp, Middelheim and Hoge Beuken, Antwerp, Belgium; 5) Department of Neurosciences, Faculty of Medicine, University of Leuven, Leuven, Belgium; 6) Laboratory of Cognitive Neurology, Department of Neurology, University hospital Leuven Gasthuisberg, Leuven, Belgium

---

**Background** Up to 90% of Alzheimer's disease (AD) patients remain genetically unexplained. As several AD families do not show clear autosomal dominant inheritance patterns, we investigated if recessive inheritance might explain part of the missing genetic etiology.

**Methods** Whole-genome sequencing (WGS) was performed for 19 Belgian unrelated AD patients (mean AAO:  $56.7 \pm 6.6$  years) without mutations in AD genes. We analyzed the WGS data searching for homozygous variants altering protein sequences with a minor allele frequency (MAF)  $\leq 5\%$  in available databases. Targeted resequencing of candidate variants/genes was performed using multiplex-amplification resequencing.

**Results** Among the 113 investigated variants, homozygous p.V366M in VWA2 was the most promising candidate variant, since p.V366M was also present homozygous in one extra AD patient, was absent in control individuals, has a CADD score of 26.8 and VWA2 is a brain-expressed gene. We performed full exonic resequencing of VWA2 in a Belgian AD cohort (n=1253) and identified one patient homozygous for p.R69M and two patients compound heterozygous for p.V366M/p.L715F. In the Belgian control cohort (n=802) we observed one person heterozygous for p.L715F/p.R69M. To replicate our findings, we resequenced VWA2 in a European AD cohort (n=814) and identified two patients homozygous for p.L715F and two patients compound heterozygous for p.L715F/p.V728L and p.D79G/p.V366M. Heterozygous p.V366M/p.L715F, present in the Belgian AD cohort, was also observed in one European patient and one European control individual. In the European control cohort (n= 423), we identified two Spanish control individuals homozygous for p.H456Y. Taken the high 3.7% allele frequency of p.H456Y in the Spanish subgroup of the European control cohort, compared to 0.6% in the gnomAD database, we expect that p.H456Y might be a population specific polymorphism.

**Conclusion** Our data suggest that VWA2 might contribute to AD risk in an autosomal recessive manner. VWA2 is present in extracellular exosomes known for their contribution to neuronal interaction. We are using long read sequencing to check whether in heterozygous VWA2 carriers the variants are located *cis* or *trans*. Additionally, we are studying expression of VWA2 in brain material to analyze the effect of the genetic variants in VWA2 on expression.



## PgmNr 1259: Mutations in SORL1 R904W as a pathogenic mechanism in the development of Alzheimer's disease in a multigenerational Colombian family.

### Authors:

J.A. Tejada Moreno <sup>1</sup>; C.A. Villegas Lanau <sup>1,2</sup>; L.S. Madrigal Zapata <sup>2</sup>; A.Y. Baena Pineda <sup>2</sup>; O. Campo Nieto <sup>1</sup>; A. Soto Ospina <sup>1</sup>; P. Araque Marín <sup>3</sup>; L. Rishishwar <sup>4,5</sup>; E.T. Norris <sup>4,5,6</sup>; A.T. Chande <sup>4,5,6</sup>; I.K. Jordan <sup>4,5,6</sup>; G.J. Bedoya Berrio <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Genetics Research Group, University of Antioquia, Medellin, Colombia; 2) Neuroscience Research Group, University of Antioquia, Medellin, Colombia; 3) Chemical formulations Research and Innovation Group, EIA University, Medellin, Colombia; 4) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, Georgia, USA; 5) PanAmerican Bioinformatics Institute, Cali, Valle del Cauca, Colombia; 6) School of Biological Sciences, Georgia Institute of Technology, Atlanta, Georgia, USA

---

Alzheimer's disease (AD) is a complex and heterogeneous neurodegenerative disease and it is the most common cause of dementia in the elderly. Dementia due to Alzheimer's disease is characterized by noticeable memory, thinking, and behavioral symptoms that impair a person's ability to function in daily life. The genetics (AD) is heterogeneous. Early or familial forms, (EOAD, onset before 65 years), exhibit an autosomal dominant inheritance pattern while late or sporadic forms (LOAD, onset after 65 years) exhibit a complex inheritance pattern. Both early onset and late onset forms have been shown a strong genetic component. In familial forms, highly penetrating causal mutations have been identified in three specific genes, APP, PSEN1, and PSEN2, supporting the major role of amyloid  $\beta$  peptide in the development of the disease. In sporadic forms genome-wide association studies (GWAS) have found more than 30 risk genes such as APOE, BIN1, CLU, ABCA7, CR1, PICALM, MS4A6A, CD33, MS4A4E, CD2AP and SORL1 among others that are involve in new pathways related to the disease, like lipid metabolism, immune system and synaptic functioning mechanisms. We used whole-exome sequencing (WES) to evaluate a Colombian multigenerational family spanning three generations with a diagnosis of EOAD and with an inheritance pattern suggestive of autosomal dominant without mutations in APP, PSEN1 (included the most prevalent mutation in the Antioquia's population, E280A) or PSEN2 genes. We identified possibly pathogenic rare variants in SORL1 gene (R904W) and MAPT gen (R163P) segregating in the family with other potential risk variants in APOE (C130R) and ABCA7 (A877T) genes suggesting an oligogenic inheritance in this family. Also, we build a structural model to try to determine the effect on the structure and function. SORL1 Arg904Trp produces a polarity change that favors hydrophobic interactions and produces a separation between the cell membrane and the mutated protein with respect to the wild type which can generate a structural change that can affect the receptor function and in this way destabilize the normal metabolism of the protein which could explain the development of the disease in this family.

## PgmNr 1260: A coexisting case of Eaton Lambert disease and Charcot Marie Tooth Type 20 syndrome.

### Authors:

L. Mora Barreto; M. Guerra Torres

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical Genetics, Pontificia Universidad Javeriana, Bogota, Colombia

---

Neuromuscular disorders comprise many different clinical conditions that involves the injury or dysfunction of peripheral nerves or muscles. They have different etiopathogenic mechanisms that can correlate with clinical symptoms in patients. Most of them has a genetic cause, but there can be other causes for example immunological etiology in some neuromuscular junction diseases. Because of the spread spectrum of symptoms that can be shared between them, diagnosis in neuromuscular diseases can become a challenge.

In this case report we present a 64-year-old female patient, with onset of symptoms at age of 54 years old, consistent in muscular weakness in lower limbs predominantly proximal, difficulty on climbing stairs, gait abnormality, spasticity, areflexia and dry mouth. Electromyography and neuroconductions were performed at that moment, reported as normal. Then clinical symptoms progressed with increase in muscular weakness, fatigue and difficulty standing, so further exams were performed. Rheumatologic, renal and hepatic profile were reported as normal. Also, enzymatic probe for Pompe disease were negative. CPK levels were lightly high. Three years later electromyography and neuroconductions were performed again, with abnormal result showing sensorimotor axonal polyneuropathy. Then the patient had worsening of symptoms and required wheelchair for ambulation during long walks. One year later she assisted to physiatry and started clinical therapies. A Single-Fiber EMG showed presynaptic neuromuscular junction disease. Repetitive stimuli test was consistent with Eaton Lambert disease. Due to inconsistent findings in electromyography's, molecular studies were considered. Clinical exome sequencing reported a heterozygote *missense* variant of uncertain significance in gene DYNC1H1: c.1976C>T p. (Thr659Met), not previously reported in the literature and classified by bioinformatic predictors as pathogenic, related to Charcot-Marie Tooth type 2o disease. It is possible that clinical and paraclinical findings are caused by coexistence of two neuromuscular diseases.

## **PgmNr 1261: A novel GTPBP2 splicing mutation in two siblings affected with microcephaly, generalized muscular atrophy, and hypotrichosis.**

### **Authors:**

I. Abdi Rad <sup>1,2</sup>; A. Vahabi <sup>1,2</sup>; E. Akbari-Azar <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Cellular and Molecular Research Center, Urmia University of Medical Sciences, WA. Iran; 2) Department of Medical Genetics; Urmia University of Medical Sciences, WA, Iran

---

One of the challenging pitfalls in the genetic counseling of patients with syndromic microcephaly is a definite clinical diagnosis to determine its etiology. Any unknown syndromic microcephaly necessitates piling up of clinical and molecular genetics data in the hope of getting a genotype-phenotype correlation. In this study, we identified a novel splice site mutation in *GTPBP2* gene by Whole Exome Sequencing in two siblings with microcephaly and progressive generalized muscular atrophy associated with hypotrichosis. This mutation is predicted to disrupt the highly conserved acceptor splice site of exon 4 in *GTPBP2* gene. We describe a syndromic microcephaly related to *GTPBP2* gene in two siblings with an autosomal recessive pattern of inheritance, and to the best of our knowledge, is the first report of *GTPBP2* related hypotrichotic microcephaly with generalized muscular dystrophy.

Keywords: Microcephaly, Muscular dystrophy, Hypotrichosis, *GTPBP2* gene

## PgmNr 1262: Pre-symptomatic training improves motor and memory defects in a mouse model of Rett syndrome.

### Authors:

N.P. Achilly<sup>1,2</sup>; H.Y. Zoghbi<sup>1,3,4,5,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Developmental Biology, Baylor College of Medicine, Houston, TX; 2) Medical Scientist Training Program, Baylor College of Medicine, Houston, TX; 3) Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 5) Department of Neuroscience, Baylor College of Medicine, Houston, TX; 6) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 7) Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX

---

Rett syndrome is a debilitating childhood neurological disease that affects 1:10,000 young girls. Patients typically reach their developmental milestones for the first 12-18 months before rapidly regressing. Over time, Rett syndrome robs patients of their motor, learning, and social skills. Although the genetic basis of Rett syndrome was discovered two decades ago, the precise molecular mechanism by which mutations in *MECP2* cause Rett syndrome remains elusive. Nevertheless, many labs are exploring ways to augment the function of MeCP2 and improve brain function in Rett syndrome. However, these experimental therapies are many years away from clinical use. Right now, patients and their families are desperate for effective therapies.

Work from our lab has shown that deep brain stimulation rescues aspects of the disease in mice. These results indicate that enhancing neural function is sufficient to overcome the defects caused by MeCP2 dysfunction. With this knowledge, we sought to determine whether prolonged training in Rett mice could provide a similar benefit. Extended motor training in symptomatic Rett mice resulted in a modest improvement on the rotating rod, a test of balance and coordination. However, the memory performance of Rett mice did not improve with extended training in the Morris water maze, a test of spatial learning and memory. Because Rett syndrome in humans and mice begins with a period of normal development, we next assessed whether the initiation of training before symptoms develop could enhance the benefit. Surprisingly, pre-symptomatic training dramatically improved the motor and memory defects of Rett mice to near wild-type levels. The benefits were task-specific because pre-symptomatic training did not improve other motor or memory behaviors. Currently, we are using high-resolution microscopy, calcium imaging, electrophysiology, and RNA sequencing to uncover the molecular basis of this improvement.

Because patients with Rett syndrome are diagnosis after the onset of symptoms, there are significant therapeutic implications to this study. The data suggest that pre-symptomatic diagnosis and early rehabilitation may be the key to modifying aspects of Rett syndrome.

## PgmNr 1263: Autosomal recessive infection-induced encephalopathy due to biallelic pathogenic variants in *NUP214*.

### Authors:

C. Applegate <sup>1</sup>; B. Fichtman <sup>2</sup>; T. Harel <sup>3</sup>; N. Biran <sup>2</sup>; F. Zagairy <sup>2</sup>; Y. Salzberg <sup>2</sup>; T. Gilboa <sup>4</sup>; S. Salah <sup>3</sup>; A. Shaag <sup>3,5</sup>; N. Simanovsky <sup>6</sup>; H. Ayoubieh <sup>1</sup>; N. Sobreira <sup>1</sup>; G. Punzi <sup>7</sup>; C. Pierri <sup>7</sup>; A. Hamosh <sup>1</sup>; O. Elpeleg <sup>3,5</sup>; A. Harel <sup>2</sup>; S. Edvardson <sup>4,5</sup>; Baylore-Hopkins Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Genetic Medicine, Johns Hopkins Univ, Baltimore, Maryland.; 2) Azrieli Faculty of Medicine, Bar-Ilan University, Safed 1311502, Israel; 3) Department of Genetic and Metabolic Diseases, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel; 4) Pediatric Neurology Unit, Hadassah-Hebrew University Medical Center, Jerusalem 91240, Israel; 5) Monique and Jacques Roboh Department of Genetic Research, Hadassah-Hebrew University Medical Center, Jerusalem 91120, Israel; 6) Department of Medical Imaging, Hadassah Medical Center, Jerusalem 91240, Israel; 7) Laboratory of Biochemistry, Molecular and Computational Biology; Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari, 70125 Bari, Italy

---

Acute encephalopathy following a febrile illness is an uncommon finding and often thought to be autoimmune. Several genetic causes have been described, most associated with specific infectious agents. However, heterozygous missense variants in RANBP2, a component of the nuclear pore complex, cause an autosomal dominant susceptibility to acute infection induced encephalopathy (MIM 608033). We report two unrelated families identified through GeneMatcher, one consanguineous family of Palestinian descent in which cousins were homozygous for a missense variant and one US family of Northern European descent in which two sisters were compound heterozygous for a missense and a frameshift pathogenic variant in the gene encoding human nucleoporin NUP214 causing acute infection-induced encephalopathy. Clinical symptoms include fever-induced neurodevelopmental regression, seizures, myoclonic jerks, progressive microcephaly and cerebellar atrophy. Patients made some post-infection recovery, but not to baseline state. Over time there was progressive neurodegeneration. One patient died at 11 years. Assays in primary fibroblasts from affected individuals showed reduced NUP214 and NUP88 protein levels, while the total number and density of nuclear pore complexes was comparable to control fibroblasts. Nuclear transport assays exhibited defects in the classical protein import and mRNA export pathways in affected cells. Using scanning electron microscopy to perform direct surface imaging of fibroblast nuclei revealed a large increase in the presence of central particles ("plugs") in the nuclear pore channels of affected cells. Heat shock exposure of fibroblasts from affected individuals resulted in a marked delay in their stress response, followed by a surge in apoptotic cell death. These studies suggest that decreased cell survival seen in cell culture is the pathogenic mechanism underlying severe fever-induced brain damage in affected individuals. While defects in nucleoporins are implicated in several kidney and other tissue-specific diseases, this study provides evidence by direct imaging at the single nuclear pore level of functional changes in a severe neurologic condition.



## PgmNr 1264: DNA sequence variants in the *CLDN5* gene cause a syndrome characterized by seizures, microcephaly and developmental delay.

### Authors:

A.R. Deshwar<sup>1</sup>; C. Cytrynbaum<sup>1</sup>; J. Zon<sup>1</sup>; R. Newbury-Ecob<sup>2</sup>; S. Ellard<sup>3</sup>; H. Lango Allen<sup>4</sup>; E. Yu<sup>1</sup>; R. Noche<sup>1</sup>; S. Walker<sup>1</sup>; S.W. Scherer<sup>1</sup>; M. Bainbridge<sup>5</sup>; J.R. Friedman<sup>5</sup>; Y. Capri<sup>6</sup>; L.L.P. Ramos<sup>7</sup>; D. Chitayat<sup>1</sup>; S. Blaser<sup>1</sup>; J.J. Dowling<sup>1</sup>; R. Weksberg<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Clinical Genetics, University Hospitals Bristol NHS Foundation Trust, Bristol, UK; 3) Molecular Genetics, Royal Devon & Exeter NHS Foundation Trust, Exeter UK; 4) University of Cambridge, Cambridge, UK; 5) Rady Children's Institute for Genomic Medicine, San Diego, CA; 6) Département de Génétique, Assistance Publique des Hôpitaux de Paris (AP-HP) Hôpital Robert Debré, Paris, France.; 7) Mendelics Genomic Analysis, Sao Paulo, Brazil.

---

The blood brain barrier (BBB) acts to ensure central nervous system (CNS) homeostasis and protects it from injury. Claudin-5 (CLDN5) is an important component of tight junctions which are critical for the proper formation of the blood brain barrier. We have identified through whole exome sequencing six unrelated patients with *de novo* mutations in *CLDN5* who present with a common constellation of features including microcephaly, developmental delay and seizures. Brain MRI are notable for brain calcifications. Generation of loss of function mutants in zebrafish *claudin5a* resulted in animals with growth restriction and seizures. Furthermore, these mutants were found to exhibit defects in the blood brain barrier. We find evidence that sequence variants in *CLDN5* cause a novel genetic disorder likely caused by the disruption of the blood brain barrier. Elucidation of the mechanisms by which variants in *CLDN5* result in abnormal brain development will have broad implications for our understanding of neurological disease.

## PgmNr 1265: Phenotype variabilities of the patients with *GABRB1* variants.

### Authors:

N. Ishihara <sup>1</sup>; H. Inagaki <sup>2</sup>; H. Kurahashi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Pediatrics, Fujita Health University, Toyoake, Aichi, Japan; 2) Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi, Japan

[Introduction] *GABRB1* has been reported as a causative gene for epileptic encephalopathy. Epileptic encephalopathy is a disease that develops intractable epilepsy early in infancy and is accompanied by severe developmental delay. Here we report that we identified variants of the same gene in a case in which comprehensive genetic analysis was performed for the purpose of differentiating congenital hypotonia.

[Subject] Subject was a 2-year-old girl born to healthy non-consanguineous parents. She had weakness of crying and difficulty for feeding since at birth. At 2 months, she was diagnosed as laryngomalacia and oxygen supply became necessary during feeding. She developed focal seizures at 3 months, in which were well controlled by carbamazepine (CBZ) administration. Her developmental milestone revealed as following; head controlled at 7 months, able to sit at 15 months, and stand with support at 20 months of age.

[Methods] Whole exome sequencing analysis was performed with trio DNA samples extracted from white blood cells of subject and her parents. Annotation and filtering of variants in which were *de novo* and autosomal recessive (AR) forms were performed by comparing to human reference sequences, and extracted variants with less than one percent allele frequency from the frequency information of the healthy human database.

[Results] One *de novo* and 4 AR variants were filtered as candidate. Among these, we identified *de novo* variant of *GABRB1* c.860C>T (p.Thr287Ile) as a causal gene of the disease. Thr287 is completely conserved in vertebrates, and in silico analysis revealed as pathogenic.

[Discussion] *GABRB1* is a gene encoding GABA<sub>A</sub> receptor  $\beta$ 1 subunit and has been reported as a causative gene for infantile spasm related diseases (Janve et al. Ann Neurol 2016). The variant (p.Thr287Ile) was first reported by Lien et al. (Ann Neurol 2016) as a pathogenic cause of epileptic encephalopathy. In our case, the epileptic seizures had been well controlled by CBZ, thus the main phenotype was hypotonia. Although the pathogenesis of hypotonia is various, the therapeutic intervention based on the genetic variants also leads an important role for developing brain, the value of the pathogenesis identification has been increasing in recent years. Since variants in *GABRB1* alter function of GABAergic neurons in which has an important role for inhibition of epileptic activities, early antiepileptic intervention would be require for appropriate brain development.

## PgmNr 1266: Mutations in *PIGB* cause an inherited GPI biosynthesis defect.

### Authors:

B. Lace<sup>1</sup>; Y. Murakami<sup>2</sup>; T.T.M. Nguyen<sup>3</sup>; N. Baratang<sup>3</sup>; P.K. Raju<sup>3</sup>; A. Knaus<sup>4</sup>; S. Ellard<sup>12</sup>; J. Rousseau<sup>3</sup>; N.F. Ajeawung<sup>3</sup>; A. Kamei<sup>5</sup>; G. Minase<sup>6</sup>; M. Akasaka<sup>5</sup>; N. Araya<sup>5</sup>; E. Koshimizu<sup>6</sup>; S. Miyatake<sup>6</sup>; N. Matsumoto<sup>6</sup>; J. van den Ende<sup>13</sup>; F. Erger<sup>7</sup>; J. Altmüller<sup>10</sup>; J. Strautmanis<sup>8</sup>; I. Inaskina<sup>9</sup>; J. Stavusis<sup>9</sup>; A. El-Gharbawy<sup>14</sup>; J. Sebastian<sup>14</sup>; R.D. Puri<sup>11</sup>; S. Kulshrestha<sup>11</sup>; Z. Krumina<sup>8</sup>; I.C. Verma<sup>11</sup>; S. Bijarnia-Mahay<sup>11</sup>; P.M. Campeau<sup>3</sup>; PIGB research group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) CHUL-CHUQ, Quebec, Quebec, Canada, Canada; 2) Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan; 3) CHU Sainte-Justine Research Center, Montreal, QC, Canada; 4) Institute for Genomic Statistics and Bioinformatics, University Hospital Bonn, 53127 Bonn, Germany; 5) Department of Pediatrics, School of Medicine, Iwate Medical University, Morioka, Iwate 020-8505, Japan; 6) Clinical Genetics Department, Yokohama City University Hospital, Yokohama, Kanagawa 236-0004, Japan; 7) Institute of Human Genetics, University Hospital of Cologne, and Center for Molecular Medicine, University of Cologne, 50931 Cologne, Germany; 8) Children's Clinical University Hospital, Riga, LV-1004, Latvia; 9) Latvian Biomedical Research and Study Centre, Ratsupites 1, Riga LV-1067; 10) Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, Germany; 11) Institute of Medical Genetics and genomics, Sir Ganga Ram Hospital, ENW Delhi, India 110060; 12) University of Exeter Medical School, Exeter EX2 5DW, UK; 13) Centrum Medische Genetica Antwerpen, 2650 EDEGEM, Belgium; 14) Department of Medical Genetics, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA 15224, USA

---

Proteins anchored to the cell surface via glycosylphosphatidylinositol (GPI) play various key roles in the human body, particularly in development and neurogenesis. As such, many developmental disorders are caused by mutations in genes involved in the GPI biosynthesis and remodeling pathway. We describe ten unrelated families with bi-allelic mutations in *PIGB*, a gene that encodes sphosphatidylinositol glycan class B that transfers the third mannose to the GPI. Ten different *PIGB* variants were found in these individuals. Flow cytometric analysis of blood cells and fibroblasts from the affected individuals showed decreased cell surface presence of GPI anchored proteins. Lentiviral expression of wildtype *PIGB* rescued the cellular phenotype in fibroblasts, supporting the pathogenicity of the *PIGB* variants. Most of the affected individuals have global developmental and/or intellectual delay, all had seizures, two had polymicrogyria, and four had a peripheral neuropathy. Eight children passed away at a young age. Some of them had a clinical diagnosis of DOORS syndrome (Deafness, Onychodystrophy, Osteodystrophy, mental Retardation, Seizures), a condition with sensorineural deafness, shortened terminal phalanges with small finger nail and toenails, intellectual disability, and seizures, which overlaps with the severe phenotypes associated with inherited GPI deficiency. However, the main causative gene for DOORS syndrome is *TBC1D24*, which is involved in synaptic-vesicle trafficking. Most individuals tested showed elevated alkaline phosphatase, which is a characteristic of the inherited GPI deficiency, but not DOORS syndrome. It is notable that some severely affected individuals showed 2-oxoglutaric aciduria, which can be seen in DOORS syndrome, suggesting that severe cases of inherited GPI deficiency and DOORS syndrome

might share some molecular pathway disruptions.

## PgmNr 1267: Whole exome sequencing of Finnish patients with vascular cognitive impairment.

### Authors:

S. Mönkäre <sup>1,2</sup>; L. Kuuluvainen <sup>1,3</sup>; S. Carmona <sup>4</sup>; C. Rodrigues <sup>4</sup>; J. Schleutker <sup>2,5</sup>; J. Bras <sup>4</sup>; M. Pöyhönen <sup>1,3</sup>; R. Guerreiro <sup>4</sup>; L. Myllykangas <sup>6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 2) Turku University Hospital, Laboratory Division, Genomics, Medical Genetics, Turku, Finland; 3) Department of Clinical Genetics, Helsinki University Hospital, Helsinki, Finland; 4) Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, MI, USA; 5) Institute of Biomedicine, University of Turku, Turku, Finland; 6) Department of Pathology, University of Helsinki, Helsinki, Finland; 7) Helsinki University Hospital, Helsinki, Finland

**Introduction:** Cerebral small vessel disease (CSVD) is the most important cause of vascular cognitive impairment (VCI). Vascular dementia (VaD) is the most severe form of VCI and it is the second most common cause of dementia after Alzheimer's disease. Most CSVD cases are sporadic but familial monogenic forms of the disorder have also been described. CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy), caused by mutations in *NOTCH3* gene, is the most common type of familial CSVD. CARASIL (cerebral autosomal recessive arteriopathy with subcortical infarcts and leukoencephalopathy) is an autosomal recessive CSVD caused by mutations in *HTRA1* gene, although autosomal dominant form of the disease has been recently identified. Furthermore, mutations in *COL4A1* and *COL4A2* have been identified as a cause of CSVD. Despite the mutations identified, many familial and sporadic CSVD cases remain unexplained genetically. We used whole exome sequencing in an attempt to identify novel genes underlying CSVD by studying 35 Finnish VaD patients without genetic diagnosis.

**Materials and methods:** A cohort of 35 unrelated Finnish patients with suspected CADASIL was selected from an initial cohort of 385 patients referred for diagnostic testing for *NOTCH3*. Patients were screened negative for at least the most common mutations in *NOTCH3* in Finland (R133C and R182C). After examining the medical records of all the patients, 35 patients were selected for whole exome sequencing. The inclusion criteria included the presence of vascular cognitive impairment with white matter changes in magnetic resonance imaging, age at onset up to 76 years and/or family history of dementia. Whole exome sequencing was performed to search for a genetic cause of CSVD.

**Results:** Our study resulted in the detection of pathogenic or possibly pathogenic variants or variants of unknown significance in genes associated with CSVD in seven patients, accounting for 20 % of cases. Those genes included *NOTCH3*, *HTRA1*, *COL4A1* and *COL4A2*. We also identified variants with predicted pathologic effect in genes associated with other neurological conditions in five patients.

**Conclusions:** This study revealed the genetic background of vascular cognitive impairment and further supports pathogenic roles of the mutations in *COL4A1*, *COL4A2* and *HTRA1* in CSVD. Our study also showed that whole exome sequencing is a useful tool for diagnosing familial cerebral small vessel disease.

## PgmNr 1268: Screening of over 1500 Indian patients affected with neurological disorders by multi-gene testing showed an improved diagnostic yield.

### Authors:

A. Mannan <sup>1</sup>; A. Ganapathy <sup>1</sup>; A. Mishra <sup>1</sup>; M.R. Soni <sup>1</sup>; P. Kumar <sup>1</sup>; V.K. Gowda <sup>2</sup>; S. Nampoothiri <sup>3</sup>; R.K. Menon <sup>4</sup>; D. Pachat <sup>5</sup>; V. Udani <sup>6</sup>; N. Naik <sup>7</sup>; M. Kamate <sup>8</sup>; A.R.R. Devi <sup>9</sup>; P.A.M. Kunju <sup>10</sup>; M. Nair <sup>11</sup>; A.U. Hedge <sup>12</sup>; M.P. Kumar <sup>13</sup>; S. Sundaram <sup>4</sup>; P. Tilak <sup>14</sup>; R.D. Puri <sup>15</sup>; K. Shah <sup>16</sup>; J. Sheth <sup>17</sup>; Q. Hasan <sup>18</sup>; F. Sheth <sup>17</sup>; P. Agrawal <sup>1</sup>; S. Katragadda <sup>1</sup>; V. Veeramachaneni <sup>1</sup>; V. Chandru <sup>1</sup>; R. Hariharan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Strand Life Sciences, Bangalore, India; 2) Indira Gandhi Institute of Child Health, Bangalore, India; 3) Amrita Institute of Medical Sciences and Research Centre, Cochin, India; 4) Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram, India; 5) Aster MIMS, Kozhikode, India; 6) P. D. Hinduja Hospital and Medical Research Centre, Mumbai, India; 7) EN1 Neuro Services Pvt. Ltd., Mumbai, India; 8) KLES Dr. Prabhakar Kore Hospital, Belgaum, India; 9) Rainbow Children's Hospital, Hyderabad, India; 10) SAT Hospital, Thiruvananthapuram, India; 11) Government Medical College, Kozhikode, India; 12) SRCC Children's Hospital, Mumbai, India; 13) GeneOmm Medical Centre, Coimbatore, India; 14) St. Johns Medical College Hospital, Bangalore, India; 15) Sir Ganga Ram Hospital, Delhi, India; 16) ONE - Centre for Rheumatology and Genetics, Vadodra, India; 17) FRIGE's Institute of Human Genetics, Ahmedabad, India; 18) Kamineni Hospitals, Hyderabad, India

**Background:** Neurological disorders are clinically heterogeneous group of disorders and are major causes of disability and death. Several of these disorders are caused due to genetic aberration. A precise and confirmatory diagnosis in the patients in a timely manner is essential for appropriate therapeutic and management strategies. Due to the complexity of the clinical presentations across various neurological disorders, arriving at an accurate diagnosis remains a challenge.

**Methods:** We sequenced 1521 unrelated patients from India with suspected neurological disorders, using TruSight One panel. Genetic variations were identified using the Strand NGS software and interpreted using the StrandOmics platform.

**Results:** We were able to detect mutations in 197 genes in 405 (40%) cases and 178 mutations were novel. The highest diagnostic rate was observed among patients with muscular dystrophy (64%) followed by leukodystrophy and ataxia (43%, each). In our cohort, 26% of the patients who received definitive diagnosis were primarily referred with complex neurological phenotypes with no suggestive diagnosis. In terms of mutations types, 62.8% were truncating and in addition, 13.4% were structural variants, which are also likely to cause loss-of-function.

**Conclusion:** In our study, we observed an improved performance of multi-gene panel testing, with an overall diagnostic yield of 40%. Furthermore, we show that NGS (next-generation sequencing)-based testing is comprehensive and can detect all types of variants including structural variants. It can be considered as a single-platform genetic test for neurological disorders that can provide a swift and definitive diagnosis in a cost-effective manner.

## PgmNr 1269: Novel mutations of *COL4A1* in four Japanese patients with variable CNS involvements.

### Authors:

N. Nishimura <sup>1,2</sup>; Y. Enomoto <sup>3</sup>; Y. Tsurusaki <sup>3</sup>; T. Kumaki <sup>1</sup>; H. Murakami <sup>1</sup>; Y. Kuroda <sup>1</sup>; T. Goto <sup>4</sup>; K. Kurosawa <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Department of Pediatrics, National Defense Medical College, Saitama, Japan; 3) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 4) Division of Neurology, Kanagawa Children's Medical Center, Yokohama, Japan

---

*COL4A1* encodes collagen type IV alpha 1, the main component of basement membranes. The mutation causes severe CNS abnormalities including porencephaly and schizencephaly. Despite large-scale studies, the genotype-phenotype correlation remains unclear. Here, we report novel *de novo* *COL4A1* mutations in four Japanese patients with typical or atypical CNS involvements, representing phenotypic diversity of the disorder.

All the patients were males with absence of abnormal events in the fetal period. They were admitted for developmental delay, growth retardation, and microcephaly. They presented the CNS symptoms, including spastic paralysis and epilepsy. Brain MRI revealed porencephaly (n = 1), ventriculomegaly associated with brain bleeding (n = 3), and atretic encephalocele (n = 1), respectively. Additionally, three patients had cataract, hydronephrosis, and hematuria. Using the clinical exome sequencing, we identified *COL4A1* mutations: two patients with missense mutations resulting in a glycine substitution (exon32:c.G2555A:p.G852D, exon40:c.G3407A:p.G1136D) and two with small in/del variants resulting in frameshift (exon32:c.2603\_2609delinsATCCTGA:p.A868\_G870delinsDPG, exon36:c.3054delinsTGTAGAT:p.L1018delinsFVD). Notably, the two patients with in/del mutation showed severe CNS involvements, porencephaly and more severe hydrocephaly. Further, the atretic encephalocele was a rare distinctive phenotype, which has never been described among the *COL4A1* related diseases.

These results indicate that *COL4A1* mutations show highly variable phenotype in cerebral hemorrhage and porencephaly. The molecular analysis using exome sequencing could expand the related phenotype of *COL4A1* mutations. Further analysis is required to clarify the underlying mechanisms of the clinical spectrum in the *COL4A1* related diseases.



## PgmNr 1270: Malformations of cortical development: International consensus recommendations on diagnostic work-up.

### Authors:

R. Oegema<sup>1</sup>; D. Amron<sup>2</sup>; E. Aronica<sup>3,4</sup>; N. Bahi-Buisson<sup>5</sup>; T.S. Barakat<sup>6</sup>; V. Conti<sup>7</sup>; W.B. Dobyns<sup>8,9</sup>; A. Fry<sup>10,11</sup>; T. Geis<sup>12</sup>; D. Gomez Andres<sup>13</sup>; R. Guerrini<sup>10,11</sup>; U. Hehr<sup>14</sup>; A. Jansen<sup>15</sup>; R. Leventer<sup>16</sup>; G.M.S. Mancini<sup>6</sup>; G. Mirzaa<sup>8,9</sup>; E. Parrini<sup>10,11</sup>; D. Pilz<sup>17</sup>; E. Said<sup>18</sup>; D. Soler<sup>19</sup>; K. Stouffs<sup>20</sup>; L.M. Valor<sup>21</sup>; M. Wilke<sup>6</sup>; M. Zaki<sup>22</sup>; N. Di Donato<sup>23</sup>; Neuro-MIG Network (COST Action CA16118)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands; 2) Centre Hospitalier de Luxembourg, Hôpital Universitaire des Enfants Reine Fabiola, Luxembourg; 3) Amsterdam UMC, University of Amsterdam, Dept (Neuro)pathology, Amsterdam, Amsterdam Neuroscience, the Netherlands; 4) Stichting Epilepsie Instellingen Nederland (SEIN); 5) Pediatric Neurology Necker Enfants Malades University Hospital Imagine Institute Paris France; 6) Department of Clinical Genetics, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 7) Pediatric Neurology, Neurogenetics and Neurobiology Unit and Laboratories, Department of Neuroscience, A. Meyer Children's Hospital, University of Florence, Florence, Italy; 8) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, Washington, USA; 9) Department of Pediatrics, University of Washington, Seattle, Washington, USA; 10) Institute of Medical Genetics, University Hospital of Wales, Cardiff CF14 4XW, UK; 11) Division of Cancer and Genetics, School of Medicine, Cardiff University, Cardiff CF14 4XN, UK; 12) Department of Pediatric Neurology, Klinik St. Hedwig, University Children's Hospital Regensburg (KUNO), Regensburg, Germany; 13) Child Neurology. Hospital Universitari Vall d'Hebron. Barcelona, Spain; 14) Center for and Institute of Human Genetics, University of Regensburg, Regensburg, Germany; 15) Pediatric Neurology Unit, Department of Pediatrics, UZ Brussel; Neurogenetics Research Group, Vrije Universiteit Brussel, Belgium; 16) Department of Neurology, Royal Children's Hospital, Murdoch Children's Research Institute and University of Melbourne Department of Pediatrics, Melbourne, Australia; 17) West of Scotland Clinical Genetics Service, Queen Elizabeth University Hospital, Glasgow G51 4TF, UK; 18) Section of Medical Genetics, Mater dei Hospital, Msida, Malta and Department of Anatomy and Cell Biology, University of Malta, Msida, Malta; 19) Department of Paediatrics, Mater dei Hospital, Msida, Malta; 20) Centre for Medical Genetics, UZ Brussel; Neurogenetics Research Group, Vrije Universiteit Brussel, Belgium; 21) Hospital Universitario Puerta del Mar, INIBICA. Spain; 22) Clinical Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 23) Institute for Clinical Genetics, TU Dresden, Dresden, Germany

---

Malformations of cortical development (MCD) are a group of predominantly genetic disorders due to abnormal formation of cerebral cortex during fetal brain development. Examples are lissencephaly, heterotopia, and polymicrogyria. MCD place a significant burden on affected individuals, their families and society, as patients often suffer from lifelong symptoms including drug-resistant epilepsy, cerebral palsy, and intellectual disability. Although next-generation sequencing has become broadly available worldwide, marked differences in diagnostic yield remain due to broad genetic and clinical heterogeneity and lack of guidelines in the recognition and classification of these rare disorders.



The international network on brain malformations NeuroMIG ([www.neuromig.org](http://www.neuromig.org)) has formed a multidisciplinary task force with the aim to develop a diagnostic strategy for individuals with MCD. Through literature search, expert input and the Delphi consensus method we have established recommendations on the optimal work flow. This includes a flow chart for the clinical work-up, laboratory standards, and a curated MCD gene panel list. In addition, we discuss the differential diagnosis, possible diagnostic pitfalls and special recommendations for specific MCD subtypes.

These recommendations will aid clinicians in providing the optimal diagnostic strategy, which will improve genetic and prognostic counseling and patient care.

## PgmNr 1271: Idiopathic basal ganglia calcification linked to a SLC20A2 gene mutation.

### Authors:

F. Suarez; R. Quero; M.C. Manotas; J.C. Villalta; M. Guerra

[View Session](#) [Add to Schedule](#)

**Affiliation:** Instituto de Genetica Humana, Pontificia Universidad Javeriana, Bogota, D.C., Bogota D.C., Colombia

---

Idiopathic basal ganglia calcification (MIM 213600) is a neurodegenerative disorder characterized by calcium deposits in the basal ganglia, the dentate nucleus, and other brain areas. It appears in the fourth decade of life with progressive movement disorders, which begin with bradykinesia, slow gait, parkinsonism, chorea, tremor, dystonia, athetosis, ataxia, dyskinesia, and neuropsychiatric disorders such as difficulty in concentration, lack of memory, changes of personality or behavior and dementia. We describe a 61-year-old female patient with no family history with 6 months of the onset of symptom onset consisting of vertigo, lingual fasciculations, dysarthria, bradykinesia, memory impairment, and parkinsonism. Diagnostic studies were performed and ruled out toxic, traumatic, infectious, and metabolic etiology. Magnetic resonance showed hyperintensities of the striated, dentate nuclei, the periventricular region and the white matter of the cerebellar hemispheres. Investigations with a multigene panel, including *PDGFB*, *PDGFRB*, *SLS20A2*, *XPR1*, reported a probable pathogenic variant in the heterozygous state in *SLC20A2* c.1794+1G>T associated with idiopathic basal ganglia calcification type 1 which confirms the diagnosis of this patient. Currently, the patient is in treatment and clinical and symptomatic monitoring of neurological and neuropsychiatric status. Idiopathic basal ganglia calcification type 1 is caused by mutations in the *SLC20A2* gene that encodes the sodium-dependent inorganic phosphate transporter type III associated with phosphate homeostasis, which can lead to the accumulation of inorganic phosphate in the brain. Based on the low prevalence of this disease, it is important to know the frequent findings of this entity and consider the possible etiologies of neurological and/or neuropsychiatric unspecific symptoms, in order to control the morbidity and complications due to neurological damage of this pathology.

## PgmNr 1272: MECP2 duplication syndrome resulting from X:7 translocation in a 9-month-old female with severe failure to thrive and developmental delay.

### Authors:

H. Wang <sup>1</sup>; M. Montesinos <sup>2</sup>; S. Ramanathan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetics/Department of Pediatrics, Loma Linda University Children's Hospital, California.; 2) Department of Pediatrics, Loma Linda University Children's Hospital, California

---

**INTRODUCTION:** MECP2 microduplication syndrome is a severe neurodevelopmental syndrome that has been primarily described in males with the cardinal features of hypotonia, global developmental delay, gastrointestinal symptoms, and recurrent respiratory infections. So far only about 20 cases of symptomatic MECP2 duplication syndrome have been reported in females with variable clinical symptoms. **CASE DESCRIPTION:** This is a 9-month-old girl who was admitted for acute respiratory failure with hypoxia, RSV infection. Genetics was consulted for hypotonia, failure to thrive and severe developmental delay. Her previous evaluation at age of 44 days due to positive newborn screen for X-linked adrenoleukodystrophy revealed normal levels of VLCFA and plasminogen although the decreased muscle tone was noticed. The patient had history of neglect prior to the admission. She appeared severely malnourished, with a weight of 5.5kg (Z= -3.47). No obvious dysmorphic features except the ears were relatively big. She had no psychomotor skills, and muscle tone was low. She had mild to moderate laryngomalacia, L>R hearing loss. Comprehensive metabolic workup, SMA test, Russell Silver methylation study were negative. MRI was suggestive of hypomyelination. The chromosome microarray result showed 2.6Mb terminal duplication on Xq28 and a 22kb terminal deletion on 7q36.3. Metaphase FISH indicated an unbalanced X:7 translocation. **DISCUSSION:** The MECP2 duplication syndrome caused by duplication on X28 is an established disorder. Methyl CpG binding protein 2 (MECP2), functions as a transcription regulator for gene expression is associated with central nervous system development. Loss-of-function mutations in MECP2 results in the Rett syndrome, whereas gain-of-function mutations are associated with the MECP2 duplication syndrome. Different from male cases, female cases caused by X;autosome translocation are reported with more severe phenotypes due to random X chromosome inactivation (XCI). **CONCLUSIONS:** Variable clinical symptoms and poor correlation between the genotype and phenotype can cause misleading and delayed diagnosis. By literature search, this exact duplication has not been previously reported in females resulting from an X;7 translocation. The further investigation of the XCI study will help to delineate the underlying mechanism. In the clinical setting of failure to thrive with other symptoms that cannot be explained, chromosome microarray should be offered.

## PgmNr 1273: Characterization of *de novo* variants in *PPP2R5C* expands the spectrum of PP2A-related neurodevelopmental disorders.

### Authors:

E. Brimble<sup>1</sup>; D. Haesen<sup>2</sup>; L. Lenaerts<sup>2</sup>; S. Reynhout<sup>2</sup>; D. Genevieve<sup>3</sup>; J. Thevenon<sup>4</sup>; J.X. Chong<sup>5</sup>; D.A. Nickerson<sup>6</sup>; M.J. Bamshad<sup>6</sup>; E.J. Marco<sup>7</sup>; K. Foss<sup>8</sup>; W.B. Dobyns<sup>5,8,9</sup>; G. Mirzaa<sup>5,8</sup>; C. Fong<sup>10</sup>; A. Telegrafi<sup>11</sup>; S. Punj<sup>11</sup>; M.R.Z. Ruzhnikov<sup>1,12</sup>; V. Janssens<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology and Neurological Sciences, Stanford Medicine, Stanford, CA; 2) Laboratory of Protein Phosphorylation & Proteomics, Department of Cellular & Molecular Medicine, University of Leuven (KU Leuven), Belgium; 3) Département de Génétique Médicale, Maladies Rares et Médecine Personnalisée, Université Montpellier, Hopital Arnaud de Villeneuve, Unité Inserm U1183, Montpellier, France; 4) CNRS UMR 5309, INSERM U1209, Institute of Advanced Biosciences, Université Grenoble-Alpes; Service Génomique et Procréation, Centre Hospitalo-Universitaire Grenoble Alpes, Grenoble Cedex, France; 5) Division of Genetic Medicine, Department of Pediatrics, University of Washington School of Medicine, Seattle, WA; 6) Department of Genome Sciences, University of Washington, Seattle, WA; 7) Cortica Healthcare, San Rafael, CA; 8) Center for Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA; 9) Department of Neurology, University of Washington, Seattle, WA; 10) Department of Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, NY; 11) GeneDx, 207 Perry Parkway, Gaithersburg, MD; 12) Division of Medical Genetics, Department of Pediatrics, Stanford Medicine, Stanford, CA

---

**Background:** Protein phosphatase 2A (PP2A) is a ubiquitously expressed serine-threonine phosphatase with widespread involvement in multiple cellular processes. Its function is, in part, determined by its composition, with variable catalytic, scaffolding, and regulatory subunits. Of the 19 genes encoding PP2A subunits, six have been implicated in human disease. A single case report describes an overgrowth phenotype in association with a *de novo* variant in *PPP2R5C*, which encodes the PP2A-B56γ regulatory subunit. **Methods:** Through international collaboration facilitated by Matchmaker Exchange, we identified six additional individuals with candidate missense variants in *PPP2R5C*. Retrospective clinical data were collected. The functional impact of *PPP2R5C* variants was evaluated in a HEK293 cell model, in which GFP-tagged PP2A-B56γ variants were ectopically expressed. Following isolation of GFP-tagged proteins, co-precipitation of endogenous PP2A catalytic C and structural A subunits was evaluated by immunoblotting. Associated PP2A catalytic activity was measured on a general PP2A phospho-peptide substrate using a malachite green assay. **Results:** We describe and characterize four novel *PPP2R5C* missense variants in six individuals. We also include functional characterization of a previously reported variant described above. Our cohort, along with the previously published individual, have a consistent clinical presentation that includes: (relative) macrocephaly (4/7), developmental and/or intellectual disabilities (7/7), and congenital hypotonia (7/7). One subject also has refractory epilepsy. All described missense variants are absent from large population databases and affect highly conserved residues; six were confirmed to occur *de novo*. A recurrent variant, p.E122K, was observed in three unrelated individuals; the E122 residue has been previously shown to form direct interactions with the catalytic subunit, and the corresponding residue

in PP2A-B56 $\beta$  is required for target dephosphorylation. Accordingly, we detected significantly impaired binding of scaffolding and catalytic subunits and reduced PP2A catalytic activity compared to wildtype. The impact on subunit binding and phosphatase activity was also evaluated for the remaining variants. **Conclusions:** *De novo* missense variants in *PPP2R5C* cause impaired function of PP2A, providing evidence for a novel neurodevelopmental disorder with a consistent clinical phenotype.

## **PgmNr 1274: Bi-allelic variants in the autophagy gene *ATG4D* are associated with a pediatric-onset neurological disorder characterized by hypotonia, dysarthria, impaired coordination, and gait abnormalities.**

### **Authors:**

M. Morimoto <sup>1</sup>; M. Davids <sup>1</sup>; P.M. Zerfas <sup>2</sup>; V. Bhambhani <sup>3</sup>; A. Lehman <sup>4</sup>; J.L. Murphy <sup>1</sup>; E. Macnamara <sup>1</sup>; C. Wang <sup>1</sup>; U.D.N. Undiagnosed Diseases Network <sup>5</sup>; S. Reichert <sup>6</sup>; A. Thurm <sup>7</sup>; W.J. Introne <sup>8</sup>; W.A. Gahl <sup>1,8,9</sup>; C.J. Tiff <sup>1,8,9</sup>; M.C.V. Malicdan <sup>1,8,9</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) National Institutes of Health Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, MD, USA; 2) Diagnostic and Research Services Branch, Office of Research Services, National Institutes of Health, Bethesda, MD, USA; 3) Department of Medical Genetics, Children's Hospitals and Clinics of Minnesota, Minneapolis, MN, USA; 4) Department of Medical Genetics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; 5) Common Fund, Office of the Director, National Institutes of Health; 6) Department of Clinical Genetics, Virginia Commonwealth University Health, Richmond, VA, USA; 7) Neurodevelopmental and Behavioral Phenotyping Service, Office of the Clinical Director, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA; 8) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 9) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

---

**Introduction:** Macroautophagy is a dynamic and highly conserved process that regulates the degradation and recycling of cellular components. Double-membraned organelles known as autophagosomes deliver cellular components to the lysosomes for degradation. Defective autophagy has been shown to contribute to the pathogenesis of various human diseases. We describe two unrelated individuals with a neurological disorder and identify a candidate gene encoding *ATG4D*, one of the four *ATG4* isoforms that process the LC3 and GABARAP proteins required for autophagosome biogenesis and dynamics. Interestingly, homozygous missense variants in *ATG4D* in a canine model are also associated with a neurological disease.

**Methods:** Clinical evaluation was performed on each proband. Exome sequencing and family-based genomics were performed, and variants were prioritized based on their frequency in population databases, level of conservation, and predicted deleteriousness. Expression analyses were performed on primary dermal fibroblasts using quantitative PCR. Autophagic flux was assessed by western blot of autophagy markers p62 and LC3-II and transmission electron microscopy (TEM) on a skin biopsy and primary dermal fibroblasts from one proband.

**Results:** Two probands presented with a neurological phenotype characterized by hypotonia, dysarthria, impaired coordination, and gait abnormalities. Rare, conserved, and likely deleterious bi-allelic missense or frameshift variants were identified in *ATG4D* in both probands segregating with the disease. TEM of skin from one proband revealed occasional accumulation of autophagosomes,

suggesting an upregulation of basal autophagy that is similar to findings from the canine model. Primary dermal fibroblasts from one proband showed decreased *ATG4D* mRNA expression, suggesting instability of the *ATG4D* mRNA. However, no significant differences in the expression of the autophagy markers p62 or LC3-II were identified at baseline or upon treatment with Bafilomycin A<sub>1</sub> (an autophagy inhibitor) and/or starvation (an autophagy inducer). Cellular morphological changes and the formation of autophagosomes in response to these treatments were also comparable to an unaffected control.

**Conclusion:** We report two unrelated individuals with a neurological disorder and hypothesize that bi-allelic variants in *ATG4D* may underlie the pathogenesis of this disorder. Further functional studies are required to delineate the molecular consequences of *ATG4D* deficiency.

## PgmNr 1275: Genetic analysis of autosomal dominant paroxysmal kinesigenic dyskinesia (PKD).

### Authors:

T. Shimojo <sup>1</sup>; S. Miura <sup>2</sup>; K. Kosaka <sup>1</sup>; T. Morikawa <sup>1</sup>; K. Aoki <sup>1</sup>; T. Kamada <sup>3</sup>; H. Shibata <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genomics, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-8582, Japan.; 2) Department of Geriatric Medicine and Neurology, Ehime University Graduate School of Medicine, Shitsukawa, Toon, Ehime 791-0295, Japan.; 3) Division of Respiratory, Neurology and Rheumatology, Department of Medicine, Kurume University School of Medicine, 67 Asahi-machi, Kurume, Fukuoka 830-0011, Japan

---

PKD is estimated to occur in one in 150,000 individuals, and characterized as involuntary movement at the start of motion. The age of disease onset is typically in childhood but the symptoms gradually ameliorates after puberty. Three genes have been reported so far to be responsible for PKD, *PRPT2*, *SLC2A1* and *MR1*. We ascertained one Japanese pedigree with PKD. This pedigree consists of two patients with five unaffected family members, suggesting an autosomal dominant inheritance. The aim of this study is to elucidate genetic basis responsible for the current familial case of PKD. We identified total of 314,953 single nucleotide variants (SNVs) by exome sequencing of two patients and two unaffected family members. First, we confirmed the absence of exonic SNVs in the three genes known to be responsible for PKD, suggesting that a novel gene is associated with the current PKD pedigree. Then we filtered SNVs detected by exome sequencing by the following fourth criteria: (i) the inheritance pattern consistent with the autosomal dominant model, (ii) deduced functional consequence, (iii) frequencies being zero or extremely low in public databases and in our in-house control samples. Through the filtering process, we selected three nonsynonymous SNVs as candidates for PKD, c.8976G>C [p.Gln2992His] in *FAT2*, c.5435T>C [p.Ser1672Pro] in *NUP98*, and c.8596C>T [p.Arg2866Trp] in *NBEA*. The SNVs in *FAT2* and *NBEA* are located in highly conserved regions among vertebrates, while the SNV in *NUP98* is located in a less conserved region among mammals. According to GTEx, the expression of *FAT2* in the cerebellum is much higher than those of *NUP98* and *NBEA*. *FAT2* and *NBEA* have been reported to be responsible for spinocerebellar ataxia 45 (SCA45) and familial epilepsy, respectively, suggesting the partial overlap of pathogenic mechanisms of PKD with the one of SCAs and/or epilepsy.



## PgmNr 1276: Analysis of DNA repair protein encoding gene, *RFC4*, as a candidate gene in two individuals with a novel neurological disorder.

### Authors:

B. Steger<sup>1</sup>; M. Morimoto<sup>1</sup>; L. Wolfe<sup>1,2</sup>; A. Huang<sup>3</sup>; H. Lee<sup>4,5</sup>; C. Lau<sup>1</sup>; S. Nelson<sup>4,5,6</sup>; D. Adams<sup>1,2</sup>; W. Gahl<sup>1</sup>; C. Toro<sup>1,2</sup>; M. Malicdan<sup>1,2</sup>; Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NIH Undiagnosed Diseases Program, NHGRI and the Common Fund, National Institutes of Health, Bethesda, MD 20892, USA; 2) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; 3) Institute for Precision Health, David Geffen School of Medicine, University of California, Los Angeles, California, USA; 4) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA; 5) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA; 6) Department of Pediatrics, David Geffen School of Medicine, University of California, Los Angeles, CA, 90095, USA

### Introduction:

DNA replication and repair play critical roles in cellular division and tolerance to DNA damage. Multiple accessory proteins are required to facilitate the binding and function of DNA polymerases during these processes. Variants in several genes encoding these DNA repair accessory proteins are associated with neurological disorders such as cerebellar ataxia, neuropathy, Vestibular Areflexia Syndrome (CANVAS) and Ataxia-Telangiectasia (AT). We present probands with an undiagnosed neurological disorder and identify a candidate gene encoding a DNA repair protein that may be associated with this novel disorder.

### Methods:

Probands were evaluated at the National Institutes of Health Clinical Center through the Undiagnosed Diseases Network's NIH clinical site, the Undiagnosed Diseases Program (UDP). Clinical genome sequencing was performed on probands and available family members. RNA sequencing identified possible splicing variants, and Sanger sequencing was used to validate candidate variants. Splicing variants were confirmed by cDNA sequencing. mRNA and protein expression were analyzed in dermal fibroblasts from probands using quantitative PCR and western blot analysis.

### Results:

Two NIH UDP probands presented with cerebellar neurodegeneration, ataxia, photophobia, dysarthria, and other striking phenotypic similarities to known DNA repair disorders. A UDN initiated collaborative analysis identified splicing variants from RNA sequencing in one proband. Research re-analysis of genome sequences revealed compound heterozygous variants in *RFC4*, including likely-deleterious splicing variants, in both probands segregating with disease. *RFC4* encodes a subunit of replication factor C (RFC), a DNA repair accessory protein that loads and activates DNA polymerases  $\delta$  and  $\epsilon$ . *RFC4* mRNA expression was either unchanged or slightly increased in patient fibroblasts compared to an unaffected control suggesting possible overcompensation of transcription. cDNA sequencing

confirmed splicing variants resulting in deletions of exons 4 or 10, respectively, in fibroblasts from the two probands.

**Conclusion:**

The changes in *RFC4* splicing seen in fibroblast cDNA from our probands, in combination with the phenotypes similar to those of AT, CANVAS, ATLD2, and other DNA repair disorders, suggest that the bi-allelic variants in *RFC4* are pathogenic. Our report implicates *RFC4* as a gene involved in a novel DNA repair disorder.

## PgmNr 1277: Deficient pseudouridylation activates autophagy resulting in postnatal microcephaly, developmental delay, and metabolic abnormalities in two siblings with *PUS7* loss of function variants.

### Authors:

S.T. Han <sup>1,2</sup>; E. Macnamara <sup>2</sup>; L.A. Schimmenti <sup>3</sup>; R.L. Proia <sup>4</sup>; M.C. Malicdan <sup>1,2</sup>; W.A. Gahl <sup>1,2</sup>; C.J. Tifft <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland.; 2) Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, Maryland.; 3) Department of Clinical Genomics, Mayo Clinic Alix School of Medicine, Rochester, Minnesota; 4) National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland.

---

Delineating the mechanism underlying disease pathology is critical for informing clinical care and remains a challenge for many genetic disorders. This is especially true for newly identified genetic disorders caused by variants in genes not previously associated with disease. In these cases, identification of pathogenic variants is often insufficient to inform clinical care. Mechanistic studies are necessary to understand normal function of candidate genes as well as the consequences of their dysfunction. Very recently, pathogenic variants in pseudouridine synthase 7 (*PUS7*) were shown to be associated with postnatal microcephaly, aggressive behavior, speech delay, and developmental delay in several unrelated individuals from consanguineous families. We report two siblings in a non-consanguineous family from the mid-western United States who are compound heterozygotes for rare splice and missense variants in *PUS7*. While these siblings showed phenotypic overlap with published reports, they also presented with hearing loss, hypoglycemia, hyperuricemia, and self-injurious behavior.

Loss of *PUS7* function was confirmed in these siblings using RNA and protein from fibroblasts. *HPRT1* was also evaluated for its possible role in the hyperuricemia and self-injurious behavior unique to our patients. Exome sequencing, SNP array, and qPCR ruled out *HPRT1* deficiency as the explanation for these Lesch-Nyhan syndrome-like findings. We therefore focused on the role of *PUS7* in disease pathophysiology. Pseudouridine synthases catalyze the conversion of uridine to pseudouridine of many RNA molecules. *PUS7* can modulate a wide range of RNAs but has been shown to predominantly target specific tRNAs, which then regulate protein translation through the action of pseudouridylated tRNA fragments (tRFs). We observed that loss of *PUS7*-generated tRFs in patient fibroblasts increased global protein synthesis, consistent with published *in vitro* studies. Incidentally, patient cells also exhibited increased expression of transcription factor EB (*TFEB*), a master regulator of autophagy. While studies are underway to confirm and better understand these findings, our results suggest that global upregulation of protein translation has aberrantly activated autophagy in patient cells, manifesting a phenotype most clearly in the CNS. These cases add to a growing body of literature describing deficient tRNA modification as a genetic cause of neurological diseases that often include microcephaly.

## PgmNr 1278: *DYNC1H1*-related disease: Reviewing the literature, describing three novel *de novo* variants, and expanding the phenotypic spectrum.

### Authors:

L. Jeffries<sup>1</sup>; S. Amabile<sup>1,2</sup>; M. Konstantino<sup>1</sup>; W. Ji<sup>1</sup>; M. Spencer-Manzon<sup>3</sup>; H. Zhang<sup>3</sup>; J. McGrath<sup>3</sup>; S. Lakhani<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatric Genomics Discovery Program, Yale University, New Haven, Connecticut; 2) Department of Molecular Medicine and Medical Biotechnologies, University of Naples Federico II, Naples, Italy; 3) Department of Genetics, Yale University, New Haven, Connecticut

The *DYNC1H1* gene is associated with different phenotypes, chiefly intellectual disability (ID), spinal muscular atrophy (SMA), Charcot-Marie-Tooth (CMT) and other hereditary neuropathies (HN), congenital myopathy (CM), and malformations of cortical development (MCD). Other phenotypes have also been described, primarily in single reports or small case series. Overlap of features has been readily observed.

Here, we report four patients identified at our institution. Three have novel, *de novo* variants (p.V1116A, p.P1511L, and p.H1412Y). One has a recurrent variant, p.G3658E, which was identified previously in a proband with similar unusual findings, such as corpus callosum hypoplasia and congenital cataracts. Our patients all fell variably within the described spectrum of *DYNC1H1*-related illnesses. One (with p.H1412Y) is the first well-described patient to present primarily with autism (ASD) and mild ID; interestingly, this patient is only the second to have documented gut dysmotility associated with *DYNC1H1*, and the first with omphalocele.

Additionally, we review previously-reported *DYNC1H1* variants and have identified 143 unique families and/or probands harboring 103 different variants associated with neurological or developmental phenotypes. Variants are found throughout the protein in the stem and motor domains and in the connecting neck/linker. We classified phenotypes into 6 major groups: SMA, CMT/HN, ID without explicit neuromuscular (NM) disease, MCDs without explicit ID, and CM. We analyzed data for phenotype/genotype relationships. Phenotypes largely cluster amongst affected family members and those with the same variant; however, there are exceptions, and other phenotype/genotype relationships are elusive. Most CMT/SMA variants are in the stem domain, while most ID variants are in the motor domain. However, there are many exceptions, and most subdomains contain at least two variants: at least one associated with NM symptoms and at least one associated with primarily MCDs.

We conclude that *DYNC1H1* can present with multiple overlapping neurodevelopmental phenotypes. *DYNC1H1*-related CMT, SMA and CM are likely a spectrum of the same disorder. We also establish gut dysmotility and confirm cataracts, arthrogryposis, congenital hip dislocation, contractures, scoliosis or other spine deformities, ADHD, ASD, dysmorphia, microcephaly, seizures, and foot deformities as recurrent features of *DYNC1H1*-related disease.

## PgmNr 1279: Development of clinical and model organism platforms to inform the phenotypic spectrum and functional characteristics of newly-identified variants in *IRF2BPL*.

### Authors:

L.D.M. Pena <sup>1,7</sup>; P.C. Marcogliese <sup>2</sup>; R.C. Spillmann <sup>3</sup>; D. Dutta <sup>2</sup>; F. Fazal <sup>2</sup>; H.J. Bellen <sup>2,4,5,6</sup>; V. Shashi <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 3) Division of Medical Genetics, Department of Pediatrics, Duke University School of Medicine, Durham, North Carolina; 4) Howard Hughes Medical Institute, Houston Texas; 5) Program in Developmental Biology, Baylor College of Medicine, Houston, Texas; 6) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas; 7) Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati OH

---

**Introduction:** *IRF2BPL* was identified in 2018 as a new disease gene that causes Neurodevelopmental Disorder with Regression, Abnormal Movements, Loss of Speech, and Seizures (NEDAMSS MIM# 618088). *De novo* truncating mutations in *IRF2BPL*, a single-exon gene of unknown function leads to severe neurological regression after an initial period of unremarkable development in affected individuals. These patients initially present with deterioration of gait and loss of intentional movements, and subsequently develop dysarthria, loss of speech, movement disorder, seizures, and loss of fine motor skills. Individuals with missense variants lack the neuroregression but may display milder symptoms of developmental delay, autism spectrum disorder, and seizures. As a newly-described condition, it is important to develop an understanding of the phenotypic spectrum of disease complications. We describe here our efforts to collect phenotypic data on newly identified cases and characterize the impact of new variants on protein function.

**Methods:** With input from various lab and clinical partners, we developed an IRB-approved protocol to collect longitudinal prospective and retrospective phenotypic data. We developed a data-collection tool that relies on participant-led data entry. Recruitment is based on online advertisements. *In vivo* functional studies using *Drosophila* as a model organism assessed *IRF2BPL* variant function.

**Results:** As a new disease gene, knowledge regarding the phenotypic spectrum is understandably limited. Our most common source of queries and referrals was from ordering physicians who had read our initial publication. In addition, we had introductions to treating clinicians through the clinical exome labs that had identified variants in *IRF2BPL*. Questions regarding known and potential complications and management were common, and we had frequent requests for additional information regarding additional observations not included in the original cohort. A recurrent question was whether variants of unclear significance in *IRF2BPL* could explain the patient's symptoms. This presented a valuable opportunity for functional analysis from model organisms to inform the biological effect of the variants. *IRF2BPL* overexpression in fruit flies allows for a variety of functional assessment depending on the tissue-specific expression of the construct. Our work illustrates the utility of prospective data collection and model organism data in phenotypic characterization.

## PgmNr 1280: Prodromal manifestations of Parkinson's disease in adults with 22q11 microdeletion syndrome.

### Authors:

G.M. Repetto <sup>1</sup>; C. Juri <sup>2</sup>; V. Kramer <sup>3</sup>; A. Cuiza <sup>1</sup>; P. Chana-Cuevas <sup>4</sup>; R. Fritsch <sup>5</sup>; C. Ornstein <sup>5</sup>; B. Rebolledo <sup>1</sup>; A. Ocampo <sup>6</sup>; K. Villanueva <sup>6</sup>; T. Cordova <sup>6</sup>; G. Encina <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetics and Genomics, Clin Alemana- Univ Desarrollo, Santiago, RM, Chile; 2) Dept of Neurology, Facultad de Medicina, P. Universidad Catolica de Chile; 3) Centro de Medicina Nuclear y PET/CT PositronMed, Santiago, Chile; 4) Centro de Trastornos del Movimiento (CETRAM), Facultad de Ciencias Médicas, Universidad de Santiago de Chile; 5) Dept of Psychiatry, Facultad de Medicina, Universidad de Chile; 6) Instituto de Ciencias Biomedicas Facultad de Medicina, Universidad de Chile

22q11 microdeletion syndrome (22q11DS) has been recently recognized as a cause of early onset Parkinson's disease (PD), presenting in approximately 4% of adult series, with onset at 35-40 years of age. Multifactorial PD usually presents after 60 years of age, and the classical motor symptoms are preceded by a prodromal phase by about a decade. This prodromal phase is characterized by anosmia, dysautonomia, REM sleep behavior disorders (RBD) and gradual loss of dopaminergic neurons in the basal ganglia.

The aim of this study was to assess the presence of prodromal PD in 22q11DS through a cross-sectional study in adults with MLPA-proven deletions, using the Movement Disorder Society (MDS) research criteria for prodromal PD (Berg et al 2015). Specifically, RBD was evaluated by polysomnography; olfactory dysfunction by Sniffin' Sticks test; dysautonomia by COMPASS31 questionnaire; subtle motor signs by the MDS-UPDRS scale, and integrity of dopaminergic pathways by PET/CT imaging with [<sup>18</sup>F]PR04.MZ, a highly selective dopamine transporter (DAT) ligand. Results of these assessments were used to calculate likelihood ratio (LR) of having prodromal PD according to the MDS criteria. In addition, psychiatric evaluation with the MINI-International Neuropsychiatric Interview version 5.0 (MINI) and cognitive assessment using WAIS-IV were performed. To date, 21 adults (11 women and 10 men) have completed the evaluations, with a median age of 22.5 years (range 18-50 years). One participant (4.1%) had a very high positive LR (>1000) for prodromal PD, based on mild motor signs and decreased DAT signaling in caudate and putamen by PET/CT (specific binding ratio < 65% of age-matched controls). In addition, reduced DAT signaling was most severe in posterior putamen and asymmetric, as seen in early or prodromal PD. This study identified a 22q11DS patient with features highly indicative of prodromal PD and suggestive of neurodegeneration. A larger number of patients and longitudinal follow up are needed to assess the progression to PD and to identify high-risk individuals who may benefit from potential neuroprotective strategies. Funded by Fondecyt Chile grant 1171014.

## PgmNr 1281: Clinical and molecular spectrum of *WWOX*-related epileptic encephalopathy (WOREE).

### Authors:

Y. Cao <sup>1,2,3,8</sup>; J. Chong <sup>3,4,8</sup>; S. Kleppe <sup>5</sup>; K. Gripp <sup>6</sup>; J. Hertecant <sup>7</sup>; J. Rosenfeld <sup>1</sup>; J. Dong <sup>2</sup>; X. Wang <sup>1</sup>; M. Miller <sup>1,2</sup>; W. Bi <sup>1,2</sup>; P. Liu <sup>1,2</sup>; F. Scaglia <sup>1,2,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 2) Baylor Genetics, Houston, TX, USA; 3) Department of Obstetrics and Gynecology, The Chinese University of Hong Kong, Hong Kong SAR, China; 4) Department of Pediatrics, The Chinese University of Hong Kong, Hong Kong SAR, China; 5) Hospital Italiano de Buenos Aires, Buenos Aires, Argentina; 6) Division of Medical Genetics, A. I. du Pont Hospital for Children/Nemours, Wilmington, DE, USA; 7) Division of Genetic and Metabolic Disorders, Departments of Pediatrics, Tawam Hospital, Al Ain, United Arab Emirates; 8) Joint BCM-CUHK Center of Medical Genetics, Prince of Wales Hospital, Hong Kong SAR, China

---

Biallelic *WWOX* pathogenic variants are reported to cause severe infantile onset epileptic encephalopathy (WOREE syndrome), or spinocerebellar ataxia. Here we present clinical and molecular data on four subjects with WOREE syndrome who were referred for clinical evaluation. All four patients were detected to have biallelic loss of function pathogenic variants and/or exonic deletions in the *WWOX* gene. WOREE syndrome is characterized by refractory epilepsy, generalized hypotonia, cerebral atrophy, hypoplastic corpus callosum, delayed myelination on brain MRI, and premature death which were also observed in our patients. Two of our patients died around 2 years of age. The causes of death were related to recurrent infections and aspiration pneumonia in early life. Moreover, our patients present additional clinical features that expand the clinical phenotypic spectrum of WOREE syndrome. Even though anticonvulsants do not control the epilepsy found in this neurogenetic syndrome, the use of a ketogenic diet improved seizure control in two of our patients. None of our patients presented with facial dysmorphic features in contrast to what has been reported in the published literature. Moreover, although these subjects exhibited normal birthweight and head circumference at birth, all of them eventually developed failure to thrive and three manifested acquired microcephaly with progressive deceleration of head growth. Unlike previously reported patients, there was an absence of prenatal findings except for club feet in one patient. Two patients also presented bilateral sensorineural hearing loss which has not been frequently reported in this syndrome. Recurrent pancreatitis was seen in one of the patients which has not been observed before. In summary, we report four patients with WOREE syndrome to further document and delineate the clinical phenotypes of this disorder. This case series expand the clinical phenotype associated with this condition as some patients responded to a ketogenic diet, facial dysmorphic features were absent and prenatal findings were not present, pancreatitis presented in one of the patients, sensorineural hearing loss was observed in two subjects, and most of these subjects developed failure to thrive and presented acquired microcephaly. Long term follow-up of a larger cohort of survivors with WOREE syndrome will provide further clinical insight into the cardinal features of this syndrome.



## PgmNr 1282: Bi-allelic *PDXK* mutations cause polyneuropathy with optic atrophy responsive to PLP supplementation.

### Authors:

V. Chelban<sup>1,6</sup>; M. Wilson<sup>2</sup>; J. Chardon<sup>3</sup>; J. Vandrovcova<sup>1</sup>; E. Zamba?Papanicolaou<sup>4</sup>; M. Osmond<sup>5</sup>; N. Wood<sup>1</sup>; K. Boycott<sup>5</sup>; P. Mills<sup>2</sup>; P. Clayton<sup>2</sup>; H. Houlden<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UCL Queen Square Institute of Neurology, University College London, London, United Kingdom; 2) Genetics and Genomic Medicine, GOS Institute of Child Health, University College London, London, United Kingdom; 3) Ottawa Hospital Research Institute, University of Ottawa, Canada; 4) The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 5) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada; 6) Department of Neurology and Neurosurgery, Institute of Emergency Medicine, Chisinau, Republic of Moldova

---

**Aim:** To identify disease-causing variants in autosomal recessive axonal polyneuropathy with optic atrophy and provide targeted replacement therapy.

**Methods:** We performed genome-wide sequencing, homozygosity mapping and segregation analysis for novel disease-causing gene discovery. We used circular dichroism to show secondary structure changes and isothermal titration calorimetry to investigate the impact of variants on ATP-binding. Pathogenicity was further supported by enzymatic assays and mass spectroscopy on recombinant protein, patient-derived fibroblasts, plasma and erythrocytes. Response to supplementation was measured with clinically validated rating scales, electrophysiology and biochemical quantification.

**Results:** We identified bi-allelic mutations in *PDXK* in five individuals from two unrelated families with primary axonal polyneuropathy and optic atrophy. The natural history of this disorder suggests that untreated, affected individuals become wheelchair-bound and blind. We identified conformational rearrangement in the mutant enzyme around the ATP-binding pocket. Low *PDXK* ATP-binding resulted in decreased erythrocyte *PDXK* activity and low pyridoxal 5'-phosphate (PLP) concentrations. We rescued the clinical and biochemical profile with PLP supplementation in one family, improvement in power, pain and fatigue contributing to patients regaining their ability to walk during the first year of PLP normalization.

**Interpretation:** We show that mutations in *PDXK* cause autosomal recessive axonal peripheral polyneuropathy leading to disease via reduced *PDXK* enzymatic activity and low PLP. We show that the biochemical profile can be rescued with PLP supplementation associated with clinical improvement. As B<sub>6</sub> is a cofactor in diverse essential biological pathways, our findings may have direct implications for neuropathies of unknown aetiology characterised by reduced PLP levels.



## PgmNr 1283: Mutational screening of genes associated with hereditary sensory and autonomic neuropathy in a Brazilian population.

### Authors:

V.P. Cintra <sup>1</sup>; P.J. Tomaselli <sup>1</sup>; F.B. Figueiredo <sup>1</sup>; S.E. Marques <sup>1</sup>; S.T. Camargos <sup>2</sup>; A. Rebelo <sup>3</sup>; L. Abreu <sup>3</sup>; M. Danzi <sup>3</sup>; S. Zuchner <sup>3</sup>; W. Marques Jr <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ribeirao Preto School of Medicine, University of Sao Paulo, Ribeirao Preto, Brazil; 2) Federal University of Minas Gerais, Belo Horizonte, Brazil; 3) Dr John T. Macdonald Foundation Department of Human Genetics, John P. Hussman Institute of Human Genomics, University of Miami Miller School of Medicine, Miami, USA

---

Hereditary sensory and autonomic neuropathies (HSAN) are a group of rare genetically and phenotypically heterogeneous disorders characterized by axonal atrophy and degeneration often complicated by ulcers and amputations, with variable motor and autonomic involvement. To date, at least 17 different genes have been associated with HSAN (*SPTLC1*, *SPTLC2*, *ATL1*, *DNMT1*, *ATL3*, *SCN11A*, *HSN2/WNK1*, *FAM134B*, *KIF1A*, *SCN9A*, *IKBKAP*, *NTRK1*, *NGF-β*, *DST*, *PRDM12*, *ZFHX2*, *CLTCL1*). The relative frequency of these disorders varies among ethnic and geographical populations probably due to founder effects. However, the epidemiology of HSAN in Brazil is still unknown, since a systematic study has not been conducted yet. In order to evaluate the frequency of mutations in these genes, we performed a systematic mutation screening of their coding sequences in 25 unrelated Brazilian patients diagnosed with HSAN. We identified 10 patients belonging to three unrelated families with mutations in three HSAN disease associated genes: *ATL3* (p.Tyr219Cys), *SCN9A* (p.Trp714Ter; p.Glu982Ter) and *SPTLC2* (p.Asn177Asp). The clinical features associated with mutations in *ATL3* and *SPTLC2* typically consisted of dominant inheritance, frequent traumas in toes with repeated painless infections, including deep foot ulcerations, osteomyelitis, osteonecrosis and acromutilations. We detected a novel mutation in *SCN9A* (p.Trp714Ter) in compound heterozygous state with the previously described p.Glu982Ter pathogenic variant, leading to an early-onset HSAN phenotype with impairment of pain, temperature and touch sensation, and autonomic involvement, such as urinary incontinence, intestinal constipation and hypohidrosis. We thus identified a pathogenic mutation in 12% of HSAN Brazilian families, which suggests that the vast majority of the cases would be explained by other unknown mutated genes that remain to be discovered. We are currently performing whole genome studies on all remaining cases. Our study characterizes the spectrum of HSAN mutations found in Brazilian patients, and provides additional insights for genotype-phenotype association of the identified mutations.

## PgmNr 1284: After the clinical exome: Identifying the genetic etiology of refractory pediatric epilepsies.

### Authors:

J.D. Calhoun<sup>1</sup>; J. Gunti<sup>1</sup>; E. Bryant<sup>2</sup>; E. Spinelli<sup>2</sup>; E. Rojas<sup>2</sup>; L. Laux<sup>2</sup>; M.A. Nobrega<sup>3</sup>; T.S. Gertler<sup>2</sup>; J.J. Millichap<sup>2</sup>; G.L. Carvill<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 3) The University of Chicago, Chicago, IL

---

The majority of early-onset, refractory epilepsies are due to *de novo* or inherited genetic variation. Genetic testing has demonstrated that pathogenic single nucleotide variants, indels and structural variants account for ~30-50% of early-onset epilepsies. Despite this strong prevailing genetic contribution, the majority of pediatric epilepsy patients never receive a genetic diagnosis. We developed a tiered approach utilizing exome reanalysis and genome sequencing to enable genetic diagnosis in exome negative trios.

We recruited a cohort of 25 exome negative pediatric epilepsy cases primarily consisting of individuals with early-onset, drug refractory seizures. By reanalysis of clinical exome data we identified likely pathogenic or candidate variants in 7 probands. The majority of these were variants in novel genes or genes recently implicated in epilepsy and/or related neurodevelopmental disorders. For instance, we identified a patient carrying a *de novo* missense variant of uncertain significance in *MAST3*. Through an international collaboration we identified an additional seven individuals with *MAST3* missense variants, two of which were recurrent and all were located in the ST kinase domain.

We have pursued genome sequencing in the remaining 18 trios to identify non-coding pathogenic variants. We have developed computational pipelines to detect all major classes of genetic variants and prioritize these with appropriate inheritance models. We hypothesize that genetic variation disrupting functional neuronal *cis*regulatory modules (CRMs) contributes to epilepsy susceptibility by modulating expression of genes involved in neuronal development or function. Promoter capture Hi-C is performed to generate neuronal CRMs in human induced pluripotent stem cell derived neuronal cell types, including neuronal progenitor cells, excitatory and inhibitory neurons. Genetic variants ablating or disrupting neuronal CRMs will be prioritized as candidates.

In summary, we applied a tiered approach to identify genetic variants in epilepsy patients with negative clinical exomes. In this cohort 7/25 (28%) individuals carried likely pathogenic or candidate variants in genic regions. We have designed a computational framework implementing CRMs to prioritize variants outside the exome. Future work will allow us to determine whether variants in CRMs account for the unsolved (72%) cases in our cohort. This framework may have applications to neurodevelopmental disorders more broadly.

## PgmNr 1285: Novel homozygous variants in *CACNA2D2* in patients with epileptic encephalopathy and cerebellar atrophy.

### Authors:

I. Herman <sup>1</sup>; J. Punetha <sup>2</sup>; E. Karaka <sup>2,3</sup>; A. Gezdirici <sup>4</sup>; R. Lamont <sup>5</sup>; D. Pehlivan <sup>1,2</sup>; D. Marafi <sup>2</sup>; J. Appendino <sup>6,7</sup>; J. Hunter <sup>6</sup>; Z. Coban-Akdemir <sup>2</sup>; J. Fatih <sup>2</sup>; S. Jhangiani <sup>8</sup>; R. Gibbs <sup>2,8</sup>; A. Innes <sup>5,6</sup>; J. Posey <sup>2</sup>; J. Lupski <sup>2,5,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Section of Child Neurology and Basic Neuroscience, Texas Children's Hospital and Baylor College of Medicine, Houston, TX.; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 3) Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; 4) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey; 5) Department of Medical Genetics and Alberta Children's Hospital Research Institute, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; 6) Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; 7) Clinical Neuroscience, Department of Pediatrics, Alberta Children's Hospital, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 9) Department of Pediatrics, Baylor College of Medicine, Houston, TX

**Background:** Developmental and epileptic encephalopathies (DEE) constitute a spectrum of disorders characterized by intractable epilepsy with prominent epileptiform activity and electroencephalography, developmental delay/intellectual disability, and neurodevelopmental regression. Many patients with DEE have an underlying genetic etiology with now more than 100 genes identified to date. Implicated genes have been shown to participate in synaptic transmission, ion channel function, transcriptional regulation, DNA damage repair, and metabolic pathways. Genes implicated in ion channel functioning are the most common etiology of DEEs, leading to channelopathies. Identification of an underlying channelopathy has important therapeutic implications. Specifically, voltage-gated calcium channel subunits encoded by the family of *CACNA* genes has been linked to epileptic encephalopathy and/or ataxia. To date, four unrelated cases with *CACNA2D2* variants have been reported. Here, we report two additional families with DEE and cerebellar atrophy who were found to have novel homozygous variants in *CACNA2D2*.

**Materials and Methods:** Two affected individuals from family #1 with clinical features of early onset epileptic encephalopathy were recruited for exome sequencing at the Centers for Mendelian Genomics to identify their molecular diagnosis. We identified family #2 through collaboration via GeneMatcher. the index had undergone panel testing for the underlying epilepsy.

**Results:** We found three individuals in two unrelated families with novel homozygous rare variants in *CACNA2D2* with clinical features of DEE and cerebellar atrophy. Family #1 includes two affected siblings with a likely damaging homozygous rare missense variant c.1778g>c.p.(Arg593Pro) and family #2 includes a proband with a homozygous rare nonsense variant c.485\_486del;p.(Tyr162Tyr) in *CACNA2D2*. We compared clinical and molecular findings from all individuals from six unrelated families reported to date and noted that cerebellar atrophy is shared among all.

**Conclusion:** Our study supports the candidacy of *CACNA2D2* as a disease gene associated with DEE

and cerebellar atrophy.

## PgmNr 1286: *GRM7* causes a severe neurodevelopmental disorder characterized by microcephaly, refractory epilepsy, and global developmental delay.

### Authors:

D.N. Marafi <sup>1</sup>; T. Mitani <sup>1</sup>; S. Isikay <sup>2</sup>; J. Hertecant <sup>3</sup>; M. Holick <sup>4,5</sup>; E. Karaca <sup>6</sup>; Y. Bayram <sup>7</sup>; J. Fatih <sup>1</sup>; J. Punetha <sup>1</sup>; J.A. Rosenfeld <sup>1</sup>; S.N. Jhangiani <sup>8</sup>; D. Pehlivan <sup>1,4,5</sup>; E. Boerwinkle <sup>8</sup>; R.A. Gibbs <sup>1,8</sup>; J.E. Posey <sup>1,5</sup>; J.R. Lupski <sup>1,5,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA; 2) Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, School of Health Sciences, Gaziantep, 27000, Turkey; 3) Pediatric Metabolic and Genetics Department, Tawam Hospital, United Arab Emirates; 4) Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA; 5) Texas Children's Hospital, Houston, Texas, 77030, USA; 6) Department of Genetics, University of Alabama at Birmingham, Birmingham, AL35294; 7) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, 77030, USA; 9) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA

### Introduction:

Defects in ion channels and neurotransmitter receptors are implicated in neurodevelopmental disorders (NDDs) and epilepsy. One such receptor is metabotropic glutamate receptor 7 (mGluR7) encoded by *GRM7*. MGLuR7 is the most highly conserved mGluR receptor and is broadly expressed in the brain. This presynaptic G-coupled receptor is critical for synaptic transmission and is recruited under high neurotransmitter concentrations to inhibit further release of excitatory neurotransmitter glutamate and the inhibitory neurotransmitter GABA.

MGLuR7-/- knockout mice develop spontaneous stimulus-provoked seizures and have reduced seizure threshold. We have previously proposed *GRM7* as a novel candidate disease gene in 4 individuals from 2 unrelated families with NDDs. One additional family has been reported since.

### Methods:

By performing family-based exome sequencing on a cohort of 185 mostly consanguineous Turkish families with NDDs, we identified 2 additional unrelated individuals with rare biallelic *GRM7* variants. Gene query to Baylor Genetics Laboratory yielded 2 more families with rare *GRM7* variants. We compare the observed clinical features and variants of 12 individuals from all 7 unrelated families, providing additional evidence to support *GRM7* as a disease-causing gene in NDDs and epilepsy.

### Results:

Novel deleterious variants in the 4 newly described families include two homozygous missense variants (c.2671G>A:p.E891K and c.1973G>A:p.R685Q), one homozygous stop-gain variant (c.1975C>T:p.R659\*) and compound heterozygous frameshift and missense variants (c.37\_38delTT:p.L13fs and c.895G>A:p.A299T). All 3 new homozygous variants were within a large Absence of Heterozygosity block (>5Mb). All 7 families had unique variants and were of different ethnicities.

Of the 12 affected individuals, neurodevelopmental delay was present in all, refractory epilepsy in 11/12 (92%), microcephaly in 9/12 (75%), and axial hypotonia in 8/12 (67%). Brain imaging showed cortical atrophy in 10/12 (83%), corpus callosum hypoplasia in 7/12 (59%), cerebellar atrophy in 5/12 (42%), and hypomyelination in 4/12 (33%) individuals. Four subjects died due to either Sudden Infant Death Syndrome, aspiration, or due to an unknown cause.

**Conclusion:**

We provide detailed clinical characterization in 7 unrelated families with rare biallelic *GRM7* variants, including 5 novel *GRM7* variants. Our study establishes *GRM7* as a potential disease-causing gene in NDDs, microcephaly, and epilepsy.

## **PgmNr 1287: *De novo* ZBTB47 missense variants in two individuals with generalized epilepsy, developmental delays, and movement abnormalities.**

### **Authors:**

J.A. Rosenfeld <sup>1</sup>; A.F. Wadley <sup>2</sup>; W.J. Craigen <sup>1</sup>; K. Fisher <sup>1</sup>; L.C. Burrage <sup>1</sup>; D.R. Murdock <sup>1</sup>; B. Lee <sup>1</sup>; A.C. Tsai <sup>2</sup>; L. Emrick <sup>1</sup>; Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Baylor College of Medicine, Houston, Texas.; 2) University of Oklahoma Health Sciences Center, Oklahoma City, OK

---

Genetic abnormalities are thought to cause a majority of epilepsies, and next-generation sequencing technologies are providing an increasing number of discoveries of such genetic causes. Genes causing epilepsies and movement disorders encode a variety of proteins that play important roles in the central nervous system, including ion channels, neurotransmitters and receptors, transcriptional regulators, and regulators of mTOR. Through trio exome sequencing we identified *de novo* missense variants in *ZBTB47*, which encodes a minimally characterized transcriptional regulator, in two females with generalized epilepsy, developmental delays, and movement abnormalities. The severity of epilepsy was variable, with one girl experiencing up to ~60 tonic seizures per day, despite the use of anti-epileptic medications. This patient also experienced chorea with onset in infancy, while the other patient showed choreiform movements, shuddering attacks and an ataxic gait. Both experienced significant hypotonia. MRIs were normal. The *ZBTB47* variants both affect conserved residues within zinc finger domains of the protein, suggesting that they may disrupt the protein's ability to properly bind DNA. Therefore, *ZBTB47* may provide yet another example of transcriptional dysregulation causing an abnormal neurodevelopmental phenotype.

## PgmNr 1288: A case of UDP-galactose transporter deficiency (SLC35A2-CDG) response to ketogenic diet to control intractable seizures.

### Authors:

S. Yano <sup>1</sup>; A. Partikian <sup>2</sup>; J. Lowe <sup>2</sup>; K. Moseley <sup>1</sup>; A. Fujimoto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics/Gen Div, LAC+USC Med Ctr, USC, Los Angeles, CA; 2) Pediatrics/Neuro Div, LAC+USC Med Ctr, USC, Los Angeles, CA

---

**Background:** Congenital disorders of glycosylation (CDG) are a group of genetic disorders characterized by abnormal glycosylation of a variety of proteins and lipids. Defects of the X-linked gene SLC35A2, which encodes the Golgi-localized UDP-galactose transporter, result in decreased UDP-galactose transport into the Golgi apparatus leading to abnormal glycosylation. Patients with this condition (formerly CDG-IIIm) show varying degrees of neurological impairments, epilepsy, and congenital anomalies. So far, approximately 60 cases have been reported. Some of the patients had early-onset epileptic encephalopathy. **Case Study:** A 23 year-old Hispanic female with a pathogenic variant of the SLC35A2 gene (c.348delA) has been followed since age 18 Mo. Hypotonia was noted at birth. Microcephaly, and fingers/toes anomalies were noted at age 18 Mo. She experienced her first seizures at age 2 Mo and did not gain any milestones for the first 18 months. She had normal transferrin glycosylation at age 14 y. Developing breakthrough tonic/convulsive seizures with recurrent status epilepticus requiring hospitalization at age 18 y, ketogenic diet was introduced emergently after failure of multiple antiepileptic drugs. The diagnosis of CDG-IIIm was made by the NGS at age 22 y. Brain MRI at age 18 y revealed diffuse, chronic cerebral/cerebellar volume loss prominent in the frontal/temporal lobes bilaterally, bilateral hippocampal dysplasia, with hyperintense MRI signal in the basal ganglia on T2 sequences. The patient's frequent and drug-resistant, generalized tonic seizures responded very well to 2.4:1 ketogenic ratio diet. Her seizure frequency dramatically improved from several tonic seizures each day to very brief spells lasting seconds occurring a few times each week on the ketogenic diet. She continued to have well-controlled tonic seizures for 3 years until age 22 y, with some improvement in attention and interactiveness, with later re-emergence of more prolonged and frequent seizures necessitating escalation of anticonvulsant interventions. **Discussion:** Galactose supplementation showed an improved glycosylation of transferrin in a case of SLC35A2. An additional UDP-galactose transporter was suggested. It is unclear why the ketogenic diet significantly improved seizure control in this patient. Collection and further investigation of other cases of SLC35A2 treated with the ketogenic diet are needed to understand the potential mechanisms of dietary therapy in this CDG.



## **PgmNr 1289: X-linked recessive Chronic Intestinal Pseudo-Obstruction (CIPO): Male patient and his unaffected mother with a frameshift/truncating variant in exon 1 of *FLNA*.**

### **Authors:**

J. Pappas <sup>1</sup>; J. Levy <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Dept Pediatrics, Clin Genetic Scvs, New York Univ, Sch Med, New York, New York.; 2) Dept Pediatrics, Gastroenterology, New York Univ, Sch Med, New York, New York.

---

Pathogenic variants in *FLNA* have been associated with different X-linked syndromes. Loss of function (LoF) variants are associated with periventricular nodular heterotopia (PVNH) in females and embryonic lethality in males (Eksioglu, Y Z et al, 1996). There are rare reports of males with LoF variants in *FLNA* in the medical literature with a combination of PVNH, cardiovascular abnormalities, Ehlers Danlos and Chronic Intestinal Pseudo-Obstruction (CIPO) or Congenital Short Bowel Syndrome (CSBS) (Gargiulo A et al., 2007; Kapur RP et al 2010; van der Werf CS et al., 2012; Oda H et al 2016; Jenkins ZA et al 2017). We present a 5 year old boy with CIPO and mild neurodevelopmental delays. He was referred to us at 2 months of age because of unexplained distention and diarrhea after feeding, treated with parenteral nutrition. He presented at 3 days with vomiting and he had surgery for pyloric stenosis. He was diagnosed with intestinal malrotation treated with exploratory laparotomy and Ladd's bands release in infancy. Two maternal uncles had similar histories. Parenteral nutrition discontinued at 14 months and his bowel function started improving. At age 5, his height was at the 23<sup>rd</sup> centile while his weight was at 93<sup>rd</sup>. He tolerated different foods by mouth with no vomiting but occasional diarrhea. His neurological examination was normal. The family resembled the report by Gargiulo A et al., 2007 and *FLNA* (NM\_001110556.1) genetic testing performed in infancy revealed an insertion of T in exon 1: c.10dupT (p.Ser4Phe\*102) by Sanger sequencing. The unaffected mother carried this variant. Our patient's variant and clinical presentation was similar to the two families reported in the medical literature with two base deletion in exon 1 leading to frameshift and stop codon (Gargiulo A et al., 2007; Jenkins ZA et al 2017). CIPO was the main abnormality of affected males in our family and both families in the literature and all the carrier females were unaffected. The frameshift variants in our family and both of the families in the medical literature were in exon 1 and predicted to disrupt the expression of *FLNA*<sup>+1</sup> isoform; the expression of *FLNA*<sup>+28</sup> was predicted normal (Jenkins ZA et al 2017). *FLNA*<sup>+1</sup> is predominantly expressed in the intestines and *FLNA*<sup>+28</sup> in many tissues including the brain. Our patient and his family provide additional evidence that frameshift/truncating variants in exon 1 of *FLNA* are associated with CIPO in males and cause no abnormality in females.

## PgmNr 1290: Identification of novel variants causing leukodystrophies in Sudan.

### Authors:

M. Amin <sup>1,2</sup>; I. Dorboz <sup>2</sup>; A. Babai <sup>1</sup>; I. Mohammed <sup>1</sup>; A. Hamed <sup>1</sup>; M. Elseed <sup>1</sup>; E. Emad <sup>1</sup>; E. Badi <sup>1</sup>; R. Abubakr <sup>1</sup>; D. Mustafa <sup>1</sup>; M. Osman <sup>1</sup>; Y. Hasan <sup>1</sup>; A. Elshafea <sup>1</sup>; M. Siddig <sup>1</sup>; S. Haron <sup>1</sup>; L. Elsayed <sup>1</sup>; A. Ahmed <sup>1</sup>; O. Boespflug-Tanguy <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculty of Medicine, University of Khartoum, Khartoum, Sudan; 2) INSERM UMR 1141 PROTECT, Université Paris Diderot- Sorbonne Paris Cité, France; 3) Neuropediatrics and Metabolic Disorders Department, reference center for leukodystrophies and rare leukoencephalopathies (LEUKOFRANCE), CHU APHP Robert-Debré, Paris, France

---

**Background:** Leukodystrophies are inherited disorders primarily affecting the white matter of the CNS with or without peripheral nerve involvement. There are currently 30 recognized forms of leukodystrophies with distinct clinical, biochemical and radiological characteristics. However, the genetics of these classical forms of leukodystrophies remain unknown in many parts of the world especially in Sub-Saharan Africa. This study aimed to shed the light on the clinical and genetic characteristics of leukodystrophies in Sudanese families.

**Methods:** Fifty two families with suspected leukodystrophy from different regions of Sudan were analyzed in the study. Genomic DNA was extracted and screened for mutations using NGS panel testing 153 leukodystrophies and leukoencephalopathies causing genes (NextSeq500 Illumina).

**Results and Discussion:** Nine novel variants were discovered: four variants causing Megalencephalic leukodystrophy with subcortical cysts (which was found to be the most common type of leukodystrophy in Sudan), two variants causing Metachromatic leukodystrophy, two variants causing Leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation and one variant causing Spastic paraplegia 51. We hypothesized that the higher percentage of MLC is caused by the adaptation of heterozygous genotype to hot, dry environments like Sudan (heterozygote superiority).

**Conclusion:** This is the first study to highlight the genetics of leukodystrophies in Sudan. Analysis of additional families is in progress in order to establish the whole spectrum of genetic variations causing inherited leukodystrophies in Sudanese families.

## PgmNr 1291: Nonsense and missense variants in *CAMTA1* gene are associated with developmental delay/intellectual disability, gait abnormalities, and hypotonia.

### Authors:

H. Al-Kateb <sup>1</sup>; M. Frank <sup>2</sup>; M. Willing <sup>3</sup>; O. Lichtarge <sup>4</sup>; P. Katsonis <sup>4</sup>; P.Y. Billie Au <sup>5</sup>; M. Shinawi <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Cincinnati Children's Hospital, Cincinnati, Ohio.; 2) Eastern Virginia Medical School, Children's Hospital of The King's Daughters 850 Southampton Ave Norfolk, VA 23510; 3) Washington University School of Medicine One Children's Place, Northwest Tower, 9132; 4) Department of Molecular and Human Genetics Baylor College of Medicine; 5) University of Calgary Alberta Children's Hospital

---

Calmodulin binding transcription activator 1 (*CAMTA1*) encodes a transcription factor and a tumor suppressor. Pathogenic intragenic copy number variation (CNVs) in *CAMTA1* have been associated with nonprogressive cerebellar ataxia with mental retardation (CANPMR; MIM# 614756). The clinical manifestations include development delay (DD)/intellectual disability (ID), ataxia, dysarthria, and poor fine motor skills with/out brain abnormalities. Here, we describe 6 additional patients from unrelated families with overlapping phenotypic features, but who were found to have missense, nonsense variants, or an intragenic duplication in *CAMTA1* gene.

### Patients:

All patients in our cohort (F/M:3/3; 4 Caucasians, 1 Arab and 1 Russian) had motor and speech delay and/or ID with variable severity as well as gait abnormalities ranging from "clumsiness" to ataxia. Five patients had hypotonia and 4 had minor dysmorphic features but without recognizable facial gestalt. Four patients had eye abnormalities, 3 had brain MRI abnormalities, and 2 had difficulties in fine motor skills. The two carrier parents (one of whom is mosaic) had overlapping but milder phenotypes.

### Material and Methods

Exome sequencing was performed on patients (1-4 and 6) and chromosomal microarray analysis was performed on patients 1-3, 5-6.

### Molecular results

Patients 1 and 2 had *de novo* missense variants (p.Y1525C and p.K141N) and patient 3 had a paternally inherited missense variant (p.G500A). The CADD scores and p values for VEST analysis for these variants were 26.2, 23.9 and 26, and 0.0060, 0.019 and 0.077, respectively. The father carrying the p.G500A variant exhibits ID and tremor. Patients 4 and 5 had a *de novo* frameshift and a maternally inherited stop-gain variants (p.E1368GfsX11, p.Gln772\*), respectively, and patient 6 had a duplication of exons 12-16 of *CAMTA1* gene resulting in a frameshift and stop codon after introducing 21 novel amino acids.

### Conclusions

The identification of pathogenic nonsense variants in *CAMTA1* gene in patients with DD/ID, hypotonia, gait and balance abnormalities provides additional evidence that *CAMTA1* has a role in normal brain and cerebellum function and is dosage-sensitive. We speculate that the missense variants cause loss of function as well but this is less clear. Larger cohorts are needed to examine the full spectrum of clinical manifestation and genotype-phenotype correlation. Further studies are required to investigate

the exact role of CAMTA1 in brain development.

## **PgmNr 1292: 11-year-old male with cerebral palsy, seizures, microcephaly, and intellectual disabilities with compound heterozygosity in the *SRCIN1* gene: Candidate gene for a new syndrome.**

### **Authors:**

S. Berger; A. Iglesias

[View Session](#) [Add to Schedule](#)

**Affiliation:** Columbia University Medical Center , New York, NY.

---

Our patient initially presented to clinic at age 11 with spastic cerebral palsy, intellectual disabilities, seizures, microcephaly, contractures, strabismus, emaciation, and reduced muscle bulk. He was born full term after an uneventful pregnancy. He is of Dominican descent and consanguinity was denied. He has a history of seizures since shortly after birth, with profound developmental delay and intellectual disability since an early age, without regression. He is nonverbal and unable to stand without support.

At the time of our consult, he was not seeing any other specialists or receiving any services. He had failure to thrive and poor seizure management with seizures 4-5 times per day. We referred him to multiple specialists at our medical center. He now has a G-tube for feeding and medication, is followed by multiple specialists, and is on multiple anti-epileptics, with an average of 1 seizure per day. Brain MRI shows encephalomalacia. EEG shows epileptic encephalopathy, likely Lennox-Gastaut syndrome.

Physical exam was positive for micro-dolicocephaly, long face, synophrys, strabismus, high arched palate, mild scoliosis, contractures of elbows, knees, wrists and fingers, reduced muscle bulk, axial hypotonia and distal spasticity.

WES identified compound heterozygous variants of uncertain significance in the *SRCIN1* gene. This is currently a candidate gene with no known genetic condition associated with variants in this gene and no known mode of inheritance. Studies have shown that this gene is expressed in the spinal cord and the brain, but further research is needed to clarify if there is a relationship between variants in this gene and his phenotype. We have reached out to a group who is researching this gene and compiling cases.

Incidentally, a microarray showed a 1.7Mb duplication of Xp22.31 which is classified as a variant of uncertain significance. Similar or larger duplications of Xp22.31 have been reported in individuals with neurodevelopmental phenotypes, congenital anomalies, and dysmorphic features. However, these duplications have also been reported in clinically healthy individuals. Therefore, the clinical significance of this variant remains uncertain at this time. No maternal sample is available to determine inheritance.

In summary, we are reporting a patient with a unique phenotype and compound heterozygosity in a

candidate gene. Further assessment of similar patients, as is currently done, will clarify the true nature of this finding.

## PgmNr 1293: Identification of a novel candidate gene in a patient presenting with a neuromuscular disease.

### Authors:

R.J. Olson <sup>1</sup>; B. Celona <sup>2</sup>; M.A. Cousin <sup>1</sup>; T. Kruisselbrink <sup>1</sup>; B. Lanpher <sup>1</sup>; A. Sigafos <sup>3</sup>; K. Clark <sup>3</sup>; B. Black <sup>2</sup>; E.W. Klee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Individualized Medicine, Mayo Clinic, Rochester MN 55905; 2) Department of Biochemistry and Biophysics, UCSF, San Francisco CA 94143; 3) Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester MN 55905

---

Pediatric neuromuscular disorders (NMD) are characterized by a genetically heterogeneous etiology with diverse pathomechanisms. NGS techniques have allowed for the discovery of novel causative genes in this family of disorders. However, many patients still remain genetically unresolved. Here we describe an adolescent patient from consanguineous, unaffected parents who clinically presented with a muscle dystrophy in a limb-girdle pattern at the age of 11. While being asymptomatic in her first decade of life, her symptoms quickly progressed to her requiring assistance with walking and eating. MRI of spine and brain were unremarkable, however EMG studies showed chronic diffuse disorder of her motor neurons. Her lab work, including creatine kinase, was normal. Targeted genetic testing of suspected genes (including *SMN1*, *SMN2*) and microarray analysis was negative. Whole-exome sequencing identified a homozygous missense variant (c.637T>G p.Phe213Val) in exon 2 of *ZNF106*, mapped to chromosome 15q15.1. The p.Phe213 residue is highly conserved across species and *in silico* analysis predicts the Val substitution to be deleterious; with a CADD score of 24.3. *ZNF106* has not been associated with disease, however, recent studies have shown that homozygous loss in mice causes a progressive degeneration of motor neurons and overexpression partially rescues the phenotype in a fruit fly model of *C9orf72* ALS. Furthermore, gene expression is largely restricted to skeletal muscle and recent studies have shown that the ZNF106 protein interacts with RNA, RNA binding proteins and is putatively involved in RNA splicing. Interestingly, pathway analysis of significantly dysregulated genes ( $P \leq 0.05$ ) from patient blood RNAseq data compared to unaffected individuals showed an association to familial amyotrophic lateral sclerosis ( $P = 0.000251$ ), further indicating the role of *ZNF106* in neuromuscular disease. Nonetheless, it is not known whether the p.Phe213Val missense variant in our patient disrupts its putative function and causes disease. We are modeling our patient's missense variant in mouse and zebrafish to understand its putative pathogenicity. In addition, *in vitro* studies will be performed to test the effect of the missense variant on localization, interactome and function of the ZNF106 protein. These studies may bring about a new genotype-phenotype correlation into the spectrum of NMDs and thereby reveal novel pathomechanisms.

## PgmNr 1294: GGC repeat expansion in *NOTCH2NLC* is the cause of neuronal intranuclear inclusion disease.

### Authors:

J. Sone <sup>1,2</sup>; S. Mitsuhashi <sup>3</sup>; A. Fujita <sup>3</sup>; K. Mori <sup>4</sup>; H. Takashima <sup>5</sup>; H. Sugiyama <sup>6</sup>; Y. Kohno <sup>7</sup>; Y. Takiyama <sup>8</sup>; K. Maeda <sup>9</sup>; F. Tanaka <sup>10</sup>; Y. Iwasaki <sup>11</sup>; M. Yoshida <sup>11</sup>; N. Matsumoto <sup>3</sup>; G. Sobue <sup>2,12</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurology, National Hospital Organization Suzuka National Hospital, Suzuka, Mie, Japan; 2) Neurology, Nagoya University Graduate School of Medicine, Nagoya, Japan.; 3) Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 4) Neurology, Oyamada Memorial Spa Hospital, Yokkaichi, Japan; 5) Neurology and Geriatrics, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan; 6) Neurology, National Hospital Organization Utano National Hospital, Kyoto, Japan; 7) Neurology, Ibaraki Prefectural University of Health Sciences, Ibaraki, Japan; 8) Neurology, University of Yamanashi, Chuo, Yamanashi, Japan; 9) Neurology, National hospital organization Higashi-Ohmi General Medical Center, Higashi-Ohmi, Japan; 10) Neurology and Stroke Medicine, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 11) Neuropathology, Institute for Medical Sciences of Aging, Aichi Medical University, Nagakute, Japan; 12) Brain and Mind Research Center, Nagoya University Graduate School of Medicine, Nagoya, Japan

---

Neuronal intranuclear inclusion disease (NIID) is a slowly progressive neurodegenerative disease characterized by eosinophilic intranuclear inclusions in the central and peripheral nervous system, and in the visceral organs. These intranuclear inclusions were ubiquitin and p62 positive by immunohistochemistry, and without limiting membrane under electron microscopic findings. These histopathological feature has been the hallmark of NIID and the basis of NIID diagnosis.

We started genetical study of NIID with familial NIID patients from two families diagnosed by autopsy. First, we studied linkage analysis with micro satellite markers and identified a peak showing highest LOD score 4.82 at 1p31.1-1q21.3, encompassing 68 Mb. We need to add another familial NIID patients to narrow down. But, the number of newly diagnosed NIID patients were few because autopsy was the only one means of diagnosis. To overcome this situation, we minutely examined NIID histopathological samples and find that skin biopsy was useful to the antemortem diagnosis of both familial and sporadic NIID patients.

After that, we diagnosed many NIID patients by skin biopsy and included them. We engaged in all exome and whole genome sequence with short-read next generation sequencer, but we could not find any nonsynonymous variations that segregate with NIID. We studied linkage analysis again with SNV data from short-read NGS, and its results showed highest LOD score 4.21 in almost same position of the result of linkage analysis with micro satellite markers. We concluded that the genetical cause of familial NIID must not be SNV and be exist within this high LOD score position, and we promoted long-read NGS study. Finally we include familial NIID patients from 9 families, and 40 sporadic NIID patients, and identified a GGC repeat expansion in *NOTCH2NLC* in all and only affected NIID patients.



We can now diagnose NIID more precisely with clinical, pathological and genetical means with the diagnostic flowchart of NIID we made of before. We will promote molecular biological study and radical treatment of NIID.

Aknowledgments: Drs. Haruki Koike, Akihiro Hashiguchi, Hiroshi Doi, Shigeru Koyano, Hideyuki Takeuchi, Michi Kawamoto, Nobuo Kohara, Tetsuo Ando, Toshiaki Ieda, Yasushi Kita, Norito Kokubun, Yoshio Tsuboi, Takeshi Mizuguchi, Kohei Hamanaka, Kazutaka Katoh, Yoshihiro Kino, Masahisa Katsuno, Ikuo K. Suzuki, Martin C. Frith are very much appreciated for their participation to this study.

## PgmNr 1295: Cohort of patients with novel *KCNK9* variants: Gain and loss of function cause *KCNK9* imprinting syndrome.

### Authors:

M.A. Cousin<sup>1,2</sup>; E.L. Veale<sup>3</sup>; M.T. Zimmermann<sup>4,5</sup>; V. Bhambhani<sup>6</sup>; M. Bialer<sup>7</sup>; M.R. Bekheirnia<sup>8</sup>; C. Boelman<sup>9</sup>; J. Carmichael<sup>10</sup>; B. Cogne<sup>11</sup>; D. Doummar<sup>12</sup>; C. Gilissen<sup>13,14</sup>; X. Latypova<sup>11</sup>; S.G. Kant<sup>15</sup>; B. Keren<sup>16</sup>; T. Kleefstra<sup>13,17</sup>; R.R. Lebel<sup>18</sup>; K. Machol<sup>8</sup>; C. Mignot<sup>16</sup>; V. Narayanan<sup>19</sup>; M. Nizon<sup>11</sup>; R. Pfundt<sup>13</sup>; F. Scaglia<sup>8,20,21</sup>; C. Smith-Hicks<sup>22</sup>; J. Taylor<sup>23</sup>; A.S. Trentesaux<sup>24</sup>; N. Zadeh<sup>25,26</sup>; D. Babovic-Vuksanovic<sup>2,27</sup>; R.A. Urrutia<sup>5,28</sup>; A. Mathie<sup>3</sup>; E.W. Klee<sup>1,2,27</sup>; *KCNK9* Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, MN, USA; 3) Medway School of Pharmacy, University of Kent and University of Greenwich, Central Avenue, Chatham Maritime, Kent, ME4 4TB, United Kingdom; 4) Clinical and Translational Sciences Institute, Medical College of Wisconsin, Milwaukee, WI, USA; 5) Bioinformatics Research and Development Laboratory, Genomic Sciences and Precision Medicine Center, Medical College of Wisconsin, Milwaukee, WI, USA; 6) Children's Hospital of Minnesota, Minneapolis, MN, USA; 7) Division of Medical Genetics, Northwell Health, Manhasset, NY, USA; 8) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 9) Division of Neurology, BC Children's Hospital, Vancouver, British Columbia, Canada; 10) Clinical Genetics, Northampton General Hospital NHS Trust, Cliftonville, Northampton NN1 5BD; 11) CHU Nantes, Service de génétique médicale, Nantes, France; 12) APHP, Department of Neuropediatrics, National Reference Center for Neurogenetic Disorders, Hôpital Armand-Trousseau, GHUEP, Paris, France; 13) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 14) Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; 15) Department of Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; 16) APHP, Département de Génétique et Centre de Référence Déficiences Intellectuelles de Causes Rares, Hôpital de la Pitié-Salpêtrière, Assistance Publique - Hôpitaux de Paris, Paris 75651, France; 17) Donders Centre for Neuroscience, Radboud University Medical Center, Nijmegen, The Netherlands; 18) Section of Medical Genetics, SUNY Upstate University Hospital, Syracuse, NY, USA; 19) Center for Rare Childhood Disorders, Translational Genomics Research Institute, Phoenix, AZ, USA; 20) Texas Children's Hospital, Houston, TX, USA; 21) Joint BCM-CUHK Center of Medical Genetics, Prince of Wales Hospital, Shatin, Hong Kong SAR; 22) Department of Neurogenetics, Kennedy Krieger Institute, Baltimore, MD, USA; 23) Oxford Medical Genetics Laboratories, Oxford University Hospitals NHS Foundation Trust, The Churchill Hospital, Oxford, UK, OX3 7LE; 24) Service de Néonatalogie, CHU de Caen, France; 25) Genetics Center, Orange, CA, USA; 26) Division of Medical Genetics, CHOC Children's Hospital, Orange, CA, USA; 27) Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA; 28) Department of Surgery, Medical College of Wisconsin, Milwaukee, WI, USA

---

The paternally imprinted *KCNK9* gene encodes a two-pore-domain potassium ion channel expressed primarily in the brain. A Gly236Arg variant in *KCNK9* has been reported to cause *KCNK9* imprinting syndrome in several families, as has a nearby Ala237Asp variant in an additional individual. Here, we

describe a cohort of 18 affected individuals from 14 families with one previously reported (Gly236Arg) and 11 novel *KCNK9* alterations. Phenotypes in these patients are generally consistent with, but demonstrating greater variability, compared to published patients with *KCNK9* imprinting syndrome. The most common phenotypes include motor delay, speech delay, intellectual disability, facial dysmorphism, hypotonia, and behavioral concerns (aggressive or compulsive behavior, ADHD, anxiety).

Protein modeling and dynamics simulations predict variants affect protein structure, dynamics, and K<sup>+</sup> ion distribution. Compared to WT, a subset of variants demonstrate an increased probability of K<sup>+</sup> near the channel pore and are associated with more K<sup>+</sup> transport events. This is in stark contrast to the decrease seen with previously described Gly236Arg, suggesting distinct channel dysregulation by these novel variants. Patient variant characterization using electrophysiological techniques showed seven of thirteen variants cause a gain of conductance compared to WT. Three variants result in a loss or reduction in conductance, and three variants show conductance similar to WT. Ten of the thirteen variants were outwardly rectifying with reversal potentials close to the WT equilibrium potential. Three variants had significantly depolarised reversal potentials. Computational modeling and molecular dynamic simulation results were consistent with experimental assays. Further, we have identified two additional families with variants impacting channel function, but for which additional evidence suggests reduced penetrance.

We describe a cohort of patients with *KCNK9* imprinting syndrome caused by novel *KCNK9* variants with varied functional impact. Both loss of function, like the classical Gly236Arg, and gain of function effects were observed. Based on this observation, we predict that channel agonists, which have been reported as efficacious in patients with the Gly236Arg variant, may not benefit patients with gain-of-function variants.

## PgmNr 1296: Multi-omic analyses in a family with Gaucher and Gaucher-Parkinson diseases reveal several potential modifier genes in *GBA1*-associated Parkinson disease.

### Authors:

E. Garcia <sup>1</sup>; F. Donovan <sup>2</sup>; B. Stubblefield <sup>1</sup>; P. Sharma <sup>1</sup>; S. Arepalli <sup>3</sup>; R. Bhatnagar <sup>4</sup>; S. Chandrasekharappa <sup>2</sup>; D. Hernandez <sup>3</sup>; G. Lopez <sup>1</sup>; E. Sidransky <sup>1</sup>; N. Tayebi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Neurogenetics Section, Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD.; 2) Cancer Genetics and Comparative Genomics Branch/Genomics Core, National Human Genome Research Institute, Bethesda, MD; 3) Genetic Tech Group, Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD; 4) Verge Genomics, Genomic, San Francisco, CA

---

Gaucher disease (GD) is an autosomal recessive lysosomal storage disorder caused by mutations in *GBA1*. Although GD is a rare disorder, mutations in *GBA1* are also the most common genetic risk factor for the more prevalent neurological disorder Parkinson disease (PD). Focusing on a nuclear family with several informative members, we performed multi-omic analyses including epigenetic studies to identify potential modifiers impacting the association between GD and parkinsonism. The proband in this family had type 1 GD and was evaluated at age 20 when she presented with typical Gaucher manifestations, and again at age 56 after she developed PD, and fibroblasts were collected on both occasions. We analyzed DNA methylation in both sets of fibroblasts, as well as fibroblasts from her two older sisters, one with GD but without PD, and the other without PD or *GBA1* mutations. Bisulfite-treated DNA extracted from the fibroblast lines and an Illumina HumanMethylationEpic BeadChip were used to analyze the methylation of CpG islands. RNA, extracted from fibroblasts from the proband at both ages and from her two sisters, was analyzed by microarray using an Affymetrix Human Clariom S Assay and by 3'-Poly(A) tail RNA-seq. Several methylated and demethylated candidate genes, identified in the proband at age 56, were confirmed in the expression array and by RNA-seq. Illumina GenomeStudio, Affymetrix Transcriptome Analysis Console, Limma and R were employed to elucidate biologically significant genes for further evaluation. Whole Exome Sequencing was also performed on DNA from each family member, and homozygous and compound heterozygous variants identified in the proband only were confirmed by Sanger sequencing and are currently undergoing further evaluation. Integrating the data from methylation studies, whole genome expression arrays, RNA-seq and whole exome sequencing using functional and protein-network analyses may help to determine whether the observed changes are directly or indirectly related to the development of parkinsonism in this patient with GD.

## PgmNr 1297: Novel variants in *LYST* associated with Chediak-Higashi disease.

### Authors:

E.R. Nicoli <sup>1</sup>; C. Kuptanon <sup>2</sup>; M. Morimoto <sup>1</sup>; D. Yarnell <sup>1</sup>; J. Serra-Vinardell <sup>2</sup>; P. Sharma <sup>1</sup>; D.L. Ross <sup>1</sup>; H.M. Dorward <sup>2</sup>; J. Burke <sup>2</sup>; C. Ciccone <sup>2</sup>; R. Fischer <sup>2</sup>; M. Huizing <sup>2</sup>; C. Toro <sup>1</sup>; K. Krzweski <sup>3</sup>; W.A. Gahl <sup>1,2</sup>; W.J. Introne <sup>2</sup>; MCV. Malicdan <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Undiagnosed Diseases Program, National Human Genome Research Institute (NHGRI), National Institutes of Health (NIH), Bethesda, MD 20892, USA; Common Fund, Office of the Director, NIH, Bethesda, MD 20892, USA.; 2) Human Biochemical Genetics Section, National Human Genome Research Institute, NIH, Bethesda, MD 20892, USA.; 3) Receptor Cell Biology Section, Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases (NIAD), National Institute of Health (NIH), Rockville, MD 20852.

---

**Introduction:** Chediak-Higashi Disease (CHD) is a rare, recessive disorder characterized by hematologic, pigmentary, neurologic and immunologic features, and can involve life-threatening hemophagocytic lymphohistiocytosis. CHD results from bi-allelic mutations in the lysosomal trafficking regulator gene (*LYST*). *LYST* contains 55 exons, encoding a large 429 kDa protein with an unknown function. *LYST* has an N-terminal ARM/HEAT domain and a C-terminal BEACH domain with WD-40 repeats. More than 75 *LYST* variants causative of CHD have been reported to date, scattered throughout the gene.

**Methods:** In our CHD Natural History study, we extensively investigated 17 CHD patients. Molecular and cellular studies were performed on fibroblast, leukocyte, natural killer (NK) cell and melanocyte cultures. Clinical, genetic and cell biology studies were used to evaluate the correlation between phenotypes and the molecular genotypes of patients with CHD.

**Results:** The patients in our cohort had variable clinical features of CHD, bi-allelic *LYST* variants and enlarged lysosomes within cell cultures. We identified 28 novel *LYST* variants. Genotype-phenotype correlations were noted. Most cases with classic severe CHD clinical features had two truncating *LYST* variants and markedly enlarged lysosomes in cultured fibroblast that were perinuclear clustered, while individuals with a milder clinical presentation had at least one homozygous missense *LYST* variant and milder enlargement and perinuclear accumulation of lysosomes. NK cells of subjects with ARM/HEAT domain mutations had larger lysosomes that polarized to the immune synapse but failed to fuse with the plasma membrane. NK cells of subjects with BEACH domain mutations had smaller and more numerous lysosomes, with impaired migration of lysosomes to the immunologic synapse. In all subjects, NK lytic activity was markedly decreased.

**Conclusion:** Our molecular studies increased the total reported number of *LYST* variants to over 100. CHD clinical severity correlates with severity of the *LYST* protein variant and cellular phenotype. Cellular studies suggest a role of *LYST* in localization and size of lysosomes as well as lytic activity of NK cells. CRISPR cell lines and mouse models are now being investigated to aid our basic investigations and design therapeutic approaches. These results are valuable for understanding the spectrum of disease in CHD and the underlying biological mechanisms of disease.

## PgmNr 1298: Correlation between HPO terms and diagnostic yield in neurodevelopmental disorders.

### Authors:

M. Lazar<sup>1</sup>; K.M. Rocha<sup>1</sup>; M.C. Varela<sup>1</sup>; J.R.M. Ceroni<sup>1,2</sup>; G.L. Yamamoto<sup>1,2</sup>; V.N. Takahashi<sup>1</sup>; K. Bagatini<sup>1</sup>; D.G. Franco<sup>1</sup>; M.O. Scliar<sup>1</sup>; J.Y.T. Wang<sup>1</sup>; M. Zatz<sup>1</sup>; M.R. Passos-Bueno<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome and Stem Cell Research Center (HUG-CELL), University of São Paulo, Brazil - Rua do Matão, travessa 13, 106 - 05508-090. São Paulo, SP, Brazil; 2) Department of Clinical Genetics, Children's Hospital, Medical School, University of São Paulo, Brazil - Av. Dr. Enéas Carvalho de Aguiar, 647 - 05403-000. São Paulo, SP, Brazil

---

HUG-CELL has implemented targeted massively parallel sequencing for diagnosis of several diseases and currently offers tests and genetic counseling for families affected with different conditions. We are conducting a preliminary analysis to verify the correlation between specific terms described in medical request and our diagnostic yield. To accomplish this goal, we have included 18 patients exhibiting features of neurodevelopmental disorders, who were referred to our center for molecular testing throughout the last 2 years with clinical indications presenting the terms: global developmental delay (**GDD**) (HP:0001263), developmental regression (**DR**) (HP:0002376) and seizures (**Se**) (HP:0001250). Using a panel with ~6700 genes (TruSight One Expanded - Illumina), we have concluded the molecular diagnosis for 9 individuals (50%) (pathogenic/likely pathogenic alterations in *ACTB*, *AHDC1*, *CTNNB1*, *DEAF1*, *DHDDS*, *DYRK1A*, *HDAC8*, *KCNB1*, *RAC1*, some of them not previously reported in the literature); 6 patients (33%) presented at least one variant of unknown significance - *VUS* (*ARID1B*, *CACNA1H*, *KCNQ3*, *SCN2A*, *SMARCC2*) or one pathogenic/likely pathogenic alteration in genes previously associated with autosomal recessive mode of inheritance (*COG1*, *COG5*); in 3 individuals (17%) no alteration was detected. In spite of the relevant sensitivity rate obtained by our laboratory, the relative frequencies taking into account the abovementioned terms are highly discrepant: considering patients who presented **GDD** (HP:0001263) associated or not with **Se** (HP:0001250) (N=8), the positive detection was attained for 75% (N=6); for individuals whose reasons for request included **DR** (HP:0002376) associated or not with **Se** (HP:0001250) (N=4), the conclusive diagnosis was achieved for 50% (N=2); nevertheless, for patients with clinical indications containing the term **Se** (HP:0001250) without stated **GDD** (HP:0001263) or **DR** (HP:0002376) (N=6) the molecular diagnosis was concluded in 17% (N=1). Although the limited size of our cohort, the low rate of positive diagnosis in the third group (**Se** without **GDD** or **DR**) could suggest that several associated-genes remain to be identified and/or its etiology is complex and fits a polygenic model instead of a Mendelian trait. Once confirmed in a large cohort, these results will be significant to prioritize patients to be investigated using whole exome sequencing, particularly when their tests are paid by public funding.

Financial support: Fapesp/CEPID and CNPq/INCT.

## PgmNr 1299: Genetics of a novel neonatal neuromuscular disorder in dogs.

### Authors:

K.J. Dillard <sup>1,2,3,4</sup>; M.K. Hytönen <sup>1,2,3</sup>; M. Arumilli <sup>1,2,3</sup>; M. Anttila <sup>4</sup>; H. Lohi <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland; 2) Department of Medical Genetics, University of Helsinki, Biomedicum, Helsinki, Finland; 3) Folkhälsan Research Institute, Helsinki, Finland; 4) Veterinary Bacteriology and Pathology Research Unit, Finnish Food Authority, Helsinki, Finland

---

Neuromuscular disorders are a group of rare heterogeneous diseases in humans and animals including dogs. In our research on the causes of high neonatal mortality in Norwich Terrier breed, we have discovered a neonatal neuromuscular disorder in two consecutive litters by the same unaffected parents. In the first litter, a five-week-old puppy was euthanized due to progressive weakness and consequent difficulties in ambulation. In the second litter, a puppy was euthanized at the age of 1.5 weeks due to dyspnea and failure to thrive. A full necropsy was performed on both puppies. Histologically both affected puppies had mild to moderate neurogenic atrophy of the striated muscles. In addition, the 1.5 weeks old puppy had severe thoracic dysplasia. Whole-genome sequencing (WGS) was performed of both affected puppies and their unaffected parents. An in-house pipeline was used for the calling of SNVs, indels, SVs, CNVs and mobile elements. The variants were filtered under recessive, de novo and compound heterozygous models against 868 whole-exome and WGS data from unaffected dogs from other breeds including seven Norwich Terriers. We also analyzed the data as two independent trios that resulted in discovery of candidate variants not shared by the affected puppies suggesting the muscle phenotype is genetically different in them. The found variants were first prioritized by in silico prediction and deleterious candidate variants are currently being validated and screened by Sanger sequencing, using DNA samples from a large cohort of Norwich Terriers stored in our (over 70.000 samples) canine biobank at the University of Helsinki. The functional consequences will be studied by utilizing tissue samples that were collected during necropsy. As dogs have emerged as an important natural model for human inherited diseases with clinicopathological and genetic similarities, our study will bring new insights to the neuromuscular disease spectrum in humans.

# PgmNr 1300: Mutations in *COA7* cause spinocerebellar ataxia with axonal neuropathy.

## Authors:

Y. Higuchi <sup>1</sup>; A. Hashiguchi <sup>1</sup>; J. Yuan <sup>1</sup>; A. Yoshimura <sup>1</sup>; Y. Okamoto <sup>1</sup>; E. Matsuura <sup>1</sup>; J. Mitsui <sup>2</sup>; H. Ishiura <sup>2</sup>; M. Nakagawa <sup>3</sup>; S. Morishita <sup>4</sup>; T. Toda <sup>2</sup>; M. Yamaguchi <sup>5</sup>; S. Tsuji <sup>2</sup>; H. Takashima <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Neurology and Geriatrics, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima city, Kagoshima, Japan; 2) Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Director of North Medical Center, Kyoto Prefectural University of Medicine, Kyoto, Japan; 4) Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; 5) Department of Applied Biology and The Center for Advanced Insect Research, Kyoto Institute of Technology, Japan

## OBJECTIVE

Inherited peripheral neuropathies (IPNs), as represented by Charcot-Marie-Tooth disease (CMT), are clinically and genetically heterogeneous. To date, more than 100 causative genes have been identified. Unexpectedly, the mutation detection rate of IPNs remains low, suggesting the necessity of further exploring unidentified disease causing genes. The purpose of this study is to identify new causes of IPNs in patients with autosomal recessive (AR) IPNs.

## METHODS

Using whole-exome sequencing (WES), we sequenced 303 unrelated patients clinically diagnosed with IPNs. We concentrated on all the recessive variants/genes shared among multiple patients manifesting with comparable clinical phenotypes. We identified cytochrome oxidase assembly factor 7 (*COA7*) gene as a novel causative gene for AR-IPNs and performed mutation screening of the *COA7* gene in 1024 additional patients with IPNs. Furthermore, we provide clinical, genetic, histopathological, and functional data to support the contention that the recessive mutations in the *COA7* gene are responsible for neurological impairment.

## RESULTS

We identified recessive mutations in the *COA7* gene in six unrelated patients. Validated mutations were located at highly conserved residues among different species and segregated with the disease in each family. All six patients with *COA7* mutations had characteristic neurological features of axonal neuropathy, and some patients showed cerebellar ataxia, extrapyramidal symptoms, cognitive impairment, spasticity or subclinical mitochondrial myopathy. MRI scans showed white matter lesions, cerebellar or spinal cord atrophy. Sural nerve biopsies showed chronic axonal degeneration with a marked loss of large and medium myelinated fibres. Mitochondrial respiratory chain enzyme assay in skin fibroblasts from the three patients showed a definitive decrease in complex I or complex IV. In addition, *Drosophila COA7* (*dCOA7*) knockdown models showed rough eye phenotype, reduced life span, impaired locomotive ability and shortened synaptic branches of motor neurons.

## CONCLUSIONS

Our results suggest that loss-of-function *COA7* mutation is responsible for the phenotype of the



presented patients, and this new entity of disease would be referred to as spinocerebellar ataxia with axonal neuropathy type 3 (SCAN3)[MIM 618387]. Mutations in *COA7* may also lead to heterogeneous neuromuscular involvement, manifesting phenotypic diversity in the future.

## **PgmNr 1301: Exome utility in phenotype/genotype association for 10K individuals: Progress toward understanding the functional biology of the ~20K genes in the human genome.**

### **Authors:**

S.N. Jhangiani <sup>1</sup>; J. Posey <sup>2</sup>; Z.C. Akdemir <sup>2</sup>; C. Boehm <sup>4</sup>; C. Carvalho <sup>2</sup>; D. Pehlivan <sup>3</sup>; N. Sobreira <sup>4</sup>; K. Doheny <sup>4</sup>; A. Hamosh <sup>4</sup>; F. Sedlazeck <sup>1</sup>; J. Hu <sup>1</sup>; H. Doddapaneni <sup>1</sup>; V.R. Sutton <sup>2</sup>; D.M. Muzny <sup>1</sup>; E. Boerwinkle <sup>1</sup>; D. Valle <sup>4</sup>; J.R. Lupski <sup>2</sup>; R.A. Gibbs <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Human Genome Sequencing Center, Baylor College Medicine, Houston, Texas; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 3) Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, Texas; 4) McKusick-Nathans Department of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland

---

Since inception in 2011, the Centers for Mendelian Genomics have recruited 23K families and generated data for more than 45K individuals from 80 countries. This effort has been a collaboration among four funded centers and >4K collaborators worldwide. The Baylor-Hopkins Center for Mendelian Genomics (BHCMG) has enrolled more than 18K individuals and contributed WES data for 10K samples. For the BHCMG, the BCM:Human Genome Sequencing Center and JHU:Center for Inherited Disease Research have generated ~90Tb of WES data. Using a multiplex strategy with an in-house developed exome capture with custom spike-in has yielded 10Gb of data per exome with 97% at ≥20X coverage. Implementation the NovaSeq platform has decreased the cost by 50% per sample.

Molecular diagnostic rates vary by phenotype/cohort from 37-85%. For undiagnosed cases several strategies are used to ascertain a molecular diagnosis including: sequencing additional family members, searching for additional patients with the same candidate genes/variants in publically accessible databases including GeneMatcher (32K submissions, 130K matches and 10K genes), VariantMatcher and internal databases; and/or alternative sequencing approaches. A large-scale re-processing effort for the BHCMG dataset is underway including alignment and variant calling on GRCh38 for a more comprehensive analysis. Also several families are being piloted for a combined approach using both short/long read WGS to identify coding/non-coding variation, copy number and copy neutral structural variants. BHCMG continues to engage private diagnostic labs to reflex undiagnosed clinical cases into research studies utilizing existing sequencing data for further analysis.

Cumulative analysis of the BHCMG dataset has yielded discoveries in 741 genes including 267 novel and 176 phenotypic expansion genes. Reported discoveries including allelic series, multi-locus pathogenic variation and other mutational burden/oligogenic models for complex traits have been published in >250 papers. BHCMG has established valuable sample acquisition approaches; generated ethnicity-specific data resources, enhanced sequencing methodology, developed PhenoDB, a highly curated, searchable database currently with 8.7K entries and encouraged collaborative efforts. Each discovery has highlighted the diagnostic capabilities in using WES and has taught

lessons in the genomics of disease that continue to drive investigation into cases that remain undiagnosed.

## PgmNr 1302: Mutation in *CADM3* cause upper limb predominant neuropathy with pyramidal signs.

### Authors:

A.P. Rebelo <sup>1</sup>; A. Cortese <sup>2</sup>; E. Peles <sup>3</sup>; L. Abreu <sup>1</sup>; C. Bacon <sup>4</sup>; S.M. Feely <sup>4</sup>; D. Castro <sup>5</sup>; M.M. Reilly <sup>2</sup>; M. Shy <sup>4</sup>; S. Zuchner <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dr. John T. Macdonald Foundation Department of Human Genetics, John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, USA; 2) MRC Centre for Neuromuscular Diseases, National Hospital for Neurology and Neurosurgery, UCL Institute of Neurology, Queen Square, London, UK; 3) Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel; 4) Department of Neurology, Carver College of Medicine, University of Iowa, Iowa City, USA; 5) Departments of Pediatrics, Neurology and Neurotherapeutics, University of Texas Southwestern Medical Center, Dallas, USA

---

The *CADM* family of proteins consists of four neuronal specific adhesion molecules (*CADM1*, *CADM2*, *CADM3* and *CADM4*) that mediate the direct contact and interaction between axons and glia. In the peripheral nerve, axon-Schwann cell interaction is essential for the structural organization of myelinated fibers and it is primarily facilitated via *CADM3*, expressed in axons, binding to *CADM4*, expressed by myelinating Schwann cells. We have identified by whole exome sequencing (WES) three families with axonal Charcot-Marie-tooth disease (CMT2) sharing the same private variant in *CADM3*, Tyr172Cys. This variant is absent in 230,000 alleles in gnomAD and predicted to be pathogenic. Although these *CADM3* families have been diagnosed with CMT2, they all share the same peculiar phenotype consisting of axonal motor neuropathy affecting mainly the upper limbs with pyramidal signs. High-resolution mass spectrometry analysis detected a new disulfide bond created in the mutant *CADM3* potentially modifying the native protein conformation. This result was supported by significant increased protein retention of the mutant in the endoplasmic reticulum leading to activation of the unfolded protein response (UPR). In addition, Stochastic Optical Reconstruction Microscopy (STORM) imaging revealed decreased co-localization of the mutant *CADM3* with *CADM4* at the plasma membrane between cell-cell contact sites. Our findings indicate a novel molecular pathway in patients with CMT involving axon-glial interaction abnormalities.

## **PgmNr 1303: Cofactors biotin, pantothenic acid, and lipoic acid dramatically improved the health of a child with *SLC5A6* gene defect, but not natural killer T-cell function, leading to an ALTE with brain damage and loss of gained skills.**

### **Authors:**

P.J. Benke <sup>1</sup>; S.A. Parikh <sup>2</sup>; B. Braffman <sup>3</sup>; C. Chang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Genetic Division, Joe Dimaggio Children's Hosp, Hollywood, Florida.; 2) Joe DiMaggio Children's Hosp, Hollywood Florida; 3) Memorial Health Care System, Hollywood Florida

---

Mutations in the gene *SLC5A6* are a newly described treatable genetic disorder in the SMVT protein, leading to decreased transport of biotin, pantothenic acid, and lipoic acid across cell membranes (Subramanian, VS, et al; Mutations in *SLC5A6* associated with brain, immune, bone and intestinal dysfunction in a young child. Human Genetics 2017 Feb;136 (2):253-261). Our patient's treatment began at age 15 months with high dose cofactors biotin, pantothenic acid, and lipoic acid to overcome the joint transport defect by diffusion. There was a dramatic improvement in areas of growth and development, head size, bone density and health (ibid). Here we report that frequent infections in this child did not improve and natural killer (NK) cell numbers and functions remained low. A year and a half after start of therapy, the child experienced a viral infection that led to an acute life threatening event (ALTE) with fever, shock, hypotension and devastating central nervous system effects. A virus was not identified. Dramatic hypoperfusion effects were found on an MRI scan in spin of inotropic measures to increase blood pressure and maintain life. The patient recovered after 2 weeks of intensive care, but he suffered loss of skills and intellectual performance. He could no longer walk or talk. Immunologic factors, including immunoglobulins which did improve after therapy, and NK cells which did not, were part of the expression of *SLC5A6* defects in this case. The defect in NK cells was a risk factor for the viral infection experienced by this child, and earlier diagnosis may have altered the outcome.

## PgmNr 1304: Novel *SON de novo* mutation in a patient with a severe neurodevelopmental disorder.

### Authors:

K. Hosoki <sup>1</sup>; S. Miyatake <sup>2</sup>; N. Miyake <sup>2</sup>; N. Matsumoto <sup>2</sup>; N. Okamoto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Molecular Medicine, Osaka Women's and Children's Hospital Research Institute; 2) Department of Human Genetics, Yokohama City University Graduate School of Medicine

---

*SON* is one of serine and arginine-rich (SR) proteins which contain a long repeat of serine and arginine as a characteristic domain, known as the RS domain, and play as splicing regulators. A heterozygous mutation of *SON* in patients with ZTTK syndrome can lead to splicing errors, such as exon skipping and intron retention. There are 29 reported mutations, and most of the mutations (28/29) are detected in exon 3 of *SON*, which is the largest exon and contains RS domain. These mutations are located in the RS domain or the upstream of RS domain; therefore, these mutations may lead to the dysfunction of RS domain. We detected a novel *de novo* heterozygous mutation in *SON* at the downstream of RS domain by whole exome sequencing, and our case shows the clinical features of ZTTK syndrome. We also detected new two *de novo* heterozygous mutations in *SON* exon 3 at the upstream of RS domain, and these are suggested the dysfunction of the RS domain as well as reported cases with ZTTK syndrome (case 2: c.2977\_2980dup and case 3: c.1881\_1882delAG, NM\_138927.2). Here, we report a novel *SON* mutation and propose that a *SON* mutation in the downstream of RS domain is associated with clinical features of ZTTK syndrome by genotype-phenotype analysis.

Our novel case showed a severe intellectual disability, motor developmental delay, hypotonia, brain malformation, autistic features, and dysmorphic features. The other two cases with a hotspot mutation in *SON* exon 3 showed the characteristic features of ZTTK syndrome patients. Compared to reported cases with *SON* hotspot mutation, our three cases mimic the phenotype of ZTTK syndrome. We also performed expression analysis using patient RNA samples, which is isolated from PBMC of our three cases. All samples showed the changes of alternative splicing patterns in several genes, such as *ADA* and *TUBG1*, which are reported the association with *SON* mutation and splicing errors. The previous report showed only the splicing errors of the mutations in the RS domain. Therefore, this is the first evidence that both mutation types of the RS-upstream and RS domain induce *SON*-mediated splicing errors; moreover, the novel mutation in the RS-downstream of *SON* also lead to splicing errors. Our data will help to understand the gene function of *SON* and the molecular mechanisms of RNA-mediated neurodevelopmental disorders.

## PgmNr 1305: Clinical features of KIF1A-related disorder: A Japanese patient with a novel missense variant and literature review.

### Authors:

K. Nakamura <sup>1,2</sup>; E. Kise <sup>1</sup>; T. Yamaguchi <sup>1,3</sup>; T. Yoshinaga <sup>2</sup>; M. Kodaira <sup>2</sup>; Y. Sekijima <sup>2</sup>; T. Kosho <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Medical Genetics, Shinshu University Hospital, Matsumoto, Nagano, Japan; 2) Department of Medicine (Neurology and Rheumatology), Shinshu University School of Medicine, Matsumoto, Nagano, Japan; 3) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Nagano, Japan; 4) Research Center for Supports to Advanced Science, Shinshu University, Matsumoto, Nagano, Japan

---

[Background] Kinesin superfamily proteins (KIFs) play an essential role in neuronal development and survival by transporting membranous organelles and protein complexes along microtubules. Kinesin family member protein 1A (KIF1A) belongs to the kinesin-3 and is a microtubule plus end-directed motor protein. Here, we report clinical and genetic findings of a patient with multiple system neurodegeneration, who was found to have a novel heterozygous missense variant in *KIF1A* (c.833T>C, p.L278P). [Method and Results] The patient is a 33-year-old Japanese man. His motor and intellectual development were delayed. At 36 months, his full-scale IQ assessed by Tanaka-Binet test was 34. At age 7, his neurological examination showed distal dominant muscle atrophy and weakness. Deep tendon reflexes were brisk in the upper and lower extremities, except Achilles tendon reflexes, which were absent. Both planter responses were extensor, and he had pes cavus. Both limb and truncal ataxia were evident. He could not walk without walking assist tools. Routine blood and cerebrospinal fluid examinations were normal. Magnetic resonance imaging (MRI) of the brain showed atrophy in the cerebellar hemispheres and vermis. An ophthalmological examination demonstrated bilateral optic neuropathy and cataract. Serial radiological studies revealed progressive cerebral cortical atrophy. Clinical exome analysis, using a MiSeq and TruSight One Sequencing Panel (Illumina, San Diego, CA, USA) including 4813 genes associated with human genetic disorders, revealed a novel de novo heterozygous missense variant in *KIF1A*. [Discussions] Our patient showed multiple system neurodegeneration, including infantile or childhood onset, intellectual disability, upper and lower motor neuron signs, and cerebellar ataxia. To date, 70 patients with genetically-confirmed KIF1A-related disorders, including ours, have been reported. Most of patients showed various degrees of neurological manifestations, including upper (89%) and lower (45%) motor neuron signs, intellectual disability (66%), cerebellar ataxia (45%), and optic neuropathy (39%), while epilepsy (13%) and microcephaly (10%) were less frequent. This study suggests the clinical utility of using panel-based TES. KIF1A-related disorder is very rare and poorly understood. Molecular genetic analysis of *KIF1A* should be considered in patients who show a combined phenotype of both upper and lower neuron sign, intellectual disability and cerebellar ataxia.

## PgmNr 1306: Disease severity in *KIF1A* Associated Neurological Disorders (KAND) is correlated with variant location.

### Authors:

L. Boyle <sup>1,2</sup>; X. Fan <sup>1</sup>; L. Hamm <sup>1</sup>; A. Thornton <sup>2</sup>; Y. Shen <sup>1</sup>; W.K. Chung <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Molecular Genetics, Columbia University Irving Medical Center, New York, NY; 2) Vagelos College of Physicians and Surgeons, New York, NY

---

### Background

*KIF1A* Associated Neurological Disorders (KAND) comprise a recently identified group of rare neurodegenerative conditions caused by mutations in *KIF1A*. *KIF1A* is a member of the kinesin-3 family of microtubule (MT) motor proteins. The phenotypic spectrum of KAND includes neurodevelopmental delay, intellectual disability, autism, microcephaly, progressive spastic paraplegia, peripheral neuropathy, optic nerve atrophy, cerebral and cerebellar atrophy, and seizures. Diagnosis is complicated by the large number of disease-causing mutations identified, predominantly heterozygous missense mutations occurring within the motor domain of the protein, many of which are unique mutations. Even among dominant variants within the motor domain, there is currently no way to predict prognosis.

### Methods

We collected caregiver or self-reported medical history and Vineland Adaptive Behavior Scales from individuals with clinically-identified *KIF1A* variants. We also aggregated data on all 82 KAND cases described in the literature, and developed a heuristic severity score by a weighted sum of common symptoms which we applied to all cases with sufficient data. To compare variant severity across dominant motor domain mutations, we further characterized genetic variants using deleterious scores, disease severity, (averaged across a given mutation) and location in the 3 dimensional protein structure.

### Results

Our cohort includes 71 individuals, 58% male and 42% female, 2m-37y old (mean: 10 years, median: 7 years). We describe 32 mutations not previously published. Our analysis of motor domain mutations focused on 133 cases (54 from our cohort, 79 from the literature). We saw a modest but significant correlation between disease severity and CADD scores ( $p=0.27$ ,  $p\text{-value}=0.002$ ), and found that disease severity is strongly associated with location in MT-binding regions, P-loop, switch I, switch II, and neck linker. We found 23 out of 28 (82%) of the most severe KAND cases (score  $>5$ ) had mutations located in the MT-binding regions (vs 10/30 of the least severe), and cases with mutations in these regions had significantly higher severity scores ( $p=2.1e-5$ ;  $\Delta\text{mean}=3$ ) than those with mutations outside the MT-binding regions.

### Discussion

We describe the largest natural history study of KAND to date and the first KAND severity score. We identified the functional subdomains of the protein associated with the most severe disease, providing useful targets for future therapeutic development.



## PgmNr 1307: ***BICRA***, a SWI/SNF complex member, is associated with Coffin-Siris-like phenotypes in humans and flies.

### Authors:

S.D. Barish<sup>1,2</sup>; B. Ugur<sup>1,2,3,4</sup>; A. Valencia<sup>5</sup>; N. Mashtalir<sup>5</sup>; C. Kadoch<sup>5</sup>; U. Undiagnosed Diseases Network<sup>1</sup>; D. Scott<sup>1</sup>; M. Wangler<sup>1,2,4</sup>; S. Yamamoto<sup>1,2,4,6</sup>; H. Bellen<sup>1,2,4,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX 77030, USA; 3) Department of Neuroscience, Yale University, New Haven, CT 06520, USA; 4) Program in Developmental Biology, Baylor College of Medicine, Houston, TX 77030, USA; 5) Pediatrics, Harvard Medical School, Cambridge, MA 02115, USA; 6) Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA; 7) Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX 77030, USA

---

Coffin-Siris syndrome (CSS) is a rare but severe neurodevelopmental disorder that is characterized by developmental disability, hypoplasia of the fifth digit, coarse facial features, and a variety of other rarer symptoms. Mutations in the members of the SWI/SNF chromatin remodeling complex are thought to cause CSS, although there is little to no functional data characterizing the patient mutations. Through the Undiagnosed Disease Network (UDN) we identified a patient with symptoms similar to CSS, who carries a *de novo* frameshift allele in the gene BRD4 Interacting Chromatin Remodeling Complex Associated protein (*BICRA*), a non-canonical member of the SWI/SNF complex. Here we present the first functional characterization of the *Drosophila* homolog of *BICRA*, *CG11873*, and correspondingly the first functional characterization the non-canonical SWI/SNF complex *in vivo*. We demonstrate that, like other SWI/SNF complex members, loss of *CG11873* is a dominant enhancer of position effect variegation. *CG11873* mutants also exhibit climbing defects at day 1 and live only for one week, both of which can be rescued by the genomic locus. Finally, we show that *CG11873* binds to other SWI/SNF complex members. Together our data show that *CG11873* is a SWI/SNF complex member whose loss leads to a less severe version of CSS in humans.

## PgmNr 1308: Upstream missense variants approximately located in *Hikeshi* gene cause hypomyelinating leukodystrophy.

### Authors:

S. Ben-Shachar <sup>1</sup>; D. Brabbing <sup>2</sup>; N. Imamoto <sup>3</sup>; A. Fattal-Valevski <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetic Institute, Tel-Aviv Medical Center, affiliated to Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel , current location, Clalit Research institute, Tel-Aviv, Israel; 2) Genetic Institute, Tel-Aviv Medical Center, affiliated to Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel; 3) Cellular Dynamics Laboratory, RIKEN, Wako, Saitama, Japan; 4) Pediatric Neurology Institute, Dana-Dwek Children's Hospital, Tel-Aviv Medical Center, affiliated to Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel

---

**Background:** A founder Ashkenazi missense variant, Chr11:86017416 G>C (p.Val54Leu), in *Hikeshi* gene (*C11ORF73*), encoding HSP70 nuclear transporter protein was found to cause hypomyelinating leukodystrophy, associated with spasticity and intellectual disability. At present, no further pathogenic variants in the gene have been described. We detected two additional missense variants in the gene causing similar phenotype.

**Methods:** We have performed trio exome sequencing of patients with leukodystrophy and their parents as part of a clinical evaluation.

**Results:** We have detected a missense variant, CHR11: 86017488 T>C (p.Pro78Ser) in homozygous state in an 18 year-old male of Christian Arab origin, with leukodystrophy, severe intellectual impairment and spasticity.. His healthy consanguineous parents were found to be heterozygous to the variant, while his healthy two sibs were detected to be heterozygous for the variant as well. A compound heterozygous for the founder Ashkenazi variant and the CHR11:86017330 T>C (p.Phe25Ser) variant has been detected in a Russian boy of partial Ashkenazi Jewish origin), presented with similar albeit milder phenotype. The two novel variants detected, were detected only very rarely among normal population (Allele frequency of 4.064e-6 for both , gnomAD browser). Brain MRI showed a similar characterization of the leukodystrophy in all cases.

**Conclusion:** We have detected two novel missense variants in *Hikeshi* gene among patients with a characterized leukodystrophy, adding to the one pathogenic variant, previously described. These approximately located variants, suggest that this region may have a functional importance. We suggest to add *Hikeshi* gene sequencing in each gene panel related to leukodystrophy.

## PgmNr 1309: The V122I mutation in hereditary transthyretin-mediated amyloidosis is significantly associated with polyneuropathy in UK Biobank.

### Authors:

P. Nioi <sup>1</sup>; S.M. Damrauer <sup>2</sup>; D.J. Rader <sup>2</sup>; S. Ticau <sup>1</sup>; D. Erbe <sup>1</sup>; M.M. Parker <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Research, Alnylam Pharmaceuticals, Cambridge, MA; 2) University of Pennsylvania, Philadelphia, PA

Hereditary transthyretin-mediated (hATTR) amyloidosis is a progressively debilitating, fatal disease caused by mutations in the transthyretin (*TTR*) gene. Historically, patients were identified by their predominant phenotype (polyneuropathy or cardiomyopathy), but evidence now suggests that a majority develop a mixed phenotype with both polyneuropathy and cardiomyopathy. The V122I (Val122Ile; p.V142I) variant is the most common pathogenic *TTR* mutation, primarily found in individuals of West African descent, and is thought to be predominantly associated with cardiomyopathy. We aimed to test the association of the V122I genotype and all available ICD10 diagnosis codes in the black subpopulation of UK Biobank (UKB) with replication in the Penn Medicine Biobank (PMBB).

The UKBB is a prospective cohort study with genetic, physical and health data on ~500,000 individuals across the United Kingdom. A phenome-wide association study was performed to test for association between the V122I genotype and 1,229 clinical diagnoses in the black subpopulation of UKB (n=6,063).

387 individuals heterozygous or homozygous for V122I were identified and were primarily of African or Caribbean descent (minor allele frequency in black subpopulation = 2%; baseline age: 40-70 years). This analysis revealed a significant association between the V122I genotype and a clinical diagnosis of polyneuropathy (OR = 11.2; 95% CI: 3.7-26.6; p=1.08x10<sup>-6</sup>). Replication analysis was performed in 5,737 black participants of the PMBB, of whom 190 individuals carried the V122I variant. The association of V122I with polyneuropathy was replicated (OR = 1.6; 95% CI: 1.2-2.4; p=0.006). In addition, there was nominally significant evidence that V122I carriers are at increased risk of other symptoms of hATTR amyloidosis, including carpal tunnel syndrome (OR = 2.0; p=0.02) and urinary retention (OR = 2.1; p=0.05) within the UKB.

These data indicate that carriers of the V122I mutation, historically assumed to cause a cardiac phenotype, have a significantly increased risk of polyneuropathy. This highlights the utility of phenome-wide association studies of known pathogenic variants to identify previously under appreciated manifestations of Mendelian disease. Additionally, this finding supports physicians having a clinical suspicion for the multisystem manifestations of hATTR amyloidosis, including both cardiomyopathy and polyneuropathy.

## PgmNr 1310: Development of a haploid mouse embryonic stem cell model of neuronal ceroid lipofuscinosis to enable forward genetic screens for disease modifiers.

### Authors:

E. Anderson; A. Hickey

[View Session](#) [Add to Schedule](#)

**Affiliation:** Catalyst for Rare Diseases, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC.

---

Neuronal Ceroid Lipofuscinosis (NCL), also referred to as Batten disease, is an autosomal recessive neurodegenerative disease and a type of lysosomal storage disorder. The classical form of Batten disease, juvenile onset NCL3, is caused by mutations in *CLN3*. Loss of *CLN3* cause numerous abnormalities in endolysosomal trafficking, leading to accumulation of autofluorescent storage material in lysosomes. Despite considerable study of *CLN3*, its function and pathways remain unknown, and there is an urgent need to identify new therapeutic targets. To enable forward genetic screens for modifiers of the NCL phenotype, we are developing models of NCL using mouse haploid mouse embryonic stem cells (mESCs). In these cells, the common *CLN3*<sup>Δex7/8</sup> mutation deleting exon 7 and 8 is introduced using CRISPR/Cas9 gene editing. We present our characterization of NCL phenotypes in neurons derived from *CLN3*<sup>Δex7/8</sup> haploid stem cells. These phenotypes are the basis for selection methods using fluorescence activated cell sorting (FACS) to isolate wild type and mutant genotypes from a population of cells based on the accumulation of autofluorescent storage material. Simulated genetic screens, using mixtures of patient-derived *CLN3*<sup>-</sup> and control induced pluripotent stem cell derived neurons, provide a proof-of-concept for future screens in mutagenized haploid mESC neurons. The haploid mESC *CLN3*<sup>Δex7/8</sup> stem cell model will enable forward genetic screens using mammalian neurons, to identify genetic modifiers of the NCL disease phenotype, disease-related pathways and potential drug targets.

## PgmNr 1311: A mix of 'old-school' and up-to-date familial genetic approaches led to identify a new genetic modifier of the age at onset in *C9orf72* frontotemporal dementia.

### Authors:

M. Barbier <sup>1,2</sup>; A. Camuzat <sup>1,2</sup>; F. Clot <sup>3</sup>; C. Cazeneuve <sup>3</sup>; D. Rinaldi <sup>1,2,4</sup>; A. Brice <sup>1,2</sup>; A. Sabbagh <sup>5</sup>; E. Genin <sup>6</sup>; I. Le Ber <sup>1,2,4</sup>; The French research network on FTD/FTD-ALS, Predict-PGRN, PREVDEMALS, FTLD-exome, IFGC, and EU EOD.

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Assistance Publique – Hôpitaux de Paris, Hôpital Pitié-Salpêtrière, Paris, France; 2) Institut du Cerveau et la Moelle épinière (ICM), Sorbonne Universités, UPMC Univ Paris 06, Inserm U1127, CNRS UMR 7225, Hôpital Pitié-Salpêtrière, Paris, France; 3) Unité Fonctionnelle de Neurogénétique Moléculaire et Cellulaire, Département de Génétique et Cytogénétique, AP-HP, Hôpitaux Universitaires La Pitié Salpêtrière-Charles Foix, Paris, France; 4) National reference center for rare or early dementias, Institute of Memory and Alzheimer's Disease (IM2A), Department of Neurology, AP-HP - Hôpital Pitié-Salpêtrière, Paris, France; 5) Institut de Recherche pour le Développement (IRD), UMR216 – MERIT, Paris, France; 6) INSERM, UMR1078, CHU Brest, Université Bretagne Occidentale, Brest, France

Frontotemporal dementia (FTD) is a rare neurodegenerative disorder, and the second cause of early onset dementia after Alzheimer's disease. FTD patients suffered from major behavior dysfunctions, associated with amyotrophic lateral sclerosis (ALS) in 15% of cases. Genetics architecture of autosomal dominant FTD has been almost completely defined. Since the major genes of FTD has been now discovered, the search for genetic modifiers, especially modifiers of Age at Onset (AO) has developed. Indeed, individuals carrying FTD associated mutations can develop symptoms from the third decade of life to nearly incomplete penetrance in elderly mutation carriers. This extensive variability of AO remains largely unexplained so far, especially in patients with *C9orf72* mutations, the major cause of familial FTD and ALS as well.

We previously showed that AO of FTD in *C9orf72* families was highly heritable using familial observations and variance component analyses. We took advantage from the availability of familial data and samples to perform linkage and association analyses among *C9orf72* relatives with either concordant or discordant AO. Genome-wide linear mixed models adjusted for sex, kinship, and covariates were notably used to assess the association between SNPs and AO. A suggestive p-value ( $p \leq 1.10^{-5}$ ) was observed for twelve SNPs, but only one was located on a region of positive linkage (LOD score = 2.1).

Independent association analyses of these twelve SNPs were performed thanks to data from the International FTD-Genomics Consortium including unrelated *C9orf72* patients. Only the association of the SNP located in the region of linkage was replicated ( $p = 0.009$ ). Extreme phenotype sampling was considered as an additional replication. Again, the same trend was observed for this SNP in a third cohort of *C9orf72* unrelated patients with early or late onset (logistic regression,  $p = 0.004$ ).

All of these converging results prompted us to focus on this strong candidate and the neighbor gene

which encodes for a synaptic adhesion protein. Through functional studies including the use of cerebral tissue, and *in vitro* models, we are defining the effect of this variants and we try to understand how this synaptic protein can influence the disease onset of FTD/ALS.

Altogether, this work illustrates how, in the context of a rare disease, family-based designs as first-line analyses may represent an alternative approach to detect genetic variants influencing disease-related traits.

## PgmNr 1312: Metachromatic leukodystrophy: Clinical, genetic, biochemical, and neuroradiological findings in a cohort of Brazilian patients.

### Authors:

F. Borges <sup>1</sup>; M.J. Costa <sup>1</sup>; A.C. Pessoni <sup>1</sup>; L. Vagnini <sup>2</sup>; K. Tirelli <sup>1</sup>; M. Burin <sup>3</sup>; R. Giugliani <sup>3</sup>; Z. Carneiro <sup>1</sup>; C. Lourenco <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centro Universitario Estacio, RIBEIRAO PRETO, SP, Brazil; 2) Centro Paulista de Diagnostico e Pesquisa - CPDP, Ribeirao Preto, Brazil; 3) Hospital de Clinicas de Porto Alegre, Porto Alegre, Brazil

---

**Background:** Metachromatic leukodystrophy (MLD) is a lyso-somal disorder affecting mainly peripheral and central nervous system, caused by arylsulfatase A (ARSA) deficiency. Usually It is classified into three forms according to the age of onset of symptoms (late infantile, juvenile, and adult). **Material and Methods:** Retrospective clinical, biochemical and neuroimaging data analysis of 15 MLD patients (9 male, 6 females). **Results:** Mean ages at onset of symptoms and at biochemical diagnosis were, respectively, 19 and 24 months (late infantile form). The most frequently reported first clinical symptom/sign of the disease was gait disturbance and other motor abnormalities infantile form) and behavioral and cognitive alterations (adult form). Diagnosis was achieved only after 2-3 years of symptoms onset (late infantile form). Leukocyte ARSA activity level did not present significant correlation with the age of onset of symptoms. All patients had high levels of sulfatides in urine (chromatography studies). Early dementia was the main feature of the two adult patients (with no motor feature associated). **Discussion:** Our results suggest that there is a considerable delay between the age of onset of signs and symptoms and the diagnosis of MLD in Brazil. Correlation between ARSA activity and MLD clinical form was not seen. Further understanding of MLD natural history, especially with larger samples, are necessary, in particular with the availability of specific treatments in a near future.

## PgmNr 1313: Recessive variants in Lithuanian hereditary neuropathy patients.

### Authors:

B. Burnyte <sup>1</sup>; K. Grigalioniene <sup>2</sup>; A. Morkuniene <sup>1,2</sup>; L. Ambrozaityte <sup>1,2</sup>; R. Buneviciute <sup>4</sup>; A. Vaitkevicius <sup>3,4</sup>; E. Preiksaitiene <sup>1,2</sup>; A. Matuleviciene <sup>1,2</sup>; L. Cimbalistiene <sup>1,2</sup>; V. Kucinskas <sup>1</sup>; A. Utkus <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human and Medical Genetics, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius, Lithuania; 2) Centre for Medical Genetics Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania; 3) Institute of Clinical Medicine, Faculty of Medicine, Vilnius University, Vilnius, Lithuania; 4) Center of Neurology of Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania

---

**Aim:** Charcot-Marie-Tooth disease (CMT) is a heterogeneous group of hereditary neuropathies (HN) that share common phenotypic features. Here, we describe genetic and phenotypic characteristics of patients with autosomal recessive CMT in a group of Lithuanian patients.

**Materials and methods:** Among 46 previously molecularly undiagnosed patients with CMT screened using multigene NGS panel during past 5 years, recessive variants have been identified for 9 unrelated patients. The phenotypes of all of them have been well-defined using HPO codes.

**Results:** Seven patients showed homozygous or compound heterozygous variants in *HINT1* gene, including previously reported pathogenic variant c.110G>C and a novel variant c.299A>G: i) four patients harboured a homozygous pathogenic variant c.110G>C, ii) two patients had compound heterozygous variants (c.110G>C and c.299A>G), and iii) one patient had a homozygous novel variant, which was also identified in affected brother. Additionally, one patient harboured compound heterozygous variants in *FIG4* (the most common pathogenic variant c.122T>C and novel deletion c.1703\_1704delAA), and one patient showed a homozygous pathogenic variant c.844C>T in *GDAP1*. Co-segregation of these variants in affected sibs or healthy relatives confirmed a recessive inheritance. Patients with homozygous novel and previously reported *HINT1* variants shared pure axonal motor neuropathy. A patient with biallelic *FIG4* variants presented an early onset severe demyelinating sensorimotor neuropathy. A patient with homozygous *GDAP1* variant had a severe axonal sensorimotor neuropathy.

**Conclusions:** The most prevalent recessive gene in Lithuanian patients with CMT is *HINT1*. Our data expand the spectrum of variants in *HINT1* and *FIG4* and their significance in molecular diagnosis.



## **PgmNr 1314: NeuroChip analysis of patients with Gaucher disease with and without Parkinson disease reveals potential protective variants in *CTSD* and *ACP2*: A pilot study.**

### **Authors:**

A. Eghbali <sup>1</sup>; C. Blauwendraat <sup>2</sup>; P. Sharma <sup>1</sup>; G. Lopez <sup>1</sup>; S. Scholz <sup>3</sup>; E. Sidransky <sup>1</sup>; N. Tayebi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Section on Molecular Neurogenetics, Medical Genetics Branch, NHGRI/NIH, Bethesda, MD; 2) Laboratory of Neurogenetics, NIA /NIH, Bethesda, MD; 3) Neurodegenerative Diseases Research Unit, NINDS, NIH/ Bethesda, MD

---

Gaucher disease is an autosomal recessive genetic disorder which results from mutations in *GBA1*. These mutations lead to the deficiency of the enzyme glucocerebrosidase and the consequent accumulation of the enzyme's substrates. Mutations in *GBA1* are the most prevalent risk factor for Parkinson disease (PD), the second most common neurodegenerative disease. The discovery of the link between Gaucher disease (GD) and Parkinsonism has led to new insights that have the potential to enable earlier diagnosis and better management of the disease. Despite this established connection, it is still not clear why only a minority of patients with GD develop PD, and it is likely that other genes or factors are involved in the GD-PD progression. The identification of causal or protective genetic variants is one approach to identify these factors. We collected DNA from patients with GD with and without PD. We then used NeuroChip genotyping and PLINK whole genome association analysis software to screen our GD and GD+PD cohorts for genetic variants in different lysosome-related genes. By performing Principle Component Analysis (PCA) we identified and removed the outliers to create a more homogeneous cohort. Conducting logistic regression corrected for sex and age, we identified two candidate genetic variants, rs17571 in *CTSD* and rs2167079 in *ACP2*, that we suggest may be protective against progression from GD to PD. Both rs17571 and rs2167079 were significantly more common in the GD group than in the GD+PD group (p-value <0.032 and <0.00524 respectively). GeneMANIA pathway and Protein-Protein network analyses confirmed potential interactions between the two genes and other established genes involved in GD and PD pathogenesis, either directly or through intermediate proteins. It will be important to further evaluate our findings in larger cohorts and to determine the impact of the identified variants at the RNA and protein level using RNA seq, quantitative PCR, immunoprecipitation, and Western blotting. Nonetheless, this work highlights the applicability of cheap and efficient genetic screening platforms such as NeuroChip in conducting genome-wide association analyses.

## PgmNr 1315: Movement-associated and developmental phenotypes of *MBD5*-associated neurodevelopmental disorder (MAND).

### Authors:

S.R. Gillipelli <sup>1,2</sup>; S.H. Elsea <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Biosciences, Rice University, Houston, TX

---

*MBD5*-associated neurodevelopmental disorder (MAND) is caused by abnormalities involving *MBD5*, a dosage-sensitive gene within chromosome 2q23.1. Individuals may have deletion, duplication or pathogenic variants that alter expression or function of *MBD5*. While initial studies of MAND included limited data for large cohorts of patients, the natural history of this rare condition is not well-described due to the limited young age range of these initial cohorts. Clinical features of MAND include global developmental delay/intellectual disability, severe speech impairment, seizures, sleep disturbances, and abnormal behaviors. Some individuals in the current patient population, now mostly in the teen and adolescent stage, recently presented with movement-associated phenotypes that were not present in the initial studies. In this study, we investigated how phenotypes of MAND have evolved over time and their overlap with symptoms of other movement-related disorders. A survey of patient caregivers (n=59) was developed to investigate current patient characteristics, including all aspects of the patient's movement, in order to address clinical features of cerebral palsy, ataxia, and factors of movement in MAND that cause obstacles for accurate diagnosis of possible underlying movement-associated disorders. Individuals with microdeletion of *MBD5* (n=42) were 50% female and ranged from age 1-24 years. Regression from developmental milestones was reported for 26% of patients, while 72% of patients were reported to have difficulties understanding their bodies within space. Of patients over age 2, 46% cannot run, while 47% of patients who cannot run had a BMI categorized as either overweight or obese by the Center for Disease Control. While 86% of patients reported full gross limb movement, 90% reported difficulty with fine motor tasks, such as eating, writing, or buttoning a shirt, and 79% reported poor coordination. Only 74% of patients were referred for occupational therapy, illustrating a therapeutic need for this population. Results revealed pronounced neurodegenerative effects of MAND, pointing to novel phenotypes related to movement difficulty and developmental regression. These data show that fine motor tasks are generally lost in the early teen years and that patients are not currently receiving sufficient therapeutic interventions to address developmental concerns. Additional studies are necessary to further investigate the neurodegenerative effects of MAND.

## PgmNr 1316: Singleton exome sequencing is an efficient tool for diagnosis of inherited white matter disorders.

### Authors:

P. Kaur <sup>1</sup>; N. Kausthubham <sup>1</sup>; A. Shrikiran <sup>2</sup>; R. Bhat <sup>2</sup>; L. Edward Lewis <sup>2</sup>; S. Nampoothiri <sup>3</sup>; S.J. Patil <sup>4</sup>; S. Sharma <sup>5</sup>; S. Bijarnia Mahay <sup>6</sup>; S. Pai <sup>7</sup>; A. Kaur <sup>8</sup>; K.C. Rakshith <sup>7</sup>; N. Kamath <sup>9</sup>; Z.K. Misri <sup>9</sup>; S. Bajaj <sup>10</sup>; A. Kumble <sup>11</sup>; R. Shetty <sup>12</sup>; R. Shenoy <sup>13</sup>; B. Banerjee <sup>14</sup>; M. Kamate <sup>15</sup>; B.S. Baliga <sup>9</sup>; K.M. Girisha <sup>1</sup>; A. Shukla <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India; 2) Department of Paediatrics, Kasturba Medical College, Manipal, Manipal Academy of Higher Education, Manipal, India; 3) Department of Paediatric Genetics, Amrita Institute of Medical Sciences and Research Centre, Kochi, India; 4) Division of Genetics, Mazumdar Shaw Medical Center, Narayana Health City, Bangalore, India; 5) Department of Paediatrics, Lady Hardinge Medical College, New Delhi, India; 6) Institute of Medical Genetics and Genomics, Sir Ganga Ram Hospital, New Delhi, India; 7) Department of Neurology, Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, India; 8) Department of Paediatrics, Postgraduate Institute of Medical Education and Research, Chandigarh, India; 9) Department of Paediatrics, Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal, India; 10) Jaslok Hospital and Research Centre, Mumbai, India; 11) Department of Paediatrics, Indiana Hospital and Heart Institute, Mangalore, India; 12) First Neuro Hospital, Mangalore, India; 13) Department of Paediatrics, K.S. Hegde Medical Academy, NITTE University, Mangalore, India; 14) Department of Paediatrics, Manipal Hospitals, Bangalore, India; 15) Department of Paediatrics, Jawaharlal Nehru Medical College, Belgaum, India

---

Inherited white matter disorders (IWMDs) are a group of neurodegenerative conditions that include leukodystrophies and genetic leukoencephalopathies. Leukodystrophies primarily involve defects in glial cells or myelin sheath. Genetic leukoencephalopathies (gLEs) have significant white matter abnormalities but do not meet inclusion criteria of a leukodystrophy. The true prevalence and incidence remain unknown worldwide. Cost-effectiveness, increasing accessibility and growing expertise have led to widespread application of whole exome sequencing (WES) in the clinic. We evaluated ninety-five individuals from eighty-eight families with white matter abnormalities. Among them, fifty-four were males and forty-one were females with age range of new-born to forty-two-years. Ten families (11.36%) had hypomyelination while other white matter pathologies were noted in remaining seventy-eight (88.63%). Nine families with specific clinical diagnosis underwent targeted genetic testing. Seventy-nine families, in absence of clinical diagnosis, underwent exome sequencing with singleton WES being performed in seventy-five families. Molecular diagnosis was established for fifty-five families (73.33%). Nine families (16.36%) were diagnosed with leukodystrophies and forty-six families (83.63%) with gLEs. Congenital muscular dystrophies (n=3, 6.52%), metabolic (n=17, 36.95%), mitochondrial (n=18, 39.13%), and other neurodegenerative disorders (n=8, 17.39%) accounted for gLEs. Thirty-three (63.46%) among the fifty-two disease-causing variants in this cohort are novel. Majority (39, 75%) of the observed variants are homozygous, eight (15.38%) are compound heterozygous, four (7.69%) are hemizygous and one (1.92%) is heterozygous (*de novo*). Three novel conditions with pathogenic variants in *ISCA1*, *AIMP2* and *TFIP11*

were identified. Possible founder events were noted in three genes, *ISCA1*, *AIMP2* and *BCKDHA*. Our population with background of consanguinity, inbreeding and founder effect, offers a high diagnostic yield for singleton WES in individuals with IWMDs, which are mostly autosomal recessive disorders. This is in line with our earlier observations on recessive Mendelian disorders. This approach appears to be highly cost-efficient for determining allelic and genetic heterogeneity of neurometabolic disorders. This approach proved highly effective for novel disease-gene discovery as well. We suggest this approach for resource-limited setting with a similar population structure.

## PgmNr 1317: Novel *C19orf12* pathogenic insertion variant in a Taiwanese family with autosomal dominant mitochondrial membrane protein-associated neurodegeneration (MPAN).

### Authors:

H.Y. Lin <sup>1</sup>; J.S. Hsu <sup>2</sup>; P.L. Chen <sup>3</sup>; C.H. Lin <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetic Counseling Program, Institute of Molecular Medicine, College of Medicine, National Taiwan University, Taipei, Taiwan; 2) Institute of Molecular and Genomic Medicine, National Health Research Institute, Taipei, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, College of Medicine, National Taiwan University, Taipei, Taiwan; 4) Department of Neurology, National Taiwan University Hospital, Taipei, Taiwan

---

**Background/Purpose:** Mitochondrial membrane protein-associated neurodegeneration (MPAN) is a rare form of neurodegeneration with brain iron accumulation (NBIA) caused by biallelic pathogenic variants in the *C19orf12* gene. Autosomal dominant inheritance by different heterozygous pathogenic variants has recently been demonstrated in a few families. We report a first Taiwanese autosomal-dominant MPAN family with a novel insertion variant in *C19orf12* and compare the inheritance pattern within different ethnicities.

**Methods:** Exome sequencing and segregation analysis were performed to identify the disease-causing variants of the index patient. The clinical and brain MRI characteristics of the patient were compared to those reported in previous studies of mitochondrial membrane protein-associated neurodegeneration (MPAN) from 2011 to 2019.

**Results:** A 30-year-old male presented with the progressive onset of ataxia, dystonia and parkinsonism since the age of 25. Brain MRI T2-weighted images displayed symmetrical hypointensities in bilateral globus pallidus (GP) and substantia nigra (SN). His 55-year-old mother also presented with progressive dystonia and cerebellar ataxia in her early thirties and now was wheelchair bounded. There was no consanguinity in his family. Exome sequencing revealed the heterozygous c.273\_274insA (p.P92Tfs\*9) insertion in *C19orf12* caused a novel frameshift and premature stop in the protein sequence, which was predicted to be pathogenic. This novel potential pathogenic variant co-segregated within his family among 4 members. Autosomal dominant MPAN was diagnosed according to the neuroradiologic and genetic findings. The novel c.273\_274insA (p.P92Tfs\*9) insertion variant was on the third exon of *C19orf12*, which many previously reported pathogenic heterozygous variants were located.

**Conclusion:** Our findings support recent observations that monoallelic *C19orf12* variants on the last exon may contribute to autosomal dominant MPAN. Variants occurred in final or penultimate exons typically escape from nonsense-mediated decay, suggesting haploinsufficiency might play a role in dominant disease-causing mechanism in MPAN.

## PgmNr 1318: Statistical profiling of clinical data in *GBA1* mutation carriers, an at-risk population for Parkinson disease.

### Authors:

G. Lopez <sup>1</sup>; A. Steward <sup>1</sup>; C. Groden <sup>1</sup>; E. Wiggs <sup>1</sup>; N. Tayebi <sup>1</sup>; J. Kim <sup>1</sup>; G. Monestime <sup>1</sup>; E. Ryan <sup>1</sup>; L. Segala <sup>2</sup>; L. Koehly <sup>3</sup>; E. Sidransky <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MGB, NHGRI (NIH) Bethesda, Maryland.; 2) OCD, NIMH (NIH), Bethesda, Maryland; 3) SBRB, NHGRI (NIH) Bethesda, Maryland.

---

The association of mutations in *GBA1*, the gene coding for the lysosomal enzyme glucocerebrosidase, and the development of Parkinson disease (PD) is well established, although only a subset of mutation carriers will develop PD. Several non-motor features of PD manifest years before motor manifestations become clinically apparent, but do not exclusively predict future development of parkinsonism. *GBA1* mutation carriers provide a risk-enriched population, ideal for longitudinal assessment of non-motor and motor features that could help identify individuals earlier in the disease process. Patients with homozygous and heterozygous *GBA1* mutations with and without Parkinson disease (n=128) were evaluated longitudinally at the NHGRI Genetics Clinic over a period of 12 years. Evaluations included complete physical and neurological examinations, validated mood scales (Geriatric Depression Scale, Fatigue Severity Scale, Epworth Sleepiness Scale) and motor evaluations (9-hole peg), olfactory testing (University of Pennsylvania Smell Identification Test), transcranial sonography, and neuropsychiatric testing. Between-group comparisons were used to assess clinical variation within the cohort. Differences in olfaction, cognitive function, and areas of echogenicity on transcranial sonography were found between the PD and non-PD groups. Hidden Markov models were explored as a hypothesis-blind method to identify latent states with unique clinical profiles that might identify patients on a neurodegenerative trajectory. Mixed effects logistic regressions were then employed to identify additional variables associated with state assignment. Longitudinal evaluation of patients at risk for Parkinson disease may help identify important clinical features and potential biomarkers associated with disease development. They may also elucidate pathways associated with neuroprotection.

## PgmNr 1319: Modulating phosphorylation of ATXN1's serine 776 residue: A potential therapeutic approach for spinocerebellar ataxia type 1.

### Authors:

L. Nitschke<sup>1,2,3</sup>; S.L. Coffin<sup>2,3</sup>; E. Xhako<sup>2,3</sup>; J.P. Orengo<sup>3,4</sup>; D. El-Najjar<sup>2,3,5</sup>; E. Alcala<sup>2,3</sup>; H.T. Orr<sup>6</sup>; H.Y. Zoghbi<sup>1,2,3,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Integrative Molecular and Biomedical Science, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Jan and Dan Duncan Neurological Research Institute, Houston, TX; 4) Department of Neurology, Baylor College of Medicine, Houston, TX; 5) Department of Biochemistry and Cell Biology, Rice University, Houston, TX; 6) Institute for Translational Neuroscience, University of Minnesota, Minneapolis, MN; 7) Howard Hughes Medical Institute, Houston, TX

Spinocerebellar Ataxia Type 1 (SCA1) is a fatal adult-onset neurodegenerative disorder characterized by motor incoordination and early lethality. It is caused by an expansion of CAG repeats encoding a polyglutamine (polyQ) tract in ATAXIN-1 (ATXN1). The polyQ expansion stabilizes ATXN1 leading to its toxic accumulation primarily in cerebellar and brainstem neurons. Thus far, no treatment options are available to prevent ATXN1's accumulation and the subsequent neuronal degeneration.

Previous studies identified several serine residues on ATXN1 that are phosphorylated. Phosphorylation of one of these sites, Serine 776 (S776), was shown to enable binding to 14-3-3, which in turn stabilizes ATXN1. Given the importance of ATXN1 levels in SCA1, we hypothesized that disrupting S776 phosphorylation would decrease ATXN1 levels in affected brain regions and potentially rescue SCA1 phenotypes. To test this, we modified an existing SCA1 knock-in mouse model (Atxn1<sup>154Q/2Q</sup>), which, similar to SCA1 patients, carries a CAG repeat expanded Atxn1 allele and recapitulates all major features of SCA1. Using CRISPR/Cas9, we replaced S776 with an alanine (S776A) generating S776 phosphorylation-deficient SCA1 mouse models. We found that the disruption of S776 phosphorylation does indeed disrupt 14-3-3 binding and leads to a ~20% decrease in ATXN1 levels throughout the entire brain. Furthermore, using molecular and behavioral assays, we demonstrate that disruption of S776 phosphorylation ameliorates the molecular changes in the cerebellum, reduces cerebellar Purkinje cell degeneration, and rescues the motor incoordination. More importantly, abolishing S776 phosphorylation also delays neuromuscular respiratory failure, an end-stage feature of SCA1 that contributes to premature death, and increases the lifespan of SCA1 animals by six weeks thus far.

This study highlights the importance of phosphorylation at S776 on ATXN1 stability and emphasizes the potential for targeting this phosphorylation site as a treatment option for SCA1. As S776 can be phosphorylated via multiple kinases, future studies are required to elucidate the predominant kinases to target in each brain region.

## PgmNr 1320: Molecular analysis of *PANK2* gene in Korean patient with suspected pantothenate kinase associated neurodegeneration.

### Authors:

D. Yang <sup>1</sup>; H. Gu <sup>2</sup>; M.J. Kim <sup>3</sup>; S.I. Cho <sup>3</sup>; M. Seong <sup>3,4</sup>; S.S. Park <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Laboratory Medicine, Kosin Gospel University Hospital, Busan, Korea; 2) Department of Laboratory Medicine, Kyung-Hee University Hospital, Seoul, Korea; 3) Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea; 4) Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea

---

**Background :** Neurodegeneration with brain iron accumulation (NBIA) describes a group of rare heterogeneous progressive neurodegenerative disorder characterized by excessive iron accumulation in the basal ganglia region. Pantothenate kinase associated neurodegeneration (PKAN) is a major form of this disease, which is caused by mutation of *PANK2* located on chromosome 20p13. Although the exact pathology of PKAN has not been established, it is assumed that defective *PANK2* enable to sufficiently convert pantothenate to 4'-phosphopantothenate, leading mitochondrial cysteine accumulation. In this study, we analyzed the *PANK2* gene in 26 Koreans among suspected PKAN patients.

**Method:** From August 2009 to February 2018, 26 unrelated Korean patients with suspicious PKAN were referred to the SNUH. Genomic DNA was extracted from the patient's EDTA peripheral blood. PCR was performed using the in-house primer. MLPA was performed using SALSA P120-B2 *PANK2* / *PLA2G6* Kits. Deletion breakpoint was re-evaluated using the in-house primers. To predict the mechanism of large deletion, RepeatMasker, and Clustal W were used.

**Result:** We identified nine variants in seven patients with Sanger sequencing. Seven of the nine identified variants were previously reported. 2 novel variants c.1210\_1214dupAATTA and c.1676C>G were respectively considered as likely pathogenic according to 2015 AGMG guidelines. As a result of MLPA, heterozygous exon 3-4 deletion was confirmed and in order to find the exact deletion breakpoint, a pair of in-house primers for sequencing was constructed. Sequence analysis revealed that exon 3-4 deletion was 5,016 bp deletion. The sequence of the breakpoint was found to be AluSx3 and AluSz6, and the sequence homology was about 97.3%.

**Conclusion:** In this study, we analyzed the *PANK2* gene in 26 Koreans among suspected PKAN patients. Exonic deletion of *PANK2* found in PKAN were mediated through *Alu*-mediated nonhomologous recombination.



## PgmNr 1321: *NEFH* founder mutation in Japanese autosomal dominant Charcot-Marie-Tooth disease and spinal muscular atrophy families.

### Authors:

M. Ando; Y. Okamoto; A. Yoshimura; Y. Higuchi; Y. Sakiyama; A. Hashiguchi; H. Takashima

[View Session](#) [Add to Schedule](#)

**Affiliation:** neurology and geriatrics, Kagosima university, Kagoshima, Kagoshima, Japan

### [Objective]

Mutations in neurofilaments genes are associated several neuromuscular disorders. Neurofilament heavy (*NEFH*) gene was known as a risk gene of sporadic ALS. In 2016, it was identified as the causative gene of Charcot-Marie-Tooth disease type 2. And mechanism was suggested toxic gain of function, which was caused by translation and aggregation of cryptic amyloidogenic element (CAE) in the 3' untranslated region (UTR) of *NEFH* gene. But the clinical and genetic spectrum of *NEFH* mutation is still unclear in Japan.

### [Method]

We analyzed whole exome sequencing data from the 891 patients in part of our genetic studies in neuromuscular disorders, included CMT, Spinal Muscular Atrophy (SMA). And we evaluated the identified *NEFH* variant according to the American College of Medical Genetics and Genomics (ACMG) standards and guidelines.

### [Result]

In this study, we identified novel *NEFH*c.3017dup p.Pro1007Alafs\*56 variant in three Japanese autosomal dominant CMT families and one Spinal Muscular Atrophy (SMA) family. And haplotype analysis of all four patients suggested founder event. This frameshift variant was absent in global, Japanese, and in-house control databases. In the segregation study, this frameshift variant was confirmed in other affected family members.

### [Conclusion]

Our study is the first report that referred to the spinal muscular atrophy (SMA) phenotype and the founder mutation in *NEFH*. The clinical diversity in our study, included CMT, CMT with pyramidal sign, and SMA, suggested that the causative lesions are peripheral nerve or spinal cord neuron or both. And it is important for future clinical and genetic analysis of *NEFH* mutation.

## PgmNr 1322: Analysis of the *TMEM230* gene in familial Parkinson's disease from south Italy.

### Authors:

G. Annesi <sup>1</sup>; R. Procopio <sup>1,2</sup>; M. Gagliardi <sup>1</sup>; G. Nicoletti <sup>1</sup>; M. Morelli <sup>2</sup>; A. Quattrone <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Molecular Bioimaging and Physiology, National Research Council, Section of Germaneto, Catanzaro, Italy; 2) Institute of Neurology, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy; 3) Neuroscience Research Center, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy

To date, more than 20 genes have been reported as involved in familial Parkinson's disease (PD). Recently, mutations in *TMEM230* exon 5 have been associated to PD in a large family from North America and in seven small Chinese families by *Deng et al.*

*TMEM230* encodes transmembrane protein 230, which belongs to the TMEM134/TMEM230 family, and it is involved in endocytosis and trafficking secretory/recycling vesicles. It is located on chromosome 20p13-p12.3 and contains five exons. Disease-linked mutations impair normal synaptic vesicle trafficking and the presence of TMEM230 protein in alpha-synuclein-positive Lewy bodies and Lewy neurites in midbrain and neocortex sections from sporadic PD cases gave supporting evidence for a role of this gene in PD pathology. However, in 15 studies published only approximately 0.28% PD patients were found to harbor potential PD-related variants with full detection information of coding regions of the *TMEM230* gene.

In order to clarify the relationship between *TMEM230* and southern Italy PD population, we performed a mutational screening of this gene in 168 autosomal dominant familial PD patients, defined by the presence of PD in at least other first degree, second degree, and third degree relative, and a control group consisting of 500 subjects from the same geographical area. After obtaining informed consent, genomic DNA was extracted from peripheral blood by standard method. All the 5 exons of *TMEM230* and intron-exon boundaries region were sequenced using an ABI 3500 Genetic Analyzed (Life Technologies, Carlsbad, CA, USA).

The screening highlighted the presence of four synonymous variants (rs186628284, rs6116651, rs763383477, rs6107576) with the same frequency in both population (PD cohort and control subjects) and the p.Ile125Met (c.375A>G, rs148033002) missense variant in a 64-years old woman. The patient (AAO: 39 years) who carries this variant in heterozygous state presented bradykinesia, rigidity and postural instability with a good response to levodopa treatment. Family history was positive with her mother and her two sisters with a diagnosis of PD. The genetic testing was extended to three sisters of the proband and their mother, and it confirmed the segregation of the c.375A>G variant in the affected members of the family.

Further studies are needed to evaluate the role of *TMEM230* in the pathogenesis of Parkinson's disease.

## PgmNr 1323: Exploring the molecular basis of hereditary spinocerebellar degeneration in a large Sudanese family.

### Authors:

A. Babai <sup>1,2</sup>; L. Elsayed <sup>2</sup>; I. Mohammed <sup>3</sup>; A. Hamed <sup>3</sup>; M. Elseed <sup>3</sup>; E. Alturabi <sup>4</sup>; I. Eltazi <sup>4</sup>; A. Mohammed <sup>5</sup>; S. Taha <sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Medicine, Institute of Endemic Diseases, Khartoum, Khartoum, Sudan; 2) Department of Biochemistry, Faculty of Medicine, University of Khartoum, Sudan; 3) Department of paediatrics and child health, Faculty of Medicine, University of Khartoum, Sudan; 4) Faculty of Medicine, University of Khartoum, Sudan; 5) Institut du Cerveau et de la Moelle épinière, INSERM U1127, CNRS UMR7225, Sorbonne Universités UMR\_S1127; 6) Department of Radiology, Dar Al Elaj specialized hospital, Khartoum, Sudan

---

**Background:** Spinocerebellar neurodegenerative disorders (SCD) are known for their complex phenotypic and genetic heterogeneity forming a heterogeneous spectrum of disorders with hereditary spastic paraplegias (HSP) on one end and hereditary ataxias (HA) on the other. In clinical practice, limb spastic weakness and cerebellar ataxia are frequently found together and present the hallmark of SCD. The genetics of SCD has been a target for extensive researches in many parts of the world, yet little is known about the genetics of SCD in Sub-Saharan African population.

**Methods:** In this study, we recruited a large consanguineous Sudanese family with five affected siblings. Genomic DNA was extracted and screened for genetic variations using whole exome sequencing (WES). Analysis was done to identify the culprit variations using bioinformatics tools and in-silico prediction of variants pathogenicity.

**Results:** Clinical results showed a complex phenotype of progressive spastic-ataxia complicated with deafness. Microcephaly was detected in the two eldest patients. Analysis of WES data and variant prioritization suggested two homozygous missense variants in two candidate genes (*MYO15A* and *SEMA5D*) that were not reported to be linked to similar disease before. The first variant in *MYO15A* gene (NM\_016239.3: c.1634C>T) was reported to cause autosomal recessive hearing loss but was not reported to similar neurological disease. The second variant (NM\_006378.3:c.1588G>A) was in *SEMA4D* gene which involved in brain development but not reported to be associated with inherited neurological conditions before. Both variants were extremely rare and highly conserved. They were predicted to be highly pathogenic using bioinformatics tools.

**Conclusion:** The scarcity of genetic data in the highly consanguineous Sudanese population makes whole exome sequencing a powerful and cost effective strategy to identify both known and new pathogenic variations and genes. Sanger sequencing and further functional studies are recommended to prove the association of *MYO15A* gene and *SEMA4D* gene with the complex clinical phenotype of deafness, spasticity and ataxia.

## PgmNr 1324: Autosomal recessive cerebellar ataxia (ARCA): Clinic and genetic aspects in Mali in the Department of Neurology, Teaching Hospital of Point G, Bamako, Mali.

### Authors:

C. Cisse <sup>1,2</sup>; S.H. Diallo <sup>1,3</sup>; A. Tamega <sup>1,2</sup>; T. Coulibaly <sup>1,2</sup>; S. Diallo <sup>1,3</sup>; S. Diarra <sup>1,2,4</sup>; D. Coulibaly <sup>5</sup>; K. Dembélé <sup>2</sup>; M. Ouologuem <sup>6</sup>; H.O. Bah <sup>7</sup>; A. Simaga <sup>8</sup>; C.O. Guinto <sup>1,2</sup>; K. Fischbeck <sup>4</sup>; G. Landouré <sup>1,2,4</sup>; H3Africa Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurogenetic, University of Sciences Techniques and Technology of Bamako, Bamako, Mali; 2) Neurology service, teaching hospital of point g, Bamako, Mali; 3) Neurology service, teaching hospital of Gabriel Touré, Bamako, Mali; 4) Neurogenetics Branch, NINDS, NIH, Bethesda, MD, Bamako, Mali; 5) Internal medicine, teaching hospital of Mère-Enfant le Luxembourg, Bamako, Mali; 6) Internal medicine, hospital of Kati, Bamako, Mali; 7) Cardiology service, teaching hospital of Gabriel Touré, Bamako, Mali; 8) African Ophtalmological and Tropical Institut (IOTA) Bamako, Mali

**Introduction:** Autosomal recessive cerebellar ataxia (ARCA) is a group of rare heterogeneous neurodegenerative disorders characterized by the involvement of both central and peripheral nervous system often associated with non-neurological signs such as pes cavus, diabetes, and ophthalmological features. They are complex diseases with an onset before the age of 30, worsening progressively and causing a major handicap. Very few studies have been reported in Sub-Saharan-Africa, and none in Mali.

**Objectives:** We aim to characterize clinically patients with ARCA, and to determine the underlying genetic defect.

**Methods:** Patients with ARCA phenotype were seen in our Neurogenetics Clinic and by ophthalmologist, cardiologist and ENT specialists. DNA was collected for genetic analysis. Brain and spinal cord imaging, electromyography and blood chemistries were performed to exclude other causes.

**Results:** We have enrolled six families totaling twelve patients, and the age of onset ranged from 1 to 21 years old. Alpha fetoprotein levels were normal in all patients. Brain MRI showed cerebellar atrophy in some patients while other had hearing impairment. *FXN* gene was sequenced in three patients (from three families) with Friedreich's ataxia-like clinical presentation, and only one patient was tested positive (999 and 766 repetition of GAA in the respective alleles of the gene). Subsequently, whole exome sequencing was performed in two families, and identified six homozygous variants in three genes in one family. Three variants were located in the *ATM* gene causing ataxia telangiectasia, two in *ANO10* causing autosomal recessive cerebellar ataxia type 3, and one in *SIL1* causing Marinesco-Sjögren syndrome that also manifests as recessive ataxia. All these variants segregated with the disease in the sequenced family members (patient, father, maternal aunt, brother). Analysis of the remaining families is underway.

**Conclusion:** We have diagnosed a rare Friedreich's ataxia case in sub-Saharan Africa. In addition, there is potential to expand the clinical presentation and genetic epidemiology of ARCA. Our future studies may indicate which of the three genes causes the disease in the family.

**Key words:** ARCA, genetic testing, variants, Mali.



## PgmNr 1325: Pediatric neurotransmitter diseases in Japan.

### Authors:

H. Fujioka <sup>1,2</sup>; H. Shintaku <sup>2,3</sup>; N. Nakano <sup>2</sup>; T. Sakaguchi <sup>2</sup>; T. Hattori <sup>2,3</sup>; K. Tanaka <sup>2,3</sup>; T. Hamazaki <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Health Science, Kansai University of Welfare Sciences, Kashiwara, Osaka, Japan; 2) Department of Pediatrics, Osaka City University Graduate School of Medicine, Japan; 3) Donated Course Disability medicine and Regenerative medicine , Osaka City University Graduate School of Medicine, Japan

---

**Background:** Segawa disease is a very rare inherited disease characterized by dopa-responsive dystonia with diurnal fluctuation. Partial defects of GTP cyclohydrolase (GTPCH I), which is the rate-limiting enzyme of bipterin synthesis is the cause of this disease. In Segawa disease patients, the amount of dopamine in their central nervous system (CNS) is diminished because shortage of bipterin decrease tyrosine hydroxylase (TH) activity and it suppress production of L-dopa from tyrosine. Some other pediatric neurotransmitter diseases such as sepiapterin reductase (SR) deficiency and TH deficiency also show dopa-responsive dystonic symptom.

**Method:** We examined 137 patients with dystonia and/ or some other involuntary movements from January 2012 to December 2016. We measured bipterin and neopterin concentration in their blood and cerebrospinal fluids. Genetic analysis of *GCH1* gene which was coding GTPCH I was performed in 50 patients, respectively. Some patients who wanted to diagnose SP deficiency or TH deficiency, we also performed genetic analysis of *SPR* and *TH* genes, which are coding genes of SR and TH enzymes.

**Results:** We detected 22 Segawa disease patients with mutations in *GCH1* gene. Segawa disease patients comprised 18 females and 4 males. The average age of patients who underwent genetic analysis was 19.35 years (Range: 7 to 62). Most of the patients who were diagnosed as adults showed dystonic symptoms from a very young age. No common mutation was observed in Segawa disease patients. In this study, Among the rest of the patients, two patients of SR deficiency and one patient of TH deficiency were also detected. In this study, we examined patients from 32 of the 47 prefectures in Japan (total population of these areas was 106,058,400 in 2016). The calculated incidence rate of Segawa disease was 4.1/100,000,000 per year in this study. The prevalence rate of Segawa disease was considered to be 1.6/1,000,000 when the average of duration of this disease was 40 years.

**Conclusion:** Over the course of five years, we genetically diagnosed patients with Segawa disease, SR deficiency and TH deficiency. No common mutation was observed. In Segawa disease the incidence rate was 4.1 per hundred million people per year and the prevalence rate was 1.6 per million people in Japan.

## PgmNr 1326: Mutational analysis of *SOST* in the first Italian family with sclerosteosis.

### Authors:

M. Gagliardi <sup>1</sup>; F. Bono <sup>2</sup>; R. Procopio <sup>1,2</sup>; A. Quattrone <sup>1,3</sup>; G. Annesi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Molecular Bioimaging and Physiology, National Research Council, Catanzaro, Italy Germaneto (CZ), Italy; 2) Institute of Neurology, Magna Graecia University of Catanzaro, Italy; 3) Neuroscience Research Center, Department of Medical and Surgical Sciences, Magna Graecia University of Catanzaro, Italy

Sclerosteosis is a rare autosomal recessive condition characterized by progressive bone thickening due to increased bone formation. Mutations in *SOST* gene coding for sclerostin are linked to sclerosteosis. There are the clinical and radiographic similarities between sclerosteosis and van Buchem disease, another autosomal recessive bone dysplasia. Sclerosteosis and van Buchem disease are linked to chromosome 17q12-q21, more just van Buchem disease is associated with a 52 kb deletion at downstream of the *SOST* gene, sclerosteosis is caused by loss-of-function mutation of sclerostin. This protein is the *SOST* gene product, it secreted by osteocytes and transported to the bone surface where it inhibits osteoblastic bone formation by antagonizing Wnt signalling. Very few sporadic cases have been described around the world mainly in Africa and South America. Here we described for the first time a South Italy patient with a mutation sclerosteosis.

Informed consent was obtained from the patients. The diagnosis of sclerosteosis was based on a history of chronic headache and nocturnal snoring documented by neurological and audiometric examination, Brain MRI, MR venography, polysomnography and 1-hour lumbar CSF pressure monitoring. Genomic DNA was extracted from peripheral blood by standard methods. The *SOST* purified PCR products were analyzed on 3500 Genetec Analyzer.

The two brothers of 45 and 36 years old presented macrocephaly and an enlarged mandible, they complained unilateral visual acuity loss and unilateral deafness since adolescence. Moreover the youngest one had a bilateral facial nerve palsy since childhood. The first brother was not interest to continue the neurological and genetic investigations. A nocturnal polysomnography showed a condition of Obstructive Sleep Apnea Syndrome (OSAS) of moderate entity. We detected a nonsense mutation, p.Gln24X (c.70C>T), in homozygous in our patient and also in heterozygous state in his parents, two brothers, one sister, two sons and one nephew. This nonsense mutation is predicted to exert a deleterious effect by Mutation Taster Server, and lead to premature termination of the protein. The bioinformatics analysis using amino acidic alignment attested that the Glutamine at position 24 is evolutionary conserved in Vertebrates. The clinical features of this patient show the presence of OSAS causing intracranial hypertension in Sclerosteosis, these symptoms enlarge the clinical spectrum of Sclerosteosis.



## PgmNr 1327: Recurrent *de novo* MAPK8IP3 variants cause neurological phenotypes.

### Authors:

S. Iwasawa <sup>1</sup>; K. Yanagi <sup>2</sup>; A. Kikuchi <sup>1</sup>; Y. Kobayashi <sup>3,4</sup>; K. Haginoya <sup>4,5</sup>; H. Matsumoto <sup>6</sup>; K. Kurosawa <sup>7</sup>; M. Ochiai <sup>8</sup>; Y. Sakaki <sup>8</sup>; A. Fujita <sup>9</sup>; N. Miyake <sup>9</sup>; T. Niihori <sup>10</sup>; M. Shirota <sup>11</sup>; R. Funayama <sup>12</sup>; S. Nonoyama <sup>6</sup>; S. Ohga <sup>8</sup>; H. Kawame <sup>13</sup>; K. Nakayama <sup>12</sup>; Y. Aoki <sup>10</sup>; N. Matsumoto <sup>9</sup>; T. Kaname <sup>2</sup>; Y. Matsubara <sup>2</sup>; W. Shoji <sup>14</sup>; S. Kure <sup>1,13</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Tohoku University, Sendai Aobaku, Miyagi, Japan; 2) National Center for Child Health and Development, Tokyo, Japan; 3) Department of Pediatrics, National Hospital Organization Sendai-Nishitaga Hospital, Sendai, Japan; 4) Department of Pediatric Neurology, Takuto Rehabilitation Center for Children, Sendai, Japan; 5) Department of Pediatric Neurology, Miyagi Children's Hospital, Sendai, Japan; 6) Department of Pediatrics, National Defense Medical College, Saitama, Japan; 7) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 8) Department of Pediatrics, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 9) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 10) Department of Medical Genetics, Tohoku University Graduate School of Medicine, Sendai, Japan; 11) Division of Interdisciplinary Medical Sciences, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 12) Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 13) Tohoku Medical Megabank organization, Tohoku University, Sendai, Japan; 14) Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai, Japan

Kinesin superfamily proteins play a fundamental role in intracellular transport such as axonal transport. Disturbed axonal transport has been associated with various human neurodegenerative diseases. JIP3, encoded by *MAPK8IP3*, is an adaptor protein of the kinesin-1 complex and essential for axonal transport in neurons. However, an association between *MAPK8IP3* variants and human disease has not been established. We identified two recurrent *de novo* *MAPK8IP3* variants in all five patients from four unrelated families. A *de novo* variant, c.1732C>T(p.Arg578Cys), was found in three individuals, including two siblings. The other variant, c.3436C>T(p.Arg1146Cys), was found in two individuals. The core phenotype includes spastic diplegia, intellectual disability, cerebral atrophy, and corpus callosum hypoplasia.

The two variants, p.Arg578Cys and p.Arg1146Cys, are absent in ExAC, gnomAD, Human Genetic Variation Database (HGVD), and ToMMo 3.5KJPNv2. Both missense variants also affect evolutionarily conserved residues and are predicted to be pathogenic by different *in silico* prediction tools (SIFT, Polyphen-2, MutationTaster, and CADD).

To test the pathogenicity of the two variants, we utilized a zebrafish model. The pLL nerve emerging from the ganglion beside the otocyst results in long axons along the mid-body line toward the tail of the zebrafish. We examined if the expression of JIP3 variants affects developing pLL axons because an abnormality of the lateral line

reportedly occurs in JIP3<sup>-/-</sup> zebrafish. A smooth long axon was observed when wild-type human JIP3 tagged with mCherry was transiently expressed in a single pLL neuron. When the variant JIP3<sup>R578C</sup> or



JIP3<sup>R1146C</sup> was expressed in a pLL neuron, abnormal varicosities were exhibited in the axon. These findings support the pathogenicity of the variants.  
Thus, we propose that *MAPK8IP3* missense variants cause a human neurodevelopmental disease.

## PgmNr 1328: Cerebral microbleeds and the risk of intracranial hemorrhage in Taiwanese patients with CADASIL.

### Authors:

Y. Liao <sup>1,2</sup>; Y. Hu <sup>1,2</sup>; F. Chang <sup>3</sup>; Y. Lee <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurology department, Taipei Veterans General Hosp, Taipei, Taiwan; 2) Department of Neurology, National Yang-Ming University School of Medicine, Taipei, Taiwan;; 3) Department of Radiology, Taipei Veterans General Hospital, Taipei, Taiwan;

### Background

Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) is the most common monogenic cerebral small vessel disease worldwide. Intracerebral hemorrhage (ICH) was rare in Caucasian patients with CADASIL; however, 12-25% of CADASIL patients in East Asian populations suffered from ICH. The present study aims at investigating the frequency and predictors of ICH in CADASIL patients.

### Method

One hundred and nine genetically confirmed CADASIL patients from 86 families were included. All of them had underwent brain MRI with T2 star weighted (T2\*) or susceptibility-weighted angiography (SWAN) images. Neuroradiological features, including cerebral microbleeds (CMBs), lacunar infarcts, white matter hyperintensity and dilated perivascular space, were analyzed by visual inspection. Multivariate logistic regression was conducted to identify factors predicting the occurrence of ICH in CADASIL.

### Results

Twenty four of the 109 CADASIL patients (22.0%) had ICH on T2\* or SWAN images. A total of 33 ICH lesions were identified, including 15 symptomatic lesions with corresponding neurological deficits and 18 asymptomatic ICH. Ten of the 18 asymptomatic lesions and three of the 15 symptomatic lesions were lobar hemorrhage (55.6% vs. 20.0%,  $p = 0.04$ ), suggesting lobar hemorrhage more likely to be clinically unnoticed. CADASIL patients with ICH more frequently had a history of hypertension, lacunes in basal ganglia, thalami and infratentorial regions, and CMBs in all brain regions than those without ICH. Up to 83.3% of the CADASIL patients with ICH and 42.2% of patients without ICH had the number of CMBs higher than 10 ( $p = 4.2 \times 10^{-4}$ ). In the multi-variate regression analysis, presence of CMBs in the brainstem and age at examination were the strongest predictors for the occurrence of ICH in CADASIL patients.

### Conclusion

In contrast with as a rare phenomenon in Caucasian, ICH occurred in one fifth of the CADASIL patients in Taiwan. There was no significant difference in the clinical features between CADASIL patients with symptomatic ICH lesions and those with asymptomatic lesions, except for a higher percentage of lobar hemorrhage in asymptomatic ones. Age and presence of CMBs in the brainstem were significant predictors for the occurrence of ICH in CADASIL patients. Clinical vigilance, strict controls of hypertension, and cautious use of antithrombotic agents/anticoagulants are warranted in CADASIL

patients with high ICH risks.

## PgmNr 1329: Biallelic pathogenic variants in *TUBGCP2* cause microcephaly and lissencephaly spectrum disorders in humans.

### Authors:

T. Mitani<sup>1</sup>; J. Punetha<sup>1</sup>; I. Akalin<sup>2</sup>; D. Pehlivan<sup>1</sup>; Z.C. Akdemir<sup>1</sup>; J.V. Hunter<sup>3</sup>; H. Hijazi<sup>1</sup>; C.M. Grochowski<sup>1</sup>; J.N. Jhangiani<sup>4</sup>; J.M. Fatih<sup>1</sup>; P. Iwanowski<sup>5</sup>; T. Gambin<sup>6,7</sup>; M. Dawidziuk<sup>7</sup>; M. Bekiesinska-Figatowska<sup>8</sup>; M. Hosseini<sup>9</sup>; S. Arzhanghi<sup>9</sup>; H. Najmabadi<sup>9</sup>; J.A. Rosenfeld<sup>1</sup>; H. Du<sup>1</sup>; S. Blaser<sup>10</sup>; R. Teitelbaum<sup>11</sup>; R. Silver<sup>11</sup>; J.E. Posey<sup>1</sup>; H. Ropers<sup>12</sup>; R.A. Gibbs<sup>1,4</sup>; W. Wiszniewski<sup>7,13</sup>; J.R. Lupski<sup>1,14,15</sup>; D. Chitayat<sup>11,16</sup>; K. Kahrizi<sup>9</sup>; P. Gawlinski<sup>7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Department of Medical Genetics, Istanbul Medeniyet University, Faculty of Medicine, Istanbul; 3) Department of Radiology, Baylor College of Medicine, Houston, Texas; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 5) Department of Medical Genetics, Children's Memorial Health Institute, Warsaw; 6) Institute of Computer Science, Warsaw University of Technology, Warsaw; 7) Department of Medical Genetics, Institute of Mother and Child, Warsaw; 8) Department of Diagnostic Imaging, Institute of Mother and Child, Warsaw; 9) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran; 10) Department of Diagnostic Imaging, Division of Pediatric Neuroradiology, The Hospital for Sick Children, University of Toronto, Toronto; 11) The Prenatal Diagnosis and Medical Genetics Program, Department of Obstetrics and Gynecology, Mount Sinai Hospital, University of Toronto, Toronto; 12) Institute for Human Genetics, University Medicine Mainz, Mainz; 13) Department of Molecular and Medical Genetics, Oregon Health & Sciences University, Portland, Oregon; 14) Department of Pediatrics, Baylor College of Medicine, Houston; 15) Texas Children's Hospital, Houston, Texas; 16) Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children, University of Toronto

---

The recent progress in DNA sequencing and brain imaging technologies has facilitated the discovery of a number of causative genes for malformations of cortical development (MCDs). Lissencephaly is a subtype of MCDs that can manifest over a range of clinical severity, including agyria, pachygyria and subcortical band heterotopia; each representing anatomical malformations of brain cortical development caused by neuronal migration defects. Genetic causes of neuronal migration anomalies are highly enriched for genes encoding microtubules and microtubule-associated proteins, which highlight the essential functions of the cytoskeleton in the developing human brain. Using exome sequencing and family-based rare variant analyses, we identified a homozygous variant c.997C>T; p.(Arg333Cys) in the gene for tubulin-gamma complex-associated protein 2, *TUBGCP2*, in two individuals from a consanguineous family presenting with microcephaly and developmental delay. *TUBGCP2* encodes GCP2 which forms the multiprotein g-tubulin ring complex (g-TuRC) together with g-tubulin and GCP3-6 to template the microtubule formation. Moreover, we have observed intrafamilial clinical variability in these individuals. Our detailed molecular studies revealed dual molecular diagnoses in the more severely affected sibling due to *TUBGCP2*-associated disease and a *de novo* 2q23.1 duplication. We found three additional families with rare biallelic variation in *TUBGCP2* and similarly affected phenotypes by querying the clinical exome sequencing database at Baylor

Genetics and through GeneMatcher. These include one homozygous variant c.1843G>C; p.(Ala615Pro) in two families and compound heterozygous variants consisting of one missense variant c.889C>T; p.(Arg297Cys) and one splice variant c.2025-2A>G in one family. Affected individuals displayed the clinical spectrum, ranging from pachygyria with mild learning disability to fatal microlissencephaly. Our data establish *TUBGCP2* as a novel disease gene for an autosomal recessive neurodevelopmental trait consisting of a neuronal migration disorder and implicate GCP2 as a core component of g-TuRC involved in both neuronal progenitor and migrating cells in humans.

## PgmNr 1330: Frontotemporal dementia: A novel *TUBA4A* variant identified in a Dutch family with TDP-43 pathology.

### Authors:

M.O. Mol; T. Hang Wong; J.G.J. van Rooij; S. Melhem; J.C. van Swieten

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dept. of Neurology, Erasmus Medical Centre, Rotterdam, The Netherlands

---

Frontotemporal dementia (FTD) is the second most common type of presenile dementia, encompassing a spectrum of clinical disorders including behavioral and language impairment. FTD shares pathophysiology and genetic etiology with amyotrophic lateral sclerosis (ALS). Approximately 40% of FTD is familial. Autosomal dominant mutations in several main genes (*C9orf72*, *GRN*, and *MAPT*) have been identified, though a substantial proportion (~45%) of familial cases remains unsolved.

Here we describe a Dutch pedigree consisting of five affected and one elderly unaffected individual. Patients presented with dementia without motor neuron disease, of which three received a definite diagnosis of behavioral FTD. Autopsy of the proband revealed frontotemporal lobar degeneration with TDP-43 pathology, not neatly fitting in a specific subtype. Genetic testing excluded mutations in *C9orf72*, *GRN*, and *MAPT*. Through whole exome sequencing (WES) on all six relatives, we identified a heterozygous c.313C>T (p.R105C) missense variant in *TUBA4A*.

This gene encodes for the Tubulin Alpha 4A protein, a major structural component of the microtubules. Mutations in *TUBA4A* have been previously implicated in familial ALS. The pathogenicity of our mutation is supported by segregation with the disease, absence in ExAc/GnomAD, and a CADD score of 33 (24 variants also passed these filters). Immunohistochemistry with *TUBA4A* antibodies revealed abundant extracellular deposits and a few cytoplasmic inclusions. Exposing cells with or without the mutation to low doses of nocodazole did not demonstrate a difference in the stability of the microtubules, which we did observe for the ALS-causing *TUBA4A* mutation (R320C). Follow-up with additional analyses, e.g. Western blotting, will be performed in order to proof the pathogenicity of the variant.

In conclusion, we describe a Dutch FTD family with predominantly behavioral symptoms and TDP-pathology possibly caused by a mutation in *TUBA4A*. The mechanism of pathogenicity is not clear and still under investigation.

## PgmNr 1331: Rare syndromes with X-linked cerebellar hypoplasia revealed by whole-exome sequencing.

### Authors:

N. Okamoto <sup>1,2</sup>; F. Miya <sup>3,4</sup>; T. Tsunoda <sup>3,4</sup>; M. Kato <sup>5</sup>; S. Saitoh <sup>6</sup>; Y. Kanemura <sup>7,8</sup>; K. Kosaki <sup>9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Medical Genetics, Osaka Women's and Children's Hospital, Osaka, Japan; 2) 2Department of Molecular Medicine, Research Institute, Osaka Women's and Children's Hospital, Osaka, Japan; 3) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 4) Laboratory for Medical Science Mathematics, Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 5) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan; 6) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 7) Department of Biomedical Research and Innovation, Institute for Clinical Research, National Hospital Organization Osaka National Hospital, Osaka, Japan; 8) Department of Neurosurgery, National Hospital Organization Osaka National Hospital, Osaka, Japan; 9) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan

---

X-linked cerebellar hypoplasia (XLCH) is a heterogeneous nonprogressive genetic disorder characterized by intellectual disability (ID) and delay in motor development. The well-known causes of XLCH include oligophrenin-1 syndrome,<sup>?</sup>and MICPCH/CASK syndrome (Mental retardation and microcephaly with pontine and cerebellar hypoplasia, calcium/calmodulin-dependent serine protein kinase. We studied four pedigrees with ID by WES and identified pathogenic variants in X-linked genes including *ABCB7*, *ATP2B3*, *HNRNPH2*, and *ZNF41*.

Patient 1 and 2 are brothers with ID and cerebellar ataxia. Brain MRI revealed atrophic cerebellar vermis. A missense variant was identified in *ABCB7*?(c.1256A>T:p.Q419L). *ABCB7* variants were reported in patients with sideroblastic anemia and ataxia. Protasova et al. [2016] found a missense mutation in the *ABCB7* in a pedigree with XLCH with neurodevelopmental phenotype. The pedigree did not show sideroblastic anemia. Patient 3 is a 5-year-old male proband with *ATP2B3* variant (c.982G>T: p.A328S). Variants in the *ATP2B3* gene, calcium-transporting ATPase predominantly expressed in the brain, are associated with X-linked spinocerebellar ataxia-1. Patient 4 had a *de novo* missense variant in *HNRNPH2* (c.89T>G: p.F30C).The *HNRNPH2* gene belongs to a member of the heterogeneous nuclear ribonucleoprotein family and is located on Xq22.1. Bain et al. [2016] reported six unrelated females with new neurogenetic syndrome with pathogenic *HNRNPH2* variants. This is the first male patient with a *HNRNPH2* mutation. Patient 5 is a 5-year old boy with cerebellar atrophy and ID. He had a mutation in *ZNF41* (c.2279C>G: p.S760W). This gene is associated with X-linked ID [Shoichet et al. 2003]. Association of *ZNF41* variants and cerebellar hypoplasia has not been reported. These genetic syndromes should be included in the differential diagnosis of XLCH.

## PgmNr 1332: Phenotypic similarities in one case with a microdeletion and two cases with single nucleotide variants in *BCL11A*: Cerebellar expressions of motor speech disorders.

### Authors:

B. Peter <sup>1,2</sup>; L. Bruce <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Speech & Hearing Science, Arizona State Univ, Tempe, Arizona.; 2) Communication Sciences and Disorders, Saint Louis University, Saint Louis, Missouri

---

*BCL11A* not only suppresses fetal hemoglobin in newborns but also influences neurogenesis via axon outgrowth. Deletions of *BCL11A* and its two closest neighbors have implicated the cerebellum, a brain region crucial for motor control and also spoken and written language. Several studies mention speech and language deficits where *BCL11A* was heterozygously deleted or altered, but only two studies provide a detailed description of these deficits, specifying childhood apraxia of speech (CAS), a severe speech disorder with cerebellar signs. Here, we describe two new cases with intragenic point mutations and one with a microdeletion.

Case 1 is a 9-year-old female with a p.Ala182Thr:c.544G>A missense mutation in exon 4 of *BCL11A*. This variant results in a non-conservative amino acid substitution, not described previously in actual cases but predicted to be highly deleterious. The patient has dysmorphic facial features, fine and gross motor delays, balance difficulties, generalized dyspraxia and dystonia, intellectual disability, ADHD, difficulties understanding and formulating sentences, and CAS. The CAS diagnosis was made at age 3 due to highly unintelligible speech and problems with oral motor tasks, chew-swallow sequences, and drooling. At age 9, she still has low intelligibility and difficulties with several consonants.

Case 2 is a 5-year-old female with a heterozygous 156 kb deletion harboring exons 1 and 2 of *BCL11A*. She was diagnosed with intellectual disability, ADHD, fine and gross motor discoordination, limited expressive language, and severe CAS. Her speech is highly unintelligible, and she is working on simple syllables and words consisting of one consonant and one vowel.

Case 3 is a 6-year-old male who carries a heterozygous *de novo* p.E611.X:c.1831 G>T variant. ClinVar lists this variant in a single case classified as pathogenic. The patient has severe fine and gross motor delays and while his expressive language is in the typical range, his speech, chewing, and swallowing are characterized by motor planning deficits.

In sum, all three cases show evidence of cerebellar dysfunction despite the different types of gene disruption. Furthermore, they share similarities including motor speech traits with a case with complete absence of one copy of only *BCL11A* described previously by us. These similarities are likely due to abnormal prenatal and early postnatal expression of *BCL11A* in brain regions crucial for motor, speech, and cognitive development.





## PgmNr 1333: Genetic mutation analysis of the COQ2 gene in Italian patients with MSA.

### Authors:

R. Procopio <sup>1,2</sup>; M. Gagliardi <sup>1</sup>; L. Brighina <sup>3</sup>; G. Nicoletti <sup>1</sup>; M. Morelli <sup>2</sup>; C. Ferrarese <sup>3</sup>; A. Quattrone <sup>1,4</sup>; G. Annesi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Molecular Bioimaging and Physiology, National Research Council, Section of Germaneto, Catanzaro, Italy; 2) Institute of Neurology, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy; 3) Department of Neurology, Milan Center for Neuroscience, San Gerardo Hospital, University of Milano-Bicocca, Monza, Italy; 4) Neuroscience Research Center, Department of Medical and Surgical Sciences, University Magna Graecia, Catanzaro, Italy

Multiple system atrophy (MSA) is an adult-onset neurodegenerative disease of unknown etiology characterized by dysautonomia accompanied by parkinsonism, cerebellar ataxia, and pyramidal signs. The neuropathologic hallmark of MSA is the presence of  $\alpha$ -synuclein-positive glial cytoplasmic inclusion. MSA is classified into two main subtypes: MSA-P, in which parkinsonism is the characteristic motor symptom; and MSA-C, in which the predominant motor presentation is cerebellar ataxia. Although considered a sporadic disorder, rare familial MSA cases have been described, supporting the presence of genetic factors in the etiology of this disease. Recently, a homozygous variant p.V393A and compound heterozygous variants in the *COQ2* gene have been described in two multiplex Japanese MSA-C families.

To extend the knowledges about the role of *COQ2* in MSA in Caucasian ethnicities, we performed genetic analysis of *COQ2* in a cohort of 100 MSA patients from Italy.

The cohort consisted of 58 patients with MSA-P and 42 with MSA-C, diagnosed by movement disorders specialists. All participants provided written informed consent and genomic DNA was collected from peripheral blood by standard protocols. All exons and intron-exon boundaries of *COQ2* were amplified and sequenced using an ABI 3500 Genetic Analyzer (Life Technologies, Carlsbad, CA, USA).

No pathogenic *COQ2* variants were detected. We identified only one missense variant (rs6818847) and two synonymous variants (rs6535454 and rs1129617), with high frequency in the population.

*COQ2* encodes para-hydroxybenzoate-polyprenyl transferase, essential for the biosynthesis of coenzyme Q10, involved in the mitochondrial respiratory chain and his deficiency results in several types of neurological disorders, ataxia and cerebellar atrophy similar to MSA.

The more common form of MSA in Asia is MSA-C (67-84%), while MSA-P is the more common manifestation in Europe and North America (58/82%). Several studies of East Asian cohorts confirmed a significant association of p.V393A with MSA-C, while studies in non-Asian ethnic cohorts were unable to replicate this finding.

We did not detect pathogenic variant in *COQ2* in our Italian MSA cohort. Therefore, our results confirmed that *COQ2* did not associate with the development of MSA in Caucasian population, supporting the hypothesis that *COQ2* tended to play a pathologic role population and subtype-specific in MSA.

## PgmNr 1334: *BPTF* mutation in siblings diagnosed with a radiation sensitivity disorder.

### Authors:

S. Teraoka <sup>1</sup>; X. Liu <sup>1</sup>; M. Gupta <sup>2</sup>; J. Wright <sup>1</sup>; A. Quinlan <sup>2</sup>; R. Gatti <sup>3</sup>; P. Concannon <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics Institute and Dept of Pathology, Immunology, and Molecular Genetics, University of Florida, Gainesville, FL; 2) Dept of Human Genetics and Dept of Biomedical Informatics, University of Utah, Salt Lake City, UT; 3) Dept of Pathology and Laboratory Medicine and Dept of Human Genetics, UCLA, Los Angeles, CA

---

A panel of 78 individuals who displayed radiation sensitivity, cancer susceptibility, immunodeficiency, neurological abnormalities and DNA double strand breaks were studied by whole exome (N=50) or whole genome (N=28) sequencing. These individuals had been previously referred for clinical testing for Ataxia-telangiectasia (A-T) or Nijmegen Breakage Syndrome (NBS) but no mutations in the causative genes *ATM* or *NBN* were identified. Whole exome sequencing uncovered a novel truncating mutation in bromodomain PHD finger transcription factor, *BPTF*, (c.2521C>T, p.Arg841\*), in a pair of siblings among 50 otherwise unrelated individuals. Although the initial expectation for the mode of inheritance for candidate genes was autosomal recessive based on A-T and NBS, autosomal dominant mutations in *BPTF* have been reported to cause a disorder with overlapping features, Neurodevelopmental Disorder with Dysmorphic Facies and Distal Limb Anomalies (NEDDFL). The siblings in our study had microcephaly, facial dysmorphology and short stature, consistent with the clinical features of NEDDFL. However, sensitivity to ionizing radiation has not been previously reported as a feature of NEDDFL nor has the *BPTF* gene been implicated in cellular responses to ionizing radiation or DNA damage. To explore the potential role of *BPTF*, cells in which siRNA had been used to knockdown *BPTF* were treated with a range of doses of ionizing radiation. A significant decrease in colony survival following treatment with ionizing radiation was observed in knockdown cells relative to controls. DNA from 28 additional radiation sensitive individuals was screened by whole genome sequencing. A second unrelated individual was identified with a 3.6Mb deletion including *BPTF*, providing a second example of a heterozygous loss of function mutation in *BPTF* associated with radiation sensitivity. Our results implicate microdeletions at 17q24.2, generally, and the *BPTF* gene, specifically, in cellular responses to ionizing radiation and highlight the utility of combining whole exome or genome sequencing with functional screens to facilitate novel gene discovery.

# PgmNr 1335: Comprehensive genetic analysis and genetic features of 1323 patients with Charcot-Marie-Tooth disease in Japan.

## Authors:

A. Yoshimura; A. Hashiguchi; M. Ando; Y. Higuchi; T. Nakamura; Y. Okamoto; H. Takashima

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Neurology and Geriatrics, Kagoshima University, Kagoshima, Kagoshima, Japan

---

## Background and Objective

Charcot-Marie-Tooth disease (CMT) is a group of Inherited peripheral neuropathies (IPNs) primarily affecting the peripheral motor and sensory nerves. Clear classification of CMT is difficult because of its clinical complexity and genetic diversity. In this study, we aim to describe the genetic features of CMT patients in Japan.

## Methods

In the period from May 2012 to December 2017, we collected 1323 cases with suspected CMT throughout Japan, whereas *PMP22* duplication/deletion were excluded in advance for demyelinating CMT cases. We performed targeted resequencing for encompassing 60 or 72 IPNs-related genes using Illumina MiSeq or Ion Proton, respectively. All suspected variants were validated using Sanger sequencing and interpreted according to the American College of Medical Genetics and Genomics standards and guidelines. We analyzed the gene-specific onset age and genetic features of these cases.

## Results

We identified pathogenic or likely pathogenic variants in 402 cases (30.4%). The most common causative genes were *GJB1* (n = 88, 21.9%), *MFN2* (n = 84, 20.9%), and *MPZ* (n = 66, 16.4%), followed by *HSPB1*, *PMP22* (point mutation), *GDAP1*, *NEFL*, *MME*. In demyelinating CMT, variants were detected in 44.6% cases, and the most common causative gene were *GJB1* and *MPZ*, and these were occupying for 66.4% in demyelinating type. Axonal CMT yielded a relatively lower detection rate (24.1%), and the most common causative gene was *MFN2* (33.8%). The most common disease onset period was first decade of life, and early-onset CMT cases were most likely to receive a molecular diagnosis.

## Conclusions

Our results updated the genetic profile within a large-scale of Japanese CMT cases. Subsequent analyses regarding onset age and geographical distribution advanced our understanding of CMT, which would be beneficial for clinicians.

## PgmNr 1336: Genotype-phenotype correlation of transthyretin-related familial amyloid polyneuropathy in Chinese population.

### Authors:

R. Zhang <sup>1,3</sup>; L. Liu <sup>1</sup>; B. Tang <sup>2</sup>; S. Zuchner <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurology, the Third Xiangya Hospital. Central South University, Changsha, Hunan, China; 2) National Clinical Research Center for Geriatric Disorders, Changsha, Hunan, China.; 3) Dr. John T. Macdonald Foundation Department of Human Genetics, John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL

---

**Objective:** Peripheral neuropathy can be caused by pathogenic TTR variants – a treatable rare neurodegenerative condition. To report the *TTR* pathogenic variants and clinical features in a cohort of 388 Chinese IPN families, and to summarize the genotypic and phenotypic spectrums in all reported Chinese TTR-FAP patients.

**Methods:** IPN families were recruited and analyzed with MLPA, gene panels and Sanger sequencing to detect *TTR* sequence variants. The published literature about TTR-FAP patients were reviewed and the phenotype-genotype spectrums were summarized.

**Results:** We detected five pathogenic *TTR* variants in five FAP families (p.Arg54Thr, p.Lys55Asn, p.Gly67Glu, homozygous and heterozygous p.Ala117Ser), including the first observation of coexistence of homozygous and heterozygous p.Ala117Ser alleles in one FAP family. The key clinical features were axonal motor and sensory neuropathies with prominent autonomic dysfunctions and organ involvements, and high inter-familial and intra-familial clinical heterogeneity. There are 88 reported TTR-FAP families (210 patients) of Chinese origin. Variants p.Ala117Ser, p.Gly103Arg, p.Val50Met and p.Val50Ala were relatively frequent. The age at onset ranged from 17 years to 80 years, with early-onset (age at onset < 50 years) accounting for 70.06% and late-onset (age at onset ≥ 50 years) accounting for 29.94%.

The autonomic symptoms alleviated after oral diflunisal treatment for 7 and 9 months in three patients, treated in the 2<sup>nd</sup> to 4<sup>th</sup> year after disease onset.

**Conclusions:** TTR-FAP accounted for 1.3% of all patients in our IPN cohort, which suggests a routine *TTR* screening is required in IPN patients. Due to the good efficacy we observed, as well as its economics and accessibility, diflunisal might be an ideal drug for Chinese FAP patients and large-scale clinic trial of oral diflunisal treatment should be carried out further.

# PgmNr 1337: NKX2-1 gene mutation: Unraveling brain-lung-thyroid syndrome in a Brazilian patient responsive to methylphenidate treatment.

**Authors:**

C. Lourenco; B.B. Dias; C.D. Gil; Z.A. Carneiro

[View Session](#) [Add to Schedule](#)

**Affiliation:** Faculdade Medicina - Centro Universitário Estácio de Ribeirão Preto, Sao Paulo, Brazil

---

## INTRODUCTION

Defects in *NKX2-1* on chromosome 14q13, which encodes thyroid transcription factor 1, produce a concurrent clinical manifestation of chorea, respiratory distress, and hypothyroidism known as “brain-lung-thyroid syndrome.”

## CASE REPORT

Male patient, 4 years, born to non consanguineous parents, after an uncomplicated pregnancy was referred for evaluation of hypotonia. He was under treatment of subclinical hypothyroidism with low doses of levothyroxin. His parents noted hypotonia and delayed motor development in first 2 years of life. Language and social development were normal. No report of respiratory problems at birth was reported. His investigation for causes of hypotonia (both central and neuromuscular hypotonia), including CK, lactate, acylcarnitine profile, organic acids chromatography and muscle biopsy turned out to be normal. Progressively, patient got better of his hypotonia, but developed irregular trunk and limb movements later recognized as chorea. The combination of extrapyramidal movement disorder and hypothyroidism pointed out to *NKX2-1* gene as cause of patient’s phenotype which was later confirmed by molecular analysis. He was started on methylphenidate showing dramatic improvement in his movement disorder.

## DISCUSSION

*NKX2-1*-related disorders range from benign hereditary chorea (BHC) to choreoathetosis, congenital hypothyroidism, and neonatal respiratory distress (also known as brain-lung-thyroid syndrome). Childhood-onset chorea, the hallmark of *NKX2-1*-related disorders, may or may not be associated with respiratory distress syndrome or congenital hypothyroidism. Chorea generally begins in early infancy or about age one year (most commonly) or in late childhood or adolescence, and progresses into the second decade after which it remains static or (rarely) remits.

## CONCLUSION

Neurological phenotype is prominent in this condition and that many patients with *NKX2-1* mutations may show hypotonia as main presentation and never fully develop the classic triad of brain-lung-thyroid syndrome

## PgmNr 1338: Brain region specific contribution of the ATXN1/CIC interaction to spinocerebellar ataxia type 1.

### Authors:

S.L. Coffin <sup>1,2</sup>; L. Nitschke <sup>1,2,3</sup>; H.T. Orr <sup>4</sup>; H.Y. Zoghbi <sup>1,2,3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, TX; 3) Program in Integrative Molecular and Biomedical Sciences, Baylor College of Medicine, Houston, TX; 4) University of Minnesota, Minneapolis, MN; 5) Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX

---

Spinocerebellar ataxia type 1 (SCA1) is a fatal adult onset neurodegenerative disease, characterized by loss of balance, slight cognitive impairment, breathing dysfunction and early lethality. This is due to selective neurodegeneration of cerebellar Purkinje cells (PCs) as well as hippocampal and brainstem neurons. The disease is caused by the expansion of CAG repeats, encoding the polyglutamine (polyQ) tract in ATAXIN-1 (ATXN1), a protein that is broadly expressed in the brain. Using biochemical and genetic studies, we discovered that the polyQ expansion stabilizes ATXN1 and mediates cerebellar degeneration through enhanced function with the transcriptional repressor Capicua (CIC). Disruption of the ATXN1/CIC interaction exclusively in cerebellar PCs rescues the motor deficits seen in a transgenic SCA1 animal model. However, the mechanisms of disease in the remainder of the cerebellum, hippocampus, and brainstem have not yet been elucidated. To investigate these unknowns, we modified the existing *Atn1*<sup>154Q/2Q</sup> knock-in SCA1 mouse model, that expresses one modified *Atn1* allele in which 154 CAG repeats replaced the endogenous 2 CAGs, and that recapitulates all known features of human SCA1. Using the CRISPR/Cas9 system we edited the 154Q allele to contain two amino acid changes (V591A and S602D), eliminating ATXN1<sup>154Q</sup>'s binding to CIC throughout the entire body. Using this mouse model, we will determine the contributions of the ATXN1/CIC complex to various SCA1 phenotypes by conducting behavioral and histopathological assays, as well as transcriptomic studies to assess levels of potential rescue. With these data, we will learn if the ATXN1-CIC complex is driving SCA1 disease pathology in regions other than Purkinje cells and can work towards discovering the mechanism of disease in these regions. This will be informative for other neurodegenerative disorders in which a broadly expressed protein is only selectively toxic, such as Alzheimer's and Parkinson's disease.

## PgmNr 1339: Somatic instability is a major determinant of clinical onset of Huntington disease.

### Authors:

G.E.B. Wright; J.A. Collins; C. Kay; N.S. Caron; H. Findlay Black; M.R. Hayden

[View Session](#) [Add to Schedule](#)

**Affiliation:** Centre for Molecular Medicine and Therapeutics, Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

---

Huntington disease (HD) is a neurodegenerative disorder that is caused by an expanded polyglutamine tract in the huntingtin (*HTT*) gene. The length of this repeat is inversely correlated with age of onset. Conventional analyses have assumed that polyglutamine length, not glutamine codon usage on the DNA level, is the most relevant variable to consider for predicting age of clinical onset in HD. This region is predominantly encoded by CAG trinucleotides, although the penultimate glutamine codon in this region is specified by CAA in the majority of HD patients [i.e. *HTT* reference: (CAG)<sub>n</sub>-CAA-CAG]. Recently, using patient samples collected over 25 years, we have shown that genetic variants that alter glutamine codon usage in this region significantly influence age of clinical onset in HD in an autosomal dominant manner. Two distinct modifiers of age of onset have been identified in this regard: (i) a variant that causes complete loss of interrupting (LOI) sequence [i.e. (CAG)<sub>n</sub>-CAG-CAG], that hastens age of onset by 13-29 years on average and (ii) a variant characterized by a gain of a CAA interruption [i.e. (CAG)<sub>n</sub>-(CAA-CAG)<sub>2</sub>], that delays age of onset by 4.2 years on average. Our previous work showed that the LOI variant is associated with increased CAG expansion ratios in whole blood, implicating increased somatic instability as a potential causal mechanism for hastening age of onset. In the current study, we assessed whether HD patients that are carriers of gain of interruption (CAA-CAG)<sub>2</sub> variant ( $n=10$ ) display differences in CAG expansion ratios in whole blood, compared to individuals with canonical repeat sequence ( $n=60$ ). These analyses revealed that somatic CAG expansion ratios were significantly reduced in (CAA-CAG)<sub>2</sub> carriers compared to individuals with only one CAA codon in this region ( $P=3.13 \times 10^{-3}$ ). For the first time, we show that reduced somatic instability may be the biological mechanism underlying the later age of onset observed in HD patients that are carriers of the (CAA-CAG)<sub>2</sub> variant. To further delineate these observations, we are currently obtaining access to brain tissue to assess the influence of the two variants on somatic instability in the tissues most relevant for the disease. The results of the current *cis*-modifier investigation help explain why certain individuals with identical polyglutamine lengths can present with HD at vastly different ages, implicating somatic instability as a chief determinant of age of onset.



## PgmNr 1340: Biallelic variants in *EXOSC5* cause phenotypic features found in exosomopathies.

### Authors:

A. Slavotinek<sup>1</sup>; D. Misceo<sup>2</sup>; L. Mathisen<sup>2</sup>; E. Frengen<sup>2</sup>; D. Schneidman<sup>3</sup>; Y. Si<sup>4</sup>; M. Fasken<sup>5</sup>; A. van Hoof<sup>6</sup>; A. Corbett<sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Pediatrics, Division Genetics, Univ California San Francisco, San Francisco, California.; 2) Dept. of Medical Genetics, Oslo University Hospital and University of Oslo, Norway; 3) Dept. of Biochemistry, The Hebrew University of Jerusalem, Israel; 4) GeneDx Inc., MD; 5) Dept. Biology, Emory University, GA; 6) Depts. of Microbiology and Molecular Genetics, University of Texas Health Science Center-Houston, TX

---

The RNA exosome complex is an essential ribonuclease complex involved in processing and surveillance of coding and non-coding RNAs and RNA degradation. The exosome comprises 10 core subunits in a barrel-like configuration with a cap of 3 proteins encoded by EXOSC1-3 that rests on a ring of 6 proteins encoded by EXOSC4-9. We present two patients with biallelic variants in EXOSC5, a novel human disease gene. The first child had motor delays and ataxia from 11 months and had cone rod dystrophy at 3 years. An MRI of the brain showed cerebellar hypoplasia and delayed myelination. At 11 years, she used a walker for ambulation. WES showed compound heterozygosity for a deletion involving exons 5-6 of EXOSC5 and a missense variant, p.Thr114Ile, affecting the RNase PH-like domain that was predicted to result in the loss of a hydrogen bond. The second patient, a male who died at 11 months, had hypotonia, esotropia and ophthalmoplegia, craniosynostosis, pectus carinatum and micrognathia. He required a tracheostomy. A muscle biopsy showed neurogenic and myopathic changes, with fiber size variation comprising hypertrophic and atrophic/hypotrophic fibers. A brain MRI showed reduced size of the brainstem and pons and slight enlargement of the 4th ventricle. He was homozygous for p.Leu206His in EXOSC5. We generated the missense variants in yeast - rrp46-Q86I, corresponding to EXOSC5-T114I, and rrp46-L191H, corresponding to EXOSC5-L206H. The rrp46-Q86I mutant shows growth similar to wildtype cells, whereas the rrp46-L191H mutant shows a moderate growth defect at 30°C that is exacerbated at 37°C compared to wildtype. Biallelic mutations have been described in four genes encoding exosome subunits. EXOSC3 mutations cause pontocerebellar hypoplasia (PCH1b) and spinal motor neuron disease, whereas EXOSC8 variants cause a different form of PCH. Missense variants in EXOSC2 cause intellectual disability, retinitis pigmentosa, hearing loss, short stature and facial anomalies, while variants in EXOSC9 cause cerebellar atrophy and an SMA-like motor neuronopathy. The two patients described here show clinical findings that overlap with other exosomopathies. Deleterious variants in the RNA exosome subunit genes could result in tissue-specific defects through a variety of mechanisms. Our findings suggest that normal function of the exosome subunit encoded by EXOSC5 is essential for CNS development and illustrate the mechanism of RNA dysregulation in motor and cerebellar degeneration.

## PgmNr 1341: The *C9orf72* repeat expansion, associated with ALS/FTD, is a rare folate-sensitive fragile site.

### Authors:

M. Mirceta<sup>1,2</sup>; M.H.M. Schmidt<sup>1,2</sup>; T. Prasolava<sup>1</sup>; S. Lanni<sup>1</sup>; P.A. Dion<sup>3</sup>; M. Mohiuddin<sup>1</sup>; M. Zhang<sup>4</sup>; M. Liang<sup>1,2</sup>; O. Beck<sup>5</sup>; S. Scheers<sup>5</sup>; L. Wiericx<sup>5</sup>; H. Raimer<sup>6</sup>; T. Abell<sup>7</sup>; E. Philips<sup>7</sup>; F. Riemsdagh<sup>8</sup>; E. Campos<sup>1,2</sup>; M. Wilson<sup>1,2</sup>; P. de Jong<sup>9</sup>; C.H. Freudenreich<sup>7</sup>; Y.H. Wang<sup>6</sup>; E. Rogaeva<sup>4</sup>; G.A. Rouleau<sup>3</sup>; R.F. Kooy<sup>5</sup>; C.E. Pearson<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics & Genome Biology, Hospital for Sick Children, Toronto, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 3) Department of Human Genetics, McGill University; Montreal Neurological Institute, McGill University; Department of Neurology & Neurosurgery, McGill University, Montreal, Canada; 4) Tanz Centre for Research of Neurodegenerative Diseases and Toronto Western Hospital, Division of Neurology, Department of Medicine, University of Toronto, Toronto, Canada; 5) Department of Medical Genetics, University of Antwerp, Universiteitsplein 1, Wilrijk, Belgium; 6) Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, Virginia 22908; 7) Department of Biology, Tufts University, 200 Boston Ave suite 4700, Medford, MA 02155, USA; 8) Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands.; 9) BACPAC Resource Center, Children's Hospital Oakland Research Institute, Oakland, CA

---

Expansion of the *C9orf72* repeat, (GGGGCC)<sub>n</sub>, is the cause of a plethora of disease symptoms, including being the most common cause of Amyotrophic Lateral Sclerosis, Frontotemporal Dementia, Parkinsonism, Huntington's disease phenocopies, multiple sclerosis, Alzheimer's, as well as clinical overlaps of these. How the *C9orf72* expansion mutation leads to the wide spectrum of clinical syndromes remains unknown. Here we revealed that the *C9orf72* expansion is a cytogenetic rare folate-sensitive chromosomal fragile site (RFSFS), similar to the ten RFSFS mapped to (CGG)<sub>n</sub> expansions, including FRAXA, the cause of fragile X syndrome (FXS), autism, fragile X associated tremor ataxia and premature ovarian failure. In total, there are >250 fragile sites, ~225 common fragile sites, and ~25 RFSFS, where only the latter have been linked with inherited disease. Our report represents the first mapped RFSFS linked to neurodegenerative disease, with all 10 of the previously mapped RFSFS associated with neurodevelopmental diseases. Like other RFSFS, *C9orf72* expansions present chromosome breaks, gaps, regions of elongated under-condensed chromatin, aberrant CpG methylation and unusual DNA structures. These *C9orf72* expansion cells are prone to micronuclei, double-strand breaks, sister-chromatid exchanges, and are a "hot spot" of genetic instability. We observed chromosomes that have lost the telomeric 9p region from *C9orf72* onwards with its lost acentric fragment. One of our patient cell lines has incurred a large 9q duplication in every cell. Thus, as with other RFSFS like FRAXA, the expanded *C9orf72* shows genomic instability. As with the mutant FRAXA allele, the mutant *C9orf72* allele assumes a striking aberrant chromatin packaging, as evidenced by extreme nuclease inaccessibility exclusively on the expanded allele. Nuclease resistance sensitivity spans ~3.3kb upstream and ~3.7kb downstream of the expanded repeat, and resistance is exacerbated by aberrant methylation. That the *C9orf72* expansion is a cytogenetic fragile site – the most profound macro-scale epigenetic mark, now broadens the repeat sequences associated with RFSFS, and sheds light on the varied genetic alterations, variable symptoms and

multiple modes of pathogenesises that can arise at the *C9orf72* locus, as observed with FXS.

## PgmNr 1342: Repeat configurations of CGG repeats in the NBPF19, a causative gene for neuronal intranuclear inclusion disease, and correlation of expanded CGG repeat sizes with age of onset.

### Authors:

S. Shibata <sup>1</sup>; J. Kikuchi <sup>1</sup>; H. Ishiura <sup>1</sup>; J. Yoshimura <sup>2</sup>; S. Morishita <sup>2</sup>; S. Tsuji <sup>3,4</sup>; T. Toda <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology, University of Tokyo Hospital, Tokyo, Japan; 2) Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; 3) Department of Molecular Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 4) Institute of Medical Genomics, International University of Health and Welfare, Chiba, Japan

---

#### [Background and Aim]

Neuronal intranuclear inclusion Disease (NIID) is an autosomal dominant disease characterized clinically by various combinations of cognitive decline, parkinsonism, cerebellar ataxia and peripheral neuropathy, and neuropathologically by eosinophilic hyaline intranuclear inclusions in the central and peripheral nervous systems and other tissues. We have recently discovered expanded CGG repeats in 5'UTR of the NBPF19 gene as the causative mutations in NIID. Given the findings that there are five paralogs including NBPF19 that have sequences with enormously high identities (>99%), we first determined the detailed repeat configurations of NBPF19 in the controls, and then investigated correlations of the sizes of expanded CGG repeats with the ages at onset of patients with NIID.

#### [Methods]

PCR primers were designed to specifically amplify genomic regions containing the CGG repeat of NBPF19. The PCR products obtained from 182 controls subjects were barcoded, and then subjected to circular consensus sequencing (CCS) employing a PacBio Sequel sequencer. To determine the size of expanded CGG repeats, genomic DNAs were extracted from 28 patients with NIID, and SacI-digested genomic DNAs were subjected to Southern blot hybridization analysis.

#### [Results]

The number of CGG repeats determined from CCS reads ranged 7–39 repeats in 182 control subjects, and there were considerable variations in the repeat configurations. In addition, three SNVs (rs1172135200, rs1258206224, and rs1436954367) were exclusively present in the allele with the repeat motif of (AGG)(CGG)9(AGG)3 in 14 control subjects. The repeat motif of (AGG)(CGG)*n*(AGG)2(CGG) was observed in the majority of the alleles and the CGG repeat lengths tended to be larger than those with the repeat motif of (AGG)(CGG)*n*(AGG)3. The onset age was available in 22 NIID patients, which ranged from 55 to 78 years (mean 64.5). The expanded CGG repeats of NBPF19 in the 28 patients with NIID ranged from 270 to 550 bp in length, based on the Southern blot analysis. The size of the expanded repeats, however, did not show significant correlation with the onset age ( $p = 0.656$ ).

#### [Conclusions]

There was no significant correlation between the size of expanded CGG repeats and the onset age in NIID patients, necessitating further investigations, including the size distribution of expanded CGG

repeats in brains, expression profiling of mutant alleles of NBPF19 in brains and involvement of methylation status of CGG repeats.

## PgmNr 1343: The phenotypic spectrum of the *LAS1L*-associated X-linked disorder in a patient cohort.

### Authors:

E.L. Macke<sup>1,2</sup>; M.A. Cousin<sup>1,2</sup>; B. Bostwick<sup>3</sup>; T. Barbaro-Dieber<sup>4</sup>; M. Field<sup>5</sup>; C. Gilissen<sup>6</sup>; A. Hackett<sup>5</sup>; M.G. Haug<sup>7</sup>; J. Juusola<sup>8</sup>; V. Kalscheuer<sup>9</sup>; P. Magoulas<sup>3</sup>; J. Rosenfeld<sup>3</sup>; X. Song<sup>4</sup>; R. Yusupov<sup>10</sup>; A. Ziegler<sup>11</sup>; P. Pichurin<sup>1,12</sup>; E.W. Klee<sup>1,2,12</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN, 55905 USA; 2) Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; 3) Dept. of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX, USA; 4) Cook Children's Medical Center, Fort Worth, TX, USA; 5) Genetics of Learning Disability Service, Hunter Genetics, Waratah, NSW, Australia; 6) Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, the Netherlands; 7) Department of Medical Genetics, St. Olav's Hospital, Trondheim, Norway; 8) JJ is an employee of GeneDx, Inc; 9) Research Group Development and Disease, Max Planck Institute for Molecular Genetics, Berlin, Germany; 10) Division of Medical Genetics, Joe DiMaggio Children's hospital, Hollywood, FL, USA; 11) Département de Biochimie et Génétique du CHU d'Angers, Angers, France; 12) Department of Clinical Genomics, Mayo Clinic, Rochester MN 55905, USA

---

*LAS1L* is required for synthesis of the 60S ribosomal subunit and maturation of 28S rRNA. Additionally, it plays a role in transcriptional regulation by affecting sumoylation status and transactivation potential of the zinc-finger transcription factor Zbp-89 (ZNF148).

Pathogenic variants in *LAS1L* have previously been associated with Wilson-Turner Syndrome (WTS, MIM#: 309585) in males. WTS is an X-linked neurological disorder typically characterized by intellectual disability (ID), dysmorphic facial features, hypogonadism, short stature, and truncal obesity. A patient with congenital lethal motor neuron disease has also been reported. We describe the phenotypic and genotypic spectrum of eleven male probands from unrelated families including three previously reported families and eight additional individuals, 7 of whom have novel *LAS1L* variants. Probands exhibit variable phenotypes overlapping with WTS, and also express additional novel phenotypes such as hypotonia and respiratory distress.

Probands in our cohort all have missense variations in *LAS1L*, with 10 unique variants identified in 11 unrelated families. Affected individuals from five families have variants clustering in exon 10, three of which impact the same amino acid residue. Remaining variants occur in exons 1, 2, 6 and 11. The variants in exons 1 and 2 fall within a conserved *LAS1L* domain. Only one variant is present in gnomAD (1/157718) in a female. Probands display variable phenotypes, including hypotonia (7/11), ID (5/11), genital abnormalities (5/11), respiratory distress (4/11), speech delay (4/11), obesity (4/11), motor delay (3/11), and abnormal gait (2/11) without a clear genotype-phenotype correlation. Respiratory distress and hypotonia have not previously been described in WTS, but some affected individuals present with both the WTS- as well as the motor neuron-related phenotypes. We hypothesize that the phenotypic spectrum for pathogenic variants within *LAS1L* is broader than previously described, and accounts for the motor delays, hypotonia, and respiratory arrest occurring

in our cohort.

Here we describe a cohort of patients with *LAS1L* variants expressing a complex and variable phenotype overlapping with WTS and including respiratory distress and hypotonia. Larger cohort and functional studies are needed to further elucidate the complex genotype-phenotype associations observed in these individuals and the mechanisms underlying the phenotypic variability.

## PgmNr 1344: Catatonia is a clinical feature of the natural progression of NLGN2-related autism spectrum disorder.

### Authors:

A. Shillington <sup>1</sup>; M. Vawter-Lee <sup>2</sup>; C. Erickson <sup>3</sup>; H. Saal <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Human Genetics, Cincinnati Children's Hospital, Cincinnati, Ohio; 2) Division of Neurology, Cincinnati Children's Hospital, Cincinnati, Ohio; 3) Division of Psychiatry, Cincinnati Children's Hospital, Cincinnati, Ohio

---

Neurexins and neuroligins are families of transmembrane synaptic cell adhesion molecule (SCAM) proteins located at the synaptic membrane which function to induce synapse formation and maturation. SHANK family proteins indirectly tether these SCAM proteins to glutamate receptor proteins creating a functional complex necessary for optimal synaptic growth and signalling. Variants in genes coding proteins in these pathways have been linked to Autism Spectrum Disorder (ASD). Catatonia is a comorbidity of ASD, and has been described in patients with variants in genes coding SCAM proteins and SHANK proteins, though to date has not been described in patients with Neuroligin-2 (*NLGN2*) mutations. We present a case of a patient with ASD, language impairment, behavior problems, hyperphagia and anxiety who presented as a teenager for evaluation of developmental regression, staring spells, weight loss and anhedonia; felt to represent catatonia. After treatment with benzodiazepines, the patient regained some developmental milestones. Genetic testing revealed a de novo missense mutation in *NLGN2*. We hypothesize that patients with pathogenic mutations in *NLGN2* diagnosed with ASD are at risk for catatonia as a comorbidity. The *NLGN2* protein, along with proteins in the SCAM and SHANK synaptic complex are present in GABA-ergic interneurons, and it is theorized that imbalances in the excitatory-inhibitory synaptic pathways cause derangement in optimal GABA and glutamate signalling at the synapse, provoking the onset of catatonia symptoms. Practitioners caring for patients with ASD second to pathogenic variants in genes coding SCAM and SHANK proteins, and more specifically in patients with disease-causing variants in *NLGN2*, should consider catatonia in the differential diagnosis for late onset developmental regression.



## PgmNr 1345: A rare cause of juvenile Parkinsonism.

**Authors:**

K. Hadzsiev; A. Till; J. Zima; R. Ripszam; B. Melegh

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medical Genetics, Medical School, University of Pécs, Pécs, Hungary

---

**Introduction:** Juvenile Parkinsonism (JP) is a rare condition with onset prior to age affects patients below 21, the genetic background of the disease is very heterogeneous. The most common genetical causes includes the alterations of PARK2, PINK1 and PARK7 genes. Symptomatology also can be very variable including classic Parkinsonian symptoms like rigidity, bradykinesia and postural instability as well as additional symptoms like depression, sleep disturbances, memory loss, constipation and urinary problems.

**Materials and Methods:** We report a 21-year-old girl with the onset of symptoms at 15 years. Initially symptoms included involuntary laugh and progressive mental decline, later associated with Parkinsonian symptoms.

**Results:** An apparently homozygous ATP13A2:c.2479G>A variant was identified in exon 22 of the *ATP13A2* gene by Whole Exome Sequencing. The variant is a novel missense variant previously not described in other patients. It is classified as a variant with unknown significance although in silico predictions suggest pathogenicity.

**Conclusions:** The above mentioned homozygous mutation confirms the diagnosis of Kufor-Rakeb syndrome, which is a rare autosomal recessive juvenile Parkinsonism syndrome. Through our case presentation we would like to shortly review the symptomatology, diagnostics and therapy of juvenile Parkinsonism.

## **PgmNr 1346: Markedly higher burden of schizophrenia associated copy number variants in very early onset psychosis compared to an adult schizophrenia population.**

### **Authors:**

C. Brownstein <sup>1,2,3</sup>; M. Hojlo <sup>1,3</sup>; E. Garvey <sup>1,3</sup>; E. D'Angelo <sup>2,3</sup>; C. Genetti <sup>1,2</sup>; M. Gorman <sup>4</sup>; D. Carroll <sup>3</sup>; A. Deo <sup>3</sup>; J. Madden <sup>1,2</sup>; A. Torres <sup>5</sup>; A. Beggs <sup>1,2</sup>; P. Agrawal <sup>1,2</sup>; D. Glahn <sup>2,3</sup>; J. Gonzalez-Heydrich <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Division of Genetics and Genomics, Boston Children's Hospital, Boston Massachusetts; 2) Department of Pediatrics, Harvard Medical School, Boston Massachusetts; 3) Developmental Neuropsychiatry Clinic, Department of Psychiatry, Boston Children's Hospital, Boston Massachusetts; 4) Department of Neurology, Boston Children's Hospital, Boston Massachusetts; 5) Department of Neurology, Boston University and Boston Medical Center, Boston Massachusetts

---

**Objectives:** Very early onset psychosis (VEOP) is defined as the onset of psychotic symptoms before 13 years of age and considered to be a more severe form of the more typical adolescent/adult onset psychosis. Therefore, we hypothesize that VEOP will have a higher rate of rare genetic mutations. The objective of this study is to compare the genetic burden of schizophrenia-related copy number variants (CNVs) in a clinically-referred VEOP cohort to that of schizophrenia variants that have been found in a Psychiatric Genomics Consortium adult onset psychosis cohort.

**Methods:** DNA and medical records were collected from VEOP probands and parents enrolled into the Manton Center for Orphan Disease Research protocol at Boston Children's Hospital (BCH).

Chromosomal Microarray (CMA) was performed on 106 probands and available parental/familial DNA samples. Probands were evaluated by clinicians in the Developmental Neuropsychiatry Program (DNP) at BCH and assigned diagnoses based on DSM-5 criteria. Diagnoses include unspecified schizophrenia spectrum disorder, schizophrenia, schizoaffective disorder, depression with psychotic features, bipolar disorder with psychotic features, and OCD with psychotic features.

**Results:** 9 probands from our cohort were found to have one of the 12 copy number variants (CNV) associated with schizophrenia with genome wide significance (Rees, 2016). These variants include 1q21.1 duplication (N=1), 15q11.2 deletion (N=1), 16p11.2 duplication (N=1), 16p13.11 duplication (N=4), and 22q11.2 deletion (N=2). Our rate of CNVs in the VEOP cohort was higher than the 0.6% rate in adult controls ( $p < 0.00001$ ) and the 1.9% rate in patients with schizophrenia ( $p < 0.0002$ ) that was reported in a recent case control study of over 20,000 patients with schizophrenia and over 20,000 controls (Bergen et al., 2018). Overall, 60% of patients in the VEOP cohort had a rare CNV.

**Conclusions:** This increased burden of rare CNVs associated with schizophrenia in our VEOP cohort supports the hypothesis that VEOP is associated with a higher rate of rare genetic mutations than adolescent/adult onset psychosis. The study of rare Mendelian forms of complex disease (such as VEOP) is an effective way to discover and confirm variants that impact more common forms of the disorder.

## PgmNr 1347: *GJB2* and *GJB6* mutations in non-syndromic childhood hearing impairment in Ghana.

### Authors:

S.M. Adadey<sup>1</sup>; N. Manyisa<sup>2</sup>; K. Mnika<sup>2</sup>; C. de Kock<sup>2</sup>; V. Nembaware<sup>2</sup>; O. Quaye<sup>1</sup>; G. Amedofu<sup>3</sup>; G.A. Awandare<sup>1</sup>; A. Wonkam<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) West African Centre for Cell Biology of Infectious Pathogens (WACCBIP), Biochemistry Cell and Molecular Biology, University of Ghana, Accra, Greater Accra, Ghana; 2) Division of Human Genetics, Faculty of Health Sciences - University of Cape Town, Cape Town, South Africa; 3) Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

---

Our study aimed to investigate *GJB2* (connexin 26) and *GJB6* (connexin 30) mutations associated with non-syndromic childhood hearing impairment (HI) as well as the environmental causes of HI in Ghana. Medical reports of 1104 students attending schools for the deaf were analyzed. Families segregating HI, as well as isolated cases of HI of putative genetic origin were recruited. DNA was extracted from peripheral blood followed by Sanger sequencing of the entire coding region of *GJB2*. Multiplex PCR and Sanger sequencing were used to analyze the prevalence of *GJB6*-D3S1830 deletion. Ninety-seven (97) families segregating HI were identified, with 235 affected individuals; and a total of 166 isolated cases of putative genetic causes, were sampled from 11 schools for the deaf in Ghana. The environmental factors, particularly meningitis, remain a major cause of HI impairment in Ghana. The male/female ratio was 1.49. Only 59.6% of the patients had their first comprehensive HI test between 6 to 11 years of age. Nearly all the participants had sensorineural HI (99.5%;  $n = 639$ ). The majority had pre-lingual HI (68.3%,  $n = 754$ ), of which 92.8% were congenital. Pedigree analysis suggested autosomal recessive inheritance in 96.9% of the familial cases. *GJB2*-R143W mutation, previously reported as founder a mutation in Ghana accounted for 25.9% (21/81) in the homozygous state in familial cases, and in 7.9% (11/140) of non-familial non-syndromic congenital HI cases, of putative genetic origin. In a control population without HI, we found a prevalent of *GJB2*-R143W carriers of 1.4% (2/145), in the heterozygous state. No *GJB6*-D3S1830 deletion was identified in any of the HI patients. *GJB2*-R143W mutation accounted for over a quarter of familial non-syndromic HI in Ghana and should be investigated in clinical practice. The large connexin 30 gene deletion (*GJB6*-D3S1830 deletion) may not account for of congenital non-syndromic HI in Ghana. We are currently employing Next Generation Sequencing approaches and functional genomics studies to identify the other genes involved in most families and isolated cases of HI in Ghana.

## **PgmNr 1348: Identification of *LOXHD1* as a deafness gene causing ski-slope hearing loss in Korean population: Genotyping of STR markers confirms strong founder effect of c.4212+1G>A in *LOXHD1*.**

### **Authors:**

B. Kim <sup>1</sup>; W. Jeon <sup>2</sup>; J. Han <sup>3</sup>; H. Park <sup>3</sup>; S. Lee <sup>3</sup>; D. Oh <sup>3</sup>; D. Kim <sup>2</sup>; B. Choi <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Otorhinolaryngology-Head and Neck Surgery, Chungnam National University, College of Medicine, Daejeon, Republic of Korea; 2) Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea; 3) Department of Otorhinolaryngology-Head and Neck Surgery, Seoul National University Bundang Hospital, Seongnam, Republic of Korea

---

*LOXHD1* (lipoxygenase homology domains 1) was reported to be a causative gene of autosomal-recessive non-syndromic hearing loss (ARNSHL, DFNB77) and it is also known as the causative gene for Fuchs Corneal Dystrophy (FCD) in the field of ophthalmology. DFNB77 is rarely documented in the literature and thus genotype-phenotype correlation is not well characterized. Further, it has never been reported in Korean deaf population. Here, we report the first *LOXHD1*-related deafness in Korean population supported by protein modeling study and the strong founder effect of a splicing region variant: c.4212+1G>A in this population.

Six pedigrees with congenital or pediatric-onset, bilateral sensorineural hearing loss (SNHL) were recruited for molecular genetic diagnosis. Audiogram showed a ski-slope pattern, which seems to progress to profound SNHL. Accordingly, two of six probands already received cochlear implant in their teens and twenties. Ophthalmologic examination did not reveal any problems. Molecular genetic diagnosis using whole-exome sequencing (WES) identified 2 missense variants, 1 frameshift variant and 2 splice region variants. Among them, c.4212+1G>A was the most commonly identified variant, involving 5 of 6 probands. 3D protein modeling study was done to prove the pathogenic potential of identified novel missense variants (c.2641G>C: p.Gly881Arg), which seems to exert pathogenicity by affecting the stability of beta sheet in the PLAT domain which comprises the entire protein encoded by *LOXHD1*. Genotyping of six c.4212+1G>A-linked Short Tandem Repeat (STR) markers revealed the strong founder effect of this variant in Korean deaf population.

Our work provides the first description of *LOXHD1*-related deafness in Korean population, as an important deafness gene causing progressive, ski-slope SNHL. Revelation of representative phenotype of DFNB77 and founder effect of c.4212+1G>A in Korean population will facilitate the genetic diagnosis in such cases, enabling more informative genetic counseling.

## PgmNr 1349: Overexpression of *CNRIP1*'s mRNA due to a genomic duplication associated with postlingual progressive hearing loss in the *DFNA58* family.

### Authors:

K. Lezirovitz <sup>1,2</sup>; G.A. Viera-Silva <sup>1,2</sup>; A.C. Batissoco <sup>1,2</sup>; D. Levy <sup>3</sup>; J.P. Kitajima <sup>4</sup>; A. Trouillet <sup>5</sup>; N. Zebajardi <sup>5</sup>; J. Sampaio-Silva <sup>1</sup>; V. Pedroso-Campos <sup>1</sup>; L.R. Nascimento <sup>1,2</sup>; C.Y. Sonoda <sup>1</sup>; L.G. Vasconcelos <sup>2</sup>; R. Beck <sup>2</sup>; S.S. Grasel <sup>2</sup>; D. Jagger <sup>6</sup>; N. Grillet <sup>5</sup>; R.F. Bento <sup>1,2</sup>; R.C. Mingroni-Netto <sup>7</sup>; J. Oiticica <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Otorhinolaryngology Lab-LIM32, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil; 2) Departamento de Otorrinolaringologia, Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo-SP, Brazil; 3) Laboratory of Genetics and Molecular Hematology LIM31, Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, São Paulo-SP, Brazil; 4) Mendelics Genomic Analysis, São Paulo-SP, Brazil; 5) Department of Otolaryngology - Head and Neck Surgery, Stanford University, Stanford, California, USA.; 6) UCL Ear Institute, University College London, London UK.; 7) Centro de Pesquisas sobre o Genoma Humano e Células-Tronco, Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

---

In 2009, we reported the mapping of the *DFNA58* locus in a large Brazilian pedigree in which autosomal dominant postlingual progressive hearing loss was segregating. Here we report the identification of an undescribed ~200Kb genomic duplication segregating with hearing loss in the *DFNA58* family: all 20 affected subjects inherited the duplication and all the 20 unaffected subjects did not, resulting in highly significant Lod score of 9.9. This novel duplication includes two entire genes (*PLEK* and *CNRIP1*) and, partially, a third gene, *PPP3R1*. One duplication breakpoint lies inside *PPP3R1*'s intron 1 and the other in intergenic region between *PLEK* and *FBXO48*, both breakpoints immersed in rich repetitive elements' regions. Analyses of the mRNA (RT-qPCR) of the three candidate genes in blood samples of affected individuals revealed significant overexpression (27.8X) of the *CNRIP1* gene in all 13 duplication carriers analyzed, but normal levels in the eight unaffected family members, when compared to six control individuals. The other two candidate genes did not show altered expression, when all affected subjects were analyzed. In situ hybridization as well as immunofluorescence in the neonatal murine cochlea showed that *Cnrip1* mRNA and the corresponding protein are largely expressed in the spiral ganglia and in the tympanic borders cells, but in 4-week-old animals, other cell types within and close to the organ of Corti also show *Cnrip1* protein expression. Clinical data such as ABR records, otoacoustic emissions and tonal audiometry from the affected duplication carriers, with varying degrees of hearing loss, suggested that the primary functional impairment resides in the cochlea. *CNRIP1* (CB1 cannabinoid receptor-interacting protein 1) is majorly known as the CB1 cannabinoid receptor interaction protein, with important functions in the nervous system such as a presynaptic modulator of neurotransmitter release. The human *CNRIP1* gene encodes two protein isoforms CRIP1a and CRIP1b. CRIP1a suppresses CB1-mediated tonic inhibition of voltage-gated Ca<sup>2+</sup> channels. *CNRIP1* has also been suggested to act as a tumor suppressor gene, since it was observed that demethylation of the *CNRIP1* promoter or overexpression

of *CNRIP1* can reduce the proliferative and migration abilities of colon cancer cells. In conclusion, our data support the causative association between the overexpression of *CNRIP1*, due to the 200Kb duplication, and autosomal dominant postlingual hearing loss.

## PgmNr 1350: Heimler syndrome: Hypomorphic mutations in the peroxisome-biogenesis genes *PEX1* and *PEX6* are in cause.

### Authors:

I. Ratbi <sup>1</sup>; K.D. Falkenberg <sup>2</sup>; M. Sommen <sup>3</sup>; N. Al-Sheqaih <sup>4</sup>; G. Vandeweyer <sup>3</sup>; J.E. Urquhart <sup>4</sup>; K.E. Chandler <sup>4</sup>; S.G. Williams <sup>4</sup>; N.A. Roberts <sup>4</sup>; M. El Alloussi <sup>5</sup>; G.C. Black <sup>4</sup>; S. Ferdinandusse <sup>2</sup>; A. Heimler <sup>7</sup>; A. Fryer <sup>8</sup>; A. Sefiani <sup>1</sup>; S.A. Lynch <sup>9</sup>; N. Cooper <sup>10</sup>; K.R. Ong <sup>10</sup>; C.E. Smih <sup>11</sup>; C.F. Inglehearn <sup>11</sup>; A. Berraho <sup>6</sup>; A.J. Mighell <sup>11</sup>; C. Elcock <sup>12</sup>; J.A. Poulter <sup>11</sup>; M. Tischkowitz <sup>13</sup>; S.J. Davies <sup>14</sup>; A.A. Mironov <sup>15</sup>; W.G. Newman <sup>4</sup>; H.R. Waterham <sup>2</sup>; G. Van Camp <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre de génomique humaine, Faculté de médecine et pharmacie, Université Mohamm, Rabat, Rabat, Morocco; 2) Laboratory Genetic Metabolic Diseases, Academic Medical Center, University of Amsterdam, Amsterdam 1105 AZ, the Netherlands; 3) Department of Medical Genetics, University of Antwerp, Antwerp 2610, Belgium; 4) Manchester Centre for Genomic Medicine, St. Mary's Hospital, Manchester Academic Health Sciences Centre, Manchester M13 9WL, UK; Manchester and Centre for Genomic Medicine, Institute of Human Development, University of Manchester, Manchester M13 9WL, UK; 5) Département de Pédiodontie-Prévention, Faculté de Médecine Dentaire, Université Mohammed V, BP 6212 Madinat Al Irfane, 10100 Rabat, Morocco and Service d'Odontologie, Hospital Militaire d'Instruction Mohamed V, Avenue des Far, Hay Riad, 10100 Rabat, Morocco; 6) Service d'Ophtalmologie B, Hospital des Spécialités, CHU Rabat, Morocco; 7) Division of Human Genetics, Schneider Children's Hospital of Long Island Jewish Medical Center, New Hyde Park, NY 11042, USA; 8) Department of Clinical Genetics, Liverpool Women's NHS Foundation Trust, Liverpool L8 7SS, UK; 9) National Centre for Medical Genetics, Our Lady's Children's Hospital, Crumlin, Dublin 12, Ireland; Department of Genetics, Children's University Hospital, Dublin 12, Ireland; 10) West Midlands Regional Genetics Service, Birmingham Women's Hospital NHS Trust, Birmingham B15 2TG, UK; 11) Leeds Institute of Biomedical and Clinical Sciences, St. James's University Hospital, University of Leeds, Leeds LS9 7TF, UK; 12) Academic Unit of Oral Health and Development, School of Clinical Dentistry, University of Sheffield, S10 2TA, UK; 13) Department of Medical Genetics and National Institute for Health Research Cambridge Biomedical Research Centre, University of Cambridge, Cambridge CB2 0QQ, UK and Department of Clinical Genetics, East Anglian Regional Genetics Service, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; 14) Institute of Medical Genetics, University Hospital of Wales, Cardiff CF14 4XW, UK; 15) Faculty of Life Sciences, University of Manchester, Manchester M13 9PL, UK

---

Heimler syndrome (HS) is a rare recessive disorder characterized by sensorineural hearing loss (SNHL), amelogenesis imperfecta, nail abnormalities, and occasional or late-onset retinal pigmentation. We firstly ascertained eight families affected by HS and, by using a whole-exome sequencing approach, identified biallelic mutations in *PEX1* or *PEX6* in six of them. Loss-of-function mutations in both genes are known causes of a spectrum of autosomal-recessive peroxisome-biogenesis disorders (PBDs), including Zellweger syndrome. PBDs are characterized by leukodystrophy, hypotonia, SNHL, retinopathy, and skeletal, craniofacial, and liver abnormalities. We demonstrate that each HS-affected family has at least one hypomorphic allele that results in

extremely mild peroxisomal dysfunction. Although individuals with HS share some subtle clinical features found in PBDs, the diagnosis was not suggested by routine blood and skin fibroblast analyses used to detect PBDs. Our findings define HS as a mild PBD, expanding the pleiotropy of mutations in *PEX1* and *PEX6*.

In a second time, we investigated another Moroccan family with Heimler syndrome with early onset, severe visual impairment and important phenotypic overlap with Usher syndrome. The patient carried a novel homozygous missense variant c.3140T > C (p.Leu1047Pro) of *PEX1* gene. As standard biochemical screening of blood for evidence of a peroxisomal disorder did not provide a diagnosis in the individuals with HS, patients with SNHL and retinal pigmentation should have mutation analysis of *PEX1* and *PEX6* genes.



## PgmNr 1351: Novel mutation in *CTNNA1* causes autosomal dominant pattern dystrophy.

### Authors:

R. Adele<sup>1,2</sup>; N. Roslin<sup>3</sup>; E. Tavares<sup>2</sup>; K. Tang<sup>2</sup>; B. Thiruvahindrapuram<sup>3</sup>; A. Paterson<sup>2</sup>; E. Heon<sup>1,2,4</sup>; A. Vincent<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Medical Science, University of Toronto, Toronto, Canada; 2) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 3) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada; 4) Ophthalmology and Vision Sciences, The Hospital for Sick Children, Toronto, Canada

---

**Introduction:** Pattern retinal dystrophy (PD) is a rare autosomal dominant disease that affects the central retinal region. In PD, pigment or lipofuscin accumulate within the retinal pigment epithelium leading to its erosion. This disease process can result in bilateral impairment of a person's central vision.

**Purpose:** To determine the genetic etiology, underlying PD in a three generation European family.

**Methods:** The proband, a 25 year old female presented to us with PD. Negative results were seen when gene panel testing was conducted. More, family members were examined by an ophthalmologist and recruited (n = 6); three other members, having a best corrected visual acuity of 20/50 or better were diagnosed with the disease. Whole genome analysis was done for three affected family members. Shared and rare variants were flagged.

**Results:** A heterozygous single nucleotide polymorphism, likely pathogenic was observed in a strong candidate gene, *CTNNA1* (NM\_001903: exon 6:c.835G>C:p.A279P). A segregation analysis was performed on all 7 family members; all affected family members were heterozygous for the novel presumed disease causing missense variant. The variant had strong pathogenicity scores (SIFT: 0.041 and polyphen:0.974) and was not seen in GnomAD. However, a family member thought to be unaffected clinically, also carried this variant.

**Conclusion and future directions:** This would be the second report linking *CTNNA1* with PD. We hypothesize that incomplete penetrance or age of onset provides an explanation of why the seemingly unaffected is a carrier of this disease causing variant. Allele specific expression will be addressed by RT-PCR and mass spectrometry on cDNA of patient derived fibroblast and sanger sequencing, to provide further support for the *CTNNA1* variant's pathogenic role. This variant discovery adds to the current theory that *CTNNA1* is involved in PD. Finding the pathogenic variant can help diagnose other patients with similar phenotypes. Furthermore, it will help expand the knowledge on the disease mechanism in PD that could help shed light to a cure in the future.

## PgmNr 1352: Exploring genetic conservation in the mechanisms of hearing and deafness between *Drosophila* and mammals.

### Authors:

D. Sutton <sup>1</sup>; J. Andrews <sup>1,2</sup>; S. Yamamoto <sup>1,2</sup>; A. Groves <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX

---

Over 150 genes that cause deafness when mutated have been identified through genetic screening in humans and mice, many of which have not been functionally characterized. Strikingly, over 90% of these genes identified have orthologs in the fruit fly, *Drosophila melanogaster*. Flies provide quick and efficient strategies to identify molecular mechanisms of disease which can be further explored in mammalian model systems. The auditory system of the fly, the Johnston's organ (JO), is located within the antenna and is comprised of ~200 stretch receptor units called scolopidia. These scolopidia contain mechanosensitive neurons that respond to gravity and sound from vibrations of the outermost antennal segment. Although the human and fly hearing organs are anatomically very different, several studies have identified functional conservation of some human deafness genes in the fruit fly.

We designed and performed a genetic screen to identify the degree of molecular and functional conservation between the fruit fly and mammals in the auditory system to provide valuable tools to help elucidate the function of deafness genes in mammals. We first used the DIOPT (DRSC Integrative Ortholog Prediction Tool) tool to generate a comprehensive list of fly genes that are orthologous to human and mouse deafness genes, and identified genetic tools to assess the expression pattern of these genes and proteins within JO. We then explored whether these genes were expressed in specific cell types within JO using T2A-GAL4 and UAS-GFP lines, and tested whether certain proteins are enriched in specific structures within the scolopidia in pupal JO using GFP or YFP protein trap lines. In addition to validating previous findings that *Myosin VIIa*, *Non-muscle myosin II*, and *Protocadherin-15* orthologs are expressed in JO, many other conserved deafness associated genes or proteins are also expressed. Interestingly, orthologs of Usher syndrome proteins Cadherin-23 and Whirlin were also expressed in JO and partially colocalize with the Myosin VIIa and Protocadherin-15 orthologs, suggesting a potential functional link. The finding that many Usher syndrome proteins are co-expressed within JO suggests a molecular conservation between the Usher complex in humans and flies. Our future work will determine the extent of this conservation and test whether loss of function of these genes in *Drosophila* is sufficient to cause deafness.

## PgmNr 1353: A novel mutation in *MERTK* for rod-cone dystrophy in a North Indian family.

### Authors:

S. Bhatia <sup>1</sup>; N. Kaur <sup>1</sup>; I.R. Singh <sup>2</sup>; V. Vanita <sup>2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Dr. Daljit Singh Eye Hospital, Amritsar, Punjab

---

**Objective:** To identify the underlying genetic defect of childhood-onset severe rod-cone dystrophy (RCD) in a consanguineous family from North India with autosomal recessive retinitis pigmentosa.

**Material and Methods:** A detailed family history, clinical data, and blood samples were collected from 11 members of the family, including 4 affected by an autosomal recessive rod-cone dystrophy (arRCD), and DNA was extracted. Whole-exome sequencing (WES) was performed on DNA samples of proband and her unaffected maternal uncle. Ion Reporter software (ver. 4.4) was used for the annotation of variants obtained by WES. The variants detected in proband were tested for validation in all other affected and unaffected family members using Sanger sequencing technique.

**Results:** We have identified a novel nonsense mutation—c.1647T>G (p.Tyr549Ter)—in the exon 11 of *MERTK* that co-segregated completely with the disease phenotype in all the 4 affected members and was not observed in the 7 unaffected members of the family. This mutation was also not detected in 120 ethnically matched controls (240 chromosomes), hence excluding it as a polymorphism.

**Conclusions:** *MERTK* has a role in retinal pigment epithelium as a regulator of rod outer segments' phagocytosis. Due to c.1647T > G substitution, the stop codon (p.Tyr549Ter) appears early in the transcript. It seems that either the altered transcript would degenerate through nonsense-mediated decay (NMD) or potentially form truncated protein lacking a functionally important domain (i.e., tyrosine kinase domain). These findings thus further expand the mutation spectrum in *MERTK* and substantiate its role in the pathogenesis of retinal dystrophy.

## PgmNr 1354: Expanding the clinical spectrum associated with *TUBB4B* mutations.

### Authors:

S. Méchaussier <sup>1</sup>; M. Descartes <sup>2</sup>; J. Kaplan <sup>1</sup>; J.-M. Rozet <sup>1</sup>; I. Perrault <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Genetics in Ophthalmology, Imagine and Paris Descartes University, Paris, France; 2) Department of Genetics, University of Alabama, Birmingham, USA

**Introduction.** Recently, we reported heterozygous mutations affecting Arg391 in the  $\beta$ -tubulin 4B isotype-encoding gene (*TUBB4B*) in four unrelated families with a distinctive neuro-sensorineural disorder consisting in early-onset and severe retinal dystrophy and sensorineural hearing loss. The mutations dampened the dynamics of microtubules (MT) lattice but ciliation and intraflagellar trafficking were unremarkable. Here, we report clinical expression and functional analysis of a novel *TUBB4B* mutation (p.Pro358Ser) identified *de novo* in a sporadic case displaying blindness, sensorineural hearing loss, short stature, chronic kidney disease and recurrent sinus and ear infections, a constellation of symptoms strongly reminiscent of ciliopathies.

**Material and methods.** Cilia from nasal brushing of the index case were analyzed by transmission electron microscopy (TEM). The dynamics of MT in normal conditions and under treatment with taxol and ciliation were analyzed by immunocytochemistry in cultured cells overexpressing FLAG-tagged wild-type or p.Pro358Ser *TUBB4B*.

**Results.** Cilia from nasal brushing of the index case revealed ultrastructural abnormalities with rare axonemes and disorganized basal bodies. Functional analysis in cultured cells overexpressing FLAG-tagged wild-type or mutant *TUBB4B* showed that the mutant *TUBB4B* have a drastic impact on microtubulation and ciliation as demonstrated by the absence of visible MT lattice and of primary cilia. Inspection of the  $\beta$ -tubulin atomic structure revealed that the Pro358 residue is located in a region that binds Taxol, an inhibitor of MT depolymerization. Interestingly, treatment of cells overexpressing the p.Pro358Ser *TUBB4B* with Taxol (25 nM) allowed the formation of a MT lattice and of primary cilia.

**Discussion and Conclusion.** The severe cilia alterations in nasal epithelial cells of the patient are consistent with a ciliopathy. Functional analysis showing that overexpression of the mutant *TUBB4B* recapitulate deciliation, strongly supports the role of the mutation in the patient phenotype.

Interestingly, the presence of MT and cilia upon treatment with Taxol, a well known chemotherapy drug, suggests that the p.Pro358Ser mutation increases the affinity for the drug, thereby reducing availability for the other  $\beta$ -tubulin isotypes that can assemble with  $\alpha$ - subunits to form the MT lattice. Whether Taxol should be used to recover ciliation in affected tissues of the patient certainly deserves consideration.

## PgmNr 1355: Inherited retinal dystrophies, a group of diseases with many faces: Mutation spectrum from a case series of Romanian patients with unexplained visual loss.

### Authors:

V. Plaiasu <sup>1</sup>; R. Viulet <sup>2</sup>; I. Stamatian <sup>2</sup>; I. Vatafu <sup>2</sup>; C. Nascutzy <sup>2</sup>; A. Ciubotaru <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Regional Center of Medical Genetics, INSMC Alessandrescu-Rusescu, Bucharest, Romania; 2) Ophthalmology Department, INSMC Alessandrescu-Rusescu, Bucharest, Romania; 3) Infosan Eye Hospital, Bucharest, Romania

---

**Background:** Retinal dystrophies (RDs) are degenerative diseases of the retina which can lead to vision loss. Common presentations among these disorders include night or colour blindness, tunnel vision and subsequent progression to complete blindness. Cases may be progressive or stationary, syndromic or non-syndromic and may be familial with autosomal recessive, autosomal dominant or X-linked modes of inheritance described, with sporadic cases also observed. More than 260 genes have been identified for the isolated forms and for syndromes. **Methods:** We present the results of genetic testing performed on rare patients experiencing visual impairment referred to Genetics Department in the last two years. Whole Exome Sequencing was performed on each patient in an international laboratory. Only for one patient Whole Genome Sequencing was applied. Cases aged between 8 months and 54 years were unrelated and their mean age of symptom onset was with range from birth to age of 30 years, with heterogeneous clinical presentation over time. Ophthalmological evaluation showed a large range of visual activity, with color vision defects and wide range of electroretinographic abnormalities. **Results:** Disease-causing mutations were identified in all analyzed patients, distributed across 10 different genes: *VCAN*, *C2orf71*, *CACNA1F*, *ABCA4*, *NDP*, *PDE6C*, *RPGRIP1*, *PDE6B*, *TUBB4B*, *CEP250*. Some sequence changes were predicted to be possibly or probably damaging and novel variants were described at the time of diagnostic testing. Few patients were identified to have many variants linked to retinal genes, interpreted as modifiers which could alter phenotype by acting as risk factors or protection factors. In our cohort, the most prevalent dystrophies were *ABCA4* gene-related disorders (5 patients). Given the large number of variants reported in *ABCA4*, most of them being polymorphisms, the identification of true disease-causing mutations was challenging in the frame of low local genetic resources which limited carrier genetic testing of the families. **Conclusions:** The clinical description remains an important first step in focusing genetic testing. A medical geneticist can help to choose the best genetic test, based on current test methodology. Eyecare professionals have an active role in caring for patients with retinal dystrophies. Next generation sequencing technologies are among the most promising approaches to identify mutations in RDs and to decide appropriate multidisciplinary management.

## PgmNr 1356: Molecular genetic analysis of rare non-syndromic prelingual hearing disorders in a Romanian patient population.

### Authors:

R. Birkenhager <sup>1</sup>; M. Trabandt <sup>1</sup>; C. Buesching <sup>1</sup>; L. Radulescu <sup>2</sup>; S. Arndt <sup>1</sup>; A. Aschendorff <sup>1</sup>; R. Laszig <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University Medical Center Freiburg, Otorhinolaryngology H&N Surg, Univ Freiburg, Freiburg, Germany; 2) University of Medicine and Pharmacy "Grigore T. Popa", Iasi, Romania

---

**Introduction:** About 1-3 / 1000 newborns are affected by a profound hearing disorder at birth or in the first two years of life. About 60% of these cases are due to genetic causes. Inherited hearing disorders are divided into syndromic or non-syndromic, non-syndromic hearing disorders (NSHL) are isolated, while syndromic hearing disorders (SHL) are associated with additional organic disorders. Approximately 70% of cases of inherited hearing disorders are non-syndromic and predominantly due to sensorineural causes. Of these, approximately 80% of cases are autosomal recessive (DFNB) and 18% are autosomal dominant (DFNA), about 2% are X-chromosome (DFNX) or mitochondrial (MT) associated. To date, 185 gene loci and 128 genes have been identified for this type of hearing impairment. Genetic changes in the *DFNB1* gene locus, in which the genes *GJB2* gene (connexin-26) and *GJB6* (connexin-30) are localized represent the main cause (~45%) of prelingual non-syndromic hearing disorders. The aim of the project is to demonstrate, in a Romanian patient group, which rare genetic hearing disorders occur.

**Methods:** So far 120 patients have been included in our studies that have been diagnosed with severe to profound non-syndromic hearing impairment in their first two years of life and which have been shown to have no alteration in the *DFNB1* gene locus, in the genes *GJB2* and *GJB6*, respectively. Initially, targeted genes were analyzed, which in some few rare cases led to hearing impairment in various European populations. The detection of genetic alterations was carried out by bi-directional sequencing of the coding exons, as well as the intron transitions.

**Results:** First of all, the genes *GRXCR1* and *ESRRB* were analyzed in this patient group followed by genes *TMIE*, *GIPC* and *LHFPL5*. By DNA sequencing, 6 novel mutations, 12 unknown polymorphisms and 14 known alterations that are already cataloged in the databases of international sequencing projects have been detected so far.

**Conclusions:** In the investigated patients mutations and previously unknown polymorphisms were occasionally identified in the genes *GRXCR1*, *ESRRB* as well as *TMIE*, *GIPC* and *LHFPL5*, however, an accumulation of changes is not detectable. Therefore, further investigations are required for a better understanding of the etiology of prelingual hearing disorders.

## PgmNr 1357: High-throughput functional genomics of hearing and vestibular disorders in Zebrafish.

**Authors:**

P. Barbalho; C. Petree; S. VijayKumar; S. Lin; P. Varshney; G. Varshney

[View Session](#) [Add to Schedule](#)

**Affiliation:** Oklahoma Medical Research Foundation, Oklahoma City, OK.

---

Approximately 360 million people worldwide (5% of the world's population) are affected by hearing loss. Indeed, one in every 500-1000 newborns is affected by sensorineural hearing loss, making it one of the most common birth defects. Hearing loss is associated with a substantially reduced quality of life and overall health. Low-cost genomic sequencing technologies continually identify new mutations associated with human hearing loss, but their functional validation is unacceptably slow. There is a need to fill this knowledge gap, which is required to develop new treatments that arrest or reverse hearing loss and improve patient outcomes. We found that 94% of human hearing loss genes have an orthologue in zebrafish, suggesting high functional conservation. Zebrafish are an ideal model organism to study hearing loss, given their external embryonic development, transparent body, accessible inner ear and the presence of lateral line neuromasts, which are functionally analogous to mechanoreceptors of the mammalian inner ear. In addition, zebrafish have been shown to effectively recapitulate disease phenotypes. Numerous publications have demonstrated the efficacy of gene targeting in zebrafish using CRISPR/Cas9 including a variety of tools and methods for guide RNA synthesis and mutant identification. While all the published techniques work, not all approaches are readily scalable to increase throughput. We recently described a CRISPR/Cas9 based high-throughput mutagenesis and phenotyping pipeline in zebrafish.

Using this pipeline, we generated loss of function alleles in 80 hearing loss genes selected from the [hereditaryhearingloss.org](http://hereditaryhearingloss.org) database. We will present the mutagenesis pipeline, and phenotyping data from a selected candidate genes showing various hearing and vestibular defects.

## PgmNr 1358: Pathogenic variants with significantly increased odds ratio for Mendelian traits are rare in real world data.

### Authors:

Xu; F. Bu; J. Cheng; M. Zhong; Y. Lu; Q. Zhang; Y. Zeng; M. He; X. Wang; H. Yuan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medical Genetics Center, Southwest Hospital, Chongqing, China

---

**Introduction:** In the practice of clinical genetic testing, the ACMG/AMP guideline was widely accepted guidance for the interpretation of sequence variants. Among listed evidence in the guideline, significantly increased prevalence of a variant in affected individuals than controls (PS4) is considered as a strong indicator of pathogenicity. However, given the ultra-low frequency of most pathogenic variants for Mendelian traits, e.g. hereditary hearing loss, the availability of PS4 is expectedly limited.

**Methods:** 10889 hearing loss cases and 3204 normal controls from 31 provinces of China were recruited by the China Deafness Genetics Consortium (CDGC). By screening 156 hearing loss causing genes, 1470 reported and 624 novel pathogenic variants were identified from 90539 rare variants that alter amino acid sequence or splicing sites. Fisher's exact test and logistic regression with population stratification adjustment were used to compare the variant frequencies in cases and controls. Relative risk (RR) and odds ratios (OR) were calculated and evaluated for the precision and recall rate.

**Results:** Fisher's exact test and logistic regression by PLINK respectively indicate only 28 and 10 pathogenic variants were enriched in the cases ( $P < 0.05$ ). Controversially, results also indicate 8138 and 6082 non-pathogenic variants false positively associated with the cases by three tests. ROC analysis for the P values of two tests, RR and OR suggest that OR had the best performance in differentiating pathogenic and benign variants with an AUC of 0.808. By grouping the variants based on MAF, results suggest that, of the PS4 labeled pathogenic variants, 1% had a relatively common  $MAF > 1.4\%$ .

**Summary:** Although PS4 strongly supports variant pathogenicity, it's only prevalent in ultra-small portion of the pathogenic variants for hearing loss. False positive association results also largely interfered the usage of PS4. Our real-world data suggest that the PS4 should be considered in the context of other genetic information instead of as a standalone term.



## PgmNr 1359: A study of hereditary non-syndromic sensorineural hearing loss in a Nigerian population.

**Authors:**

M. Ukaegbu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Biology, Microbiology and Biotechnology Department, Alex Ekwueme Federal University, Ndufu-Alike, Ebonyi State, Nigeria., Ikwo, Nigeria.

---

This study examined the mode of inheritance of non-syndromic SNHL using three generations pedigree. Data on the age, sex, region of the country where the patient comes from, age at onset of hearing loss, number of affected ears and family history of the patients were obtained through a structured questionnaire. Individuals whose hearing loss was as a result of environmental influence were excluded from the study. The frequency and severity of hearing loss was obtained from the pure tone audiometry. Syndromic sensorineural hearing loss patients were screened out during physical examination by Medical Doctors. A total of 148 patients (98.7 %) had autosomal recessive mode of inheritance while only 2 patients (1.3 %) had autosomal dominant mode of inheritance. None had either maternal or sex-linked mode of inheritance. Among the 102 patients that did audiological evaluations, 43 patients (42.1 %) had moderately severe SNHL, 28 (27.4 %) had severe SNHL, 24 (23.5 %) had profound SNHL, 7 (6.8 %) had moderate SNHL while none had mild SNHL. Fifty four patients (53 %) had high frequency SNHL, 30 (29.4 %) had middle frequency SNHL while 18 (17.6 %) had low frequency SNHL. The results of the study demonstrate that autosomal recessive mode of inheritance is the commonest mode of inheritance of non syndromic SNHL in the studied population. A novel mutation, Leu 56 His which caused non-syndromic SNHL, was discovered from this study.

## PgmNr 1360: *RIMS2*, a novel gene responsible for syndromic congenital cone-rod synaptic disorder associated with neurodevelopmental and pancreatic involvement.

### Authors:

I. Perrault<sup>1</sup>; S. Méchaussier<sup>1</sup>; B. Almoallem<sup>2</sup>; K. Van Schil<sup>2</sup>; J. Van Dorpe<sup>3</sup>; M. Polack<sup>4</sup>; N. Boddaert<sup>5</sup>; N. Bahi-Buisson<sup>6</sup>; A. Mouallem<sup>7</sup>; C. Zeitze<sup>8</sup>; I. Audo<sup>8</sup>; J. Kaplan<sup>1</sup>; J.M. Rozet<sup>1</sup>; E. De Baere<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics in Ophthalmology, INSERM UMR1163 - IMAGINE, Paris, Paris, France; 2) Center for Medical Genetics, Ghent University and Ghent University Hospital, Ghent, Belgium; 3) Department of Pathology, Ghent University Hospital, Ghent, Belgium; 4) Endocrinology, Gynecology and Pediatric Diabetology Department, University Hospital Necker-Enfants Malades, APHP, 75015 Paris, France.; 5) Department of Pediatric Radiology, Hôpital Necker-Enfants Malades, APHP, Paris, Descartes University, 75015 Paris, France.; 6) Pediatric Neurology Department, University Hospital Necker-Enfants Malades, APHP, 75015 Paris, France.; 7) Ophthalmology Department, University Hospital Henri Mondor, APHP, 94000 Creteil, France.; 8) Centre Hospitalier National d'Ophtalmologie des Quinze-Vingts, 75012 Paris, France.

---

**Introduction:** Congenital cone-rod synaptic disorder (CRSD) is a non-progressive disorder of the congenital stationary night blindness spectrum. CRSD manifests at birth by congenital nystagmus, photophobia and normal or near-normal fundus appearance. This presentation is reminiscent of Leber congenital amaurosis (LCA), the earliest and most severe retinal dystrophy. Electroretinography is pivotal to make the differential diagnosis, characterized by near-normal to subnormal rod function, and delayed and/or decreased to non-recordable cone responses in CRSD versus no rod or cone response in LCA. This examination may be challenging however in young infants. To date, mutations in *CABP4* and *CACNA1F* encoding proteins of the photoreceptor synapse have been involved in CRSD. Here, we report the identification of a third gene further involving the photoreceptor synapse in three unrelated families, two of which were initially diagnosed with LCA. **Material and methods:** WES has been performed in trio from 1 sporadic case born to non-consanguineous parents and in index cases from 2 multiplex and consanguineous families. Patients (n = 6, 5 ≤ age ≤ 32 years) were re-examined for ocular, metabolic and neurologic manifestations. **Results:** Variant analysis identified homozygosity for nonsense variants and compound heterozygosity for a nonsense and a consensus splice-site variant in *RIMS2* (p.Trp1042\*, p.Arg1170\*, p.Arg962\*/c.4363+1G>A p.?). *RIMS2* encodes the RAB3A-interacting molecule 2 (RIM) also known as RIM2, a protein regulating synaptic membrane exocytosis in photoreceptors and brain, and insulin secretion by pancreatic islets. Ophthalmological re-examination showed ERG traces consistent with the diagnosis of congenital cone-rod synaptic disorder. Neurological examination revealed autism spectrum disorder or intellectual disability in 4/5 individuals and metabolic work-up (3/5 individuals) showed abnormal glucose homeostasis in the eldest patient (32 years). *RIMS2* immunostaining was shown in the outer plexiform layer, Purkinje cells and pancreatic islets. **Discussion:** The disease presentation associated with biallelic *RIMS2* loss-of-function variants is consistent with its expression in human retina, brain and pancreas where it plays an important role in the release of synaptic and secretory vesicles. CRSD is reminiscent of

*CABP4*- and *CACNA1F*-disease that is non-syndromic however. Here, we describe the first syndromic CRSD with neurodevelopmental or pancreatic involvement.

## PgmNr 1361: Notch ligand *JAG1* is a novel candidate gene for familial exudative vitreoretinopathy.

### Authors:

Zhu <sup>1</sup>; L. Zhang <sup>2</sup>; Z. Yang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sichuan Provincial People's Hospital, Chengdu, China; 2) University of Electronic Science and Technology of China, Chengdu, China

---

### Purpose

Familial exudative vitreoretinopathy (FEVR) is a blindness causing retinal vascular disease characterized by incomplete vascularization of the peripheral retina and by the absence or abnormality of the second/tertiary capillary layers in the deep retina. Mutations in known FEVR disease genes can only explain about 50% of FEVR-affected cases. We aim to identify additional disease genes in patients with FEVR.

### Methods

We applied exome sequencing analysis in a cohort of 49 FEVR families without pathogenic variants in known FEVR genes. Functions of the affected proteins were evaluated by reporter assay. Knockout mouse models were generated by endothelial specific Cre line.

### Results

Three novel rare heterozygous variants in Notch ligand *JAG1* were identified in FEVR families: c.413C>T p. (A138V), c.1415G>A p. (R472H) and c.2884A>G p. (T962A) and verified by Sanger sequencing analysis. Notch reporter assay revealed that mutant *JAG1* proteins *JAG1*-A138V and *JAG1*-T962A almost lost all of their activities, and *JAG1*-R472H lost approximately 50% of its activity. Deletion of *Jag1* in mouse endothelial cells resulted in reduced tip cells at the angiogenic front and retarded vessel growth, reproducing FEVR-like phenotypes.

### Conclusions

Our data suggest that *JAG1* is a novel candidate gene for FEVR and pinpoints a potential target for therapeutical intervention.

## PgmNr 1362: Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations.

### Authors:

K.M. Bujakowska; E. Zampaglione; B. Kinde; E.M. Place; M. Maher; D. Navarro-Gomez; F. Jamshidi; A. Mazzone; C. Finn; J. Comander; E.A. Pierce

[View Session](#) [Add to Schedule](#)

**Affiliation:** Ocular Genomics Institute, Massachusetts Eye & Ear, Harvard Medical School, Boston, Massachusetts.

---

**Purpose:** Over two million people worldwide are affected by inherited retinal degenerations (IRDs). However, despite substantial progress in sequencing, current strategies can genetically solve only about 55-60% of IRD cases. This can partially be attributed to elusive mutations in the known IRD genes, including copy number variations (CNVs), which we believe are a major contributor of the unsolved IRD cases.

**Methods:** Five hundred IRD patients were analyzed with targeted NGS to detect mutations in the known IRD genes. The NGS data was analyzed by two bioinformatic approaches (ExomeDepth and gCNV) to detect likely causal CNVs. The CNV detection rate by the NGS methods was compared to the SNP-Array in 144 patients using PennCNV software. Likely causal CNV predictions were validated by quantitative PCR on genomic DNA.

**Results:** Single nucleotide variants (SNVs) and small indels likely leading to disease were found in 55.6% of subjects, with mutations in *USH2A*(11.6%), *RPGR*(4%) and *EYS*(4%) as the most common causes of disease. Deletions and duplication likely leading to disease were found in additional 8.8% of patients and 1.4% of patients carried the *MAK-Alu* insertion. Of the three CNV detection methods gCNV showed the highest performance (positive predictive value of 85%). One third of the remaining 171 unsolved subjects had a single likely pathogenic mutation in a recessive IRD gene and there is a high chance they carry a second elusive mutation in the same gene.

**Conclusions:** CNV detection using NGS-based software is a reliable method, which increases the genetic diagnostic rate of IRDs from 55.6% to 64.4%, with highest performance seen for gCNV. Since each of the prediction tools gave false positive results, experimental validation of CNVs is an important step in determining the genetic diagnosis. The results presented in this study suggest that additional analyses of elusive variants in known IRD genes (e.g. *Alu* insertions, deep intronic variants, other structural variations) will have a major impact on improving the diagnostic rate of IRD patients.

## PgmNr 1363: Alterations in *GRHL2-OVOL2-ZEB1* axis and aberrant activation of Wnt signaling lead to altered gene transcription in posterior polymorphous corneal dystrophy.

### Authors:

D. Chung; W. Zhang; K. Jatavallabhula; A. Barrington; J. Jung; A. Aldave

[View Session](#) [Add to Schedule](#)

**Affiliation:** Ophthalmology, Stein Eye Institute, UCLA, Los Angeles, California

---

Mutations associated with posterior polymorphous corneal dystrophy (PPCD) have been identified in three genes: *ZEB1* (zinc-finger E-box binding homeobox 1) associated with sub-type PPCD3; *OVOL2* (ovol-like zinc finger 2) associated with sub-type PPCD1; and *GRHL2* (grainyhead like transcription factor 2) associated with sub-type PPCD4. Each of these genes encodes a transcription factor that regulates cell-state transitions. While the discovery of these PPCD-associated genes has greatly expanded our knowledge of the genetic basis of PPCD, the molecular mechanisms via which mutations in these genes lead to indistinguishable disease phenotypes have yet to be elucidated. To characterize the gene expression profiles of the genetic sub-types of PPCD, RNA-seq was performed on corneal endothelium derived from an individual with PPCD1 who harbors a c.-307T>C *OVOL2* promoter mutation; an individual with PPCD3 associated with a *ZEB1* truncating mutation (c.1381delinsGACGAT); and an individual with genetically unresolved PPCD. Transcriptomic analysis revealed: *OVOL2* expression increased in PPCD1 (259 fold), unchanged in PPCD3 and slightly increased in genetically unresolved PPCD (from 0 TPM to 0.86 TPM, undefined fold change); *ZEB1* expression decreased in PPCD1 (-5.9 fold), PPCD3 (-3.95 fold) and genetically unresolved PPCD (-3.96 fold); and *GRHL2* expression increased in PPCD1 (333.5 fold), slightly increased (from 0 TPM to 0.67 TPM, undefined fold change) in PPCD3 and increased in genetically unresolved PPCD (1853 fold). As *ZEB1* can act as an activator or repressor of downstream target gene expression depending on Wnt signaling pathway activation or deactivation, we also sought to determine whether or not Wnt signaling is active in PPCD by performing immunohistochemistry in corneal tissue sections derived from an individual affected with PPCD3 and from an individual with genetically unresolved PPCD. Immunohistochemistry results demonstrated corneal endothelial nuclear accumulation of S552 phospho-β-catenin and cytosolic localization of S33/37/T42 non-phosphorylated β-catenin in PPCD, indicating aberrant activation of Wnt signaling, which was not observed in control corneal endothelium. These findings suggest that alterations in the *ZEB1-OVOL2-GRHL2* axis (caused by PPCD-associated mutations) lead to changes in corneal endothelial cell state and molecular pathways, including the aberrant activation of the Wnt signaling pathway.

## PgmNr 1364: Discovering novel rhesus macaque models of human disease through a combination of forward and reverse genetics: Examples from retinal disorders.

### Authors:

J. Wang <sup>1</sup>; M. Raveendran <sup>1</sup>; R.A. Harris <sup>1</sup>; A. Moshiri <sup>2</sup>; T. Stout <sup>3</sup>; S.M. Thomasy <sup>4</sup>; J. Rogers <sup>1</sup>; R. Chen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 2) Department of Ophthalmology and Vision Science, School of Medicine, UC Davis, Sacramento, California, USA; 3) Department of Ophthalmology, Cullen Eye Institute, Baylor College of Medicine, Houston, Texas, USA; 4) Department of Surgical and Radiological Sciences, School of Veterinary Medicine, University of California-Davis, Davis, California, USA

---

One of the major impediments to the development of efficacious therapies for human genetic disorders is the lack of animal models that closely resemble the human conditions. A prominent example is the difficulty of modeling human macular disease in animals such as rodents, as they lack maculae and thus do not provide translational validity. By contrast, the retinal architecture of nonhuman primates (NHP), particularly their cone-dominated central retina and macula, closely mimics the human retina. Therefore, well-defined NHP models that closely resemble the human condition are essential for developing and testing novel therapies. We have designed a combined forward and reverse genetics approach to take full advantage of the naturally occurring functional genetic variation in rhesus macaques (*Macaca mulatta*) to develop NHP primate models of inherited human retinal diseases (IRD). To date, we have sequenced 1485 rhesus macaques across multiple National Primate Research Centers with a custom enrichment panel targeting protein coding regions of about 264 known IRD associated genes. A large number of genetic variants have been identified in the rhesus population with over 4000 rhesus SNPs also present in human population. The majority of the overlapped SNPs belong to CG transitions, and the allele frequencies of these overlapped variants are largely discordant between rhesus and human. More importantly, about 20 LoF mutations and 14 reported human likely pathogenic mutations were identified in IRD genes in the rhesus monkeys. For example, we found predicted damaging mutations in *NPHP4*, *OPA1* and other genes known to cause IRD. The rhesus carriers of these mutations are potentially useful for establishing new human disease models. Phenotypic studies of the identified rhesus carriers are in progress. In parallel, through a forward genetics screen among individual rhesus macaques which show visual system defects, we identified a missense mutation in *PDE6C* that causes achromatopsia in macaque homozygotes (Moshiri et al. JCI 129: 863, 2019). Targeted sequencing of individuals with other phenotypes is also underway. Overall, this broad approach consisting of large-scale sequencing of uncharacterized monkeys to discover spontaneous candidate pathogenic variants and phenotyping of additional candidate animals with relevant traits or deficits will uncover numerous primate models of human genetic disease.

## PgmNr 1365: Hidden genetic variations in *PAX6*-associated aniridia revealed by whole genome sequencing and machine learning based functional annotation.

### Authors:

J. Han <sup>1,2</sup>; J. Lee <sup>1</sup>; Y. Suh <sup>1</sup>; H. Jeung <sup>1</sup>; S.H. Byeon <sup>1</sup>; G. Kim <sup>3</sup>; H.T. Lim <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Ophthalmology, Yonsei University College of Medicine, Seoul, Korea, Republic of; 2) Department of Ophthalmology, Gangnam Severance Hospital, Institute of Vision Research, Yonsei University College of Medicine, Seoul, Korea; 3) Medical Genetics Center, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 4) Department of Ophthalmology, Asan Hospital, University of Ulsan College of Medicine, Seoul, Korea

---

Aniridia (MIM #106210) is a rare, congenital anomaly with autosomal dominant inheritance. It is usually caused by genetic defects of the *PAX6*, a master control gene for eye organogenesis, located on chromosome 11p13. This study was conducted to find hidden pathogenic variants associated with aniridia who were negative for *PAX6* gene sequencing. Two patients with classic aniridia, who were regarded as unsolved after *PAX6* Sanger sequencing or targeted next-generation sequencing (NGS) accompanied with MLPA, were recruited in 2 university-based hospitals. Whole genome sequencing (WGS) was performed using HiSeq X. Variants were filtered and prioritized by Genomiser, CADD (combined annotation dependent depletion), Eigen-PC, DeepSea, FATHMM-MKL (Functional annotation through hidden Markov models-multiple kernel learning). All 2 patients had aniridia, infantile onset-nystagmus, and foveal hypoplasia. None of patients had mental retardation, ataxia and other neurological features. The 2 patients have family history of aniridia, which was consistent with autosomal dominant trait. No copy number variants and large structural variants were found using Delly, ERDS, and CNVnator. The 5' untranslated regions (5'-UTR) c.-133\_-132delTA and splice site c.-52+5G>A variants in non-coding exons were prioritized by Genomiser, CADD, Eigen-PC, DeepSea, FATHMM-MKL algorithms. These two variants were absent in gnomAD and Korean Reference Genome Database. The minigene splicing assay of c.-52+5G>A and luciferase assay of c.-133\_-132delTA confirmed defective splice and abnormal *PAX6* expression. These findings suggest that WGS with machine learning-based algorithm can effectively identify pathogenic non-coding regulatory variant and splice variant in 5' upstream non-coding regions in *PAX6* gene. It is important to find causative mutations in patients with aniridia because it will guide us whether to screen Wilm's tumor.



## PgmNr 1366: AON-mediated exon skipping to bypass protein truncation in retinal dystrophies due to the recurrent *CEP290* c.4723A > T mutation.

### Authors:

I. Barny <sup>1</sup>; I. Perrault <sup>1</sup>; C. Michel <sup>1</sup>; N. Goudin <sup>2</sup>; S. Defoort-Dhellemmes <sup>3</sup>; I. Ghazi <sup>4</sup>; J. Kaplan <sup>1</sup>; J.M. Rozet <sup>1</sup>; X. Gerard <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Genetics in Ophthalmology, Inserm UMR1163, Institute of Genetics Diseases, Imagine, and Paris Descartes University, 75015 Paris, France; 2) Cell Imaging Core Facility of the Structure Fédérative de Recherche Necker, Inserm US24/CNRS UMS3633, Imagine, and Paris Descartes University, 75015 Paris, France; 3) Service d'exploration de la Vision et Neuro-Ophthalmologie, Pôle d'Imagerie et Explorations Fonctionnelles, CHRU de Lille, 59037 Lille, France; 4) Department of Ophthalmology, IHU Necker-Enfants Malades, 75015 Paris, France; 5) Unit of Retinal Degeneration and Regeneration, Department of Ophthalmology, University of Lausanne, Hôpital Ophtalmique Jules Gonin, Fondation Asile des Aveugles, 1004 Lausanne, Switzerland

---

**Introduction and Purpose:** Mutations in *CEP290* encoding a centrosomal protein important to cilia formation and maintenance cause a spectrum of diseases, from retinal dystrophies to multivisceral ciliopathies. In recent years, splicing regulation involving variable mechanisms e.g. endogenous and selective non-canonical exon skipping have shown to contribute to *CEP290* pleiotropy. This observation led us to consider targeted exon skipping to bypass protein truncation resulting from a nonsense mutation in exon 36 (c.4723A > T, p.Lys1575\*) involved in 2% of isolated retinal ciliopathy cases.

**Results:** mRNA analysis in control human retina and fibroblasts revealed the presence of low levels of a *CEP290* isoform lacking exon 36, supporting endogenous basal skipping. This isoform was detected in fibroblasts from two unrelated individuals carrying the p.Lys1575\* mutation (P1 and P2) and affected with early-onset severe retinal dystrophy and congenital blindness, respectively. Although low, the abundance of the *CEP290* isoform lacking exon 36 was higher in patients compared to controls suggesting basal skipping and nonsense associated-altered splicing of mutant exon 36. Furthermore consistent with a more severe disease, fibroblasts from P2 contained less skipped mRNA and exhibited reduced ciliation compared to P1 cells displaying normally abundant cilia; both lines presented however significantly elongated cilia, suggesting altered axonemal trafficking. Antisense oligonucleotides (AONs)-mediated skipping of exon 36 increased the abundance of the premature termination codon (PTC)-free mRNA and protein, reduced axonemal length and improved cilia formation in P2 but not in P1 expressing higher levels of skipped mRNA.

**Discussion:** The present study shows that increasing moderately the quantity of the *CEP290* protein isoform lacking residues encoding exon 36 in mutant cells ameliorated cilia formation whereas high abundance compromised cilia formation. Whether this would occur in photoreceptor cells is an important question. AON-mediated skipping of exon 36 in iPSC-derived retinal organoids from individuals carrying the c.4723A > T mutation would certainly merit consideration to address this

burning question.

## PgmNr 1367: Genotypic profile and phenotype correlations of *ABCA4*-associated retinopathy in Koreans.

### Authors:

K. Joo <sup>1</sup>; M.W. Seong <sup>2</sup>; K.H. Park <sup>1</sup>; S.S. Park <sup>2</sup>; S.J. Woo <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, Gyeonggi-do, Korea, Republic of; 2) Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Republic of Korea

---

This study was conducted to analyze the clinical features associated with the pathogenic variants of *ABCA4* in Korean patients with inherited retinal dystrophies (IRDs). We enrolled patients with IRDs who visited a tertiary referral hospital and identified the pathogenic variants of *ABCA4* by targeted gene panel sequencing and whole exome sequencing. We analyzed the clinical characteristics and phenotypic spectrum according to genotypes. Eleven patients (from nine families) with IRDs and pathogenic variants in *ABCA4* were included. Eight patients (from seven families) with Stargardt disease, two (from one family) with cone-rod dystrophy, and one with early-onset retinitis pigmentosa (RP) were included. Two heterozygous mutations were identified in eight families and one variant was found in a patient with fundus flavimaculatus. Two variants p.Gln294Ter and p.Gln636Lys were associated with severe phenotypes such as early-onset RP and cone-rod dystrophy. Four novel pathogenic variants, p.Gln636Lys, p.Ile1114del, p.Thr1117Ala, and p.Asn1588Tyr, were identified. p.Gln294Ter, p.Leu1157Ter, and p.Lys2049ArgfsTer12 were repeatedly detected in Koreans with *ABCA4*-associated retinopathy. Various pathogenic variants of *ABCA4*, including four novel variants, were identified and *ABCA4*-associated retinopathies exhibited various phenotypes and disease severities in a Korean IRD cohort. These findings will be useful for understanding the clinical features of *ABCA4*-RD and ethnicity-specific variants in East Asians.

## PgmNr 1368: Characterization of genetic spectrum of Leber Congenital Amaurosis (LCA) in Vietnamese patients with blindness.

### Authors:

T. Nguyen <sup>1</sup>; K. Tran Trung <sup>1</sup>; V. Le Sy <sup>2</sup>; L. Nguyen Thanh <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vinmec Research Institute of Stem cell and Gene technology. 458 Minh Khai, Vinh Tuy, Hai Ba Trung, Hanoi, VietNam; 2) University of Engineering and Technology, Vietnam National University Hanoi, 144 Xuan Thuy, Cau Giay, Hanoi

---

Leber's Congenital Amaurosis (LCA) is the most severe form of the retinal dystrophy causing blindness at birth or early onset during the first year of life. To date, 33 genes have been identified to be genetically associated with LCA disease. The mutations in the gene *GUCY2D*, *RPE65*, *CRB1*, *CEP290*, *AiPL1* and *IMPDH1* represent the most frequent in LCA patients worldwide. However, little is known regarding the prevalence and genetic profile of LCA disease in Vietnam.

In Vietnam, there are about 2.000 patients expected to be diagnosed with LCA, spreading across the country from rural and urban areas. Lack of knowledge in molecular basis of LCA disease will lead to misdiagnosis and inappropriate indicated treatment. Therefore, genetic profiling specific for Vietnamese population of LCA-associated genes is essentially needed.

We aim at identifying the genetic profile of previously diagnosed LCA patients at Nguyen Dinh Chieu special school, and Ophthalmology Hospital in Hanoi and other locations in the North area of Vietnam by conducting targeted sequencing on specific genes in combination with Multiplex Ligation-dependent Probe Amplification (MLPA) for copy number variation detection. Variant spectrum and genetic frequency of LCA associated genes will provide insights to conduct important research project in the future such as (1) initiating a potential project on gene therapy, given that *RPE65* gene is an important target for a gene therapy approved by FDA for LCA treatment and (2) establishing a prenatal test for patients with high genetic risk of the disease.

## PgmNr 1369: A study of inherited retinal disease candidate genes.

### Authors:

K.H. Xiong <sup>1,2</sup>; C. Thangavel <sup>1,2</sup>; R.E. Bertrand <sup>1,3</sup>; J. Wang <sup>1</sup>; Y. Li <sup>1</sup>; R. Chen <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.; 2) Department of BioSciences, Rice University, Houston, TX; 3) Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

---

Inherited retinal degeneration (IRD) is a group of diseases that leads to incurable vision loss due to death or dysfunction of photoreceptor cells. Although over 200 genes have been implicated in IRDs, together, mutations in these genes account for about 70% of the patients, leaving about 30% of patients without molecular diagnoses. Using a combination of whole genome and whole exome sequencing data from human patients and multi-omics datasets, a list of candidate disease-related genes has been identified. To validate these candidate genes, knockout mouse models generated via the CRISPR/Cas9 gene editing method are used to evaluate the gene function in retina. Retina of the KO mice were characterized by evaluating the morphological and functional phenotype of the retina at various timepoints. To study the progression of functional retinal degeneration, electroretinograms were used to measure the response of photoreceptor cells to light stimulation. Dark-adapted and light-adapted conditions were used to distinguish rod and cone responses. Retinal morphology was analyzed by using hematoxylin-and-eosin-stained retinal sections to visualize photoreceptor cell death. Knockout mice exhibiting either histological or functional defects provide strong evidence of the association of the gene with human IRDs. Using this approach, several novel human IRD disease genes have been identified, including *CWC27*, *REEP6*, and recently *FAM57B*. Further characterization of these mouse models is currently underway, which will provide additional insights of the disease mechanisms and form the basis for developing novel therapies. Progress of evaluation of additional mouse models will also be reported.

## PgmNr 1370: Biases in arginine codon usage correlate with genetic disease risk.

### Authors:

K.V. Schulze<sup>1,2</sup>; N.A. Hanchard<sup>1,2</sup>; M.F. Wangler<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 2) USDA/ARS/Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA; 3) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, USA

---

Arginine substitutions underlie 20% of all 'Pathogenic' single nucleotide variants (SNVs) listed in ClinVar, thus making arginine the most commonly substituted amino acid. Arginine can be encoded by four codons containing CpG dinucleotides ('CGX': CGA, CGC, CGG, CGU) or two alternative codons ('AGR': AGA, AGG). The persistence of the four CGX codons in human genes, despite the hypermutability of CpG dinucleotides, suggests the possibility of negative selective pressure at these sites, and that arginine codon usage could be an indicator of disease genes. We hypothesized that genes rich in CGX codons are particularly intolerant to variation and thus result in a greater burden of human disease.

We analyzed arginine codons from all 'canonical' Ensembl protein coding gene transcripts (hg38). Although AGR codons made up 42.3% of all arginine codons, fewer than 20% of arginines were encoded by AGR codons among a subset (16.3%) of CGX-rich, 'hyper-mutable' genes. We found that the frequency of CGX codons among a gene's total arginine codon count was significantly higher in genes linked to syndromic autism spectrum disorder (ASD) compared to genes not associated with ASD ( $p=0.006$ , t-test); this was exemplified by *CHD8* (MIM#610528), which had a CGX frequency of 72% ( $n=165$  total arginines) and for which 8/12 ClinVar 'Pathogenic' SNVs involved CGX codons. Strikingly, a comparison of genes annotated as dominant or recessive in the OMIM database with control genes not matching either classification revealed a progressive increase in CGX frequency moving from control to recessive to dominant genes ( $p<2E-16$ , ANOVA). We also found that CGX frequency is positively correlated with a gene's probability of loss-of-function (pLI) score ( $r=0.099$ ,  $p<2.2E-16$ ) and negatively correlated with gnomAD's 'observed-over-expected' (o/e) ratios for both loss of function ( $r=-0.128$ ,  $p<2.2E-16$ ) and missense mutations ( $r=-0.163$ ,  $p<2.2E-16$ ). We speculated that more highly expressed genes in regions of high GC content might contribute more to these associations with CGX arginine codon frequency; indeed, we found that GC content and median Cerebellum gene expression were both positively correlated with CGX frequency ( $r=0.537$ ,  $p<2.2E-16$  and  $r=0.172$ ,  $p<2.2E-16$ , respectively).

Our findings indicate that genes utilizing CGX codons are more likely to underlie single gene disorders and thus constitute a robust set of candidate genes, particularly for the study of dominant disease, including ASD.

## PgmNr 1371: Mouse models of glycobiology-related diseases: Identification of understudied genes.

### Authors:

M. Tomczuk; C. Smith; C. Bult; the Mouse Genome Informatics Staff and Software Team

[View Session](#) [Add to Schedule](#)

**Affiliation:** Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME

---

Glycans function in a variety of structural and modulatory roles and in regulation of cell-cell and protein-protein interactions. While progress has been made in discovering causes of glycobiology-related diseases, the many roles attributed to glycans make it difficult to identify these disorders in patients who present with variable symptoms. Mice are a superb organism to model diseases: they are genetically and physiologically similar to humans, all life stages can be accessed, completely sequenced, well annotated reference genomes for multiple strains are available, and many genomic tools are available for comparative and experimental manipulation. Mice with mutations affecting glycosylation, glycan degradation and carbohydrate metabolism can aid in disease discovery and identifying patients affected by such changes.

Mouse Genome Informatics (MGI, [www.informatics.jax.org](http://www.informatics.jax.org)) is the authoritative resource for the laboratory mouse, integrating data from multiple sources to facilitate links between mouse phenotypes and human disease symptoms. We will show how users can utilize MGI to gain understanding of glycoprotein-related gene function by analyzing phenotypes of mouse mutants and how mouse models of these diseases can be found and used to gain insights into disease pathogenesis and treatment, such as the Hexb knock-out used to study enzyme replacement therapy for Sandhoff disease. The more extensively researched diseases include muscular dystrophies like muscular dystrophy-dystroglycanopathy for which 8 mouse models involving 5 genes (B4gat1, Fkrp, Fktn, Large1, Pomt1) are reported and cellular storage disorders such as Gaucher's and Tay-Sachs for which multiple models are available. In collaboration with GlyGen ([www.glygen.org](http://www.glygen.org)), a resource for assisting glycoscience research, MGI surveyed a list of 96 genes involved in glycobiology to determine where informational gaps exist. We find that mice with mutations in or knock-outs of these genes are understudied, with 81 of these genes having collectively 204 germline transmissible alleles but phenotypes reported for only 67 of the genes and only alleles in 15 of these genes reported as disease models. In addition, complete elimination of individual glycans or relevant genes, such as Pigo and Pomt1, often results in embryonic lethality, supporting their biological importance and the need to develop better models with disease symptoms such as allelic variants, hypomorphic and tissue-specific targeted alleles.

## PgmNr 1372: 49,XXXXY syndrome: Report of a Colombian patient.

### Authors:

S. Maradei; J. Acosta Guio; A. Muñoz

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, Biotecgen S.A., Bogota, Cundinamarca, Colombia

---

**Introduction:** 49,XXXXY Syndrome represents an aneuploid chromosomal aberration in which there are three extra X chromosomes in boys and men. It affects over 1/85.000 to 1/100.000 newborn males, and is a different clinical entity than Klinefelter Syndrome (KS). Here we present the case of an 11 year old Colombian patient who was diagnosed with this condition.

**Case Report:** A prepuberal boy with short stature, congenital cardiopathy, hypogonadism (micropenis, hypoplastic scrotum and cryptorchidism), bilateral radioulnar synostosis, and a personal history of developmental delay and behavioural alterations, was first evaluated by an endocrinologist who suspected KS. However, and given the co-appearance of multiple congenital anomalies, the possibility of another gonadal dysgenesis disorder or a chromosomal imbalance was on the table, so the treating physician referred the patient to the Medical Genetics consultation for further examination. A conventional Karyotype analysis was ordered, showing presence of an abnormal number of the sexual chromosomes: 49,XXXXY. Comparative Genomic Hybridization was also carried out, confirming the diagnosis and excluding other copy number variants as causative. The boy is now undergoing a series of medical evaluations in order to give him the appropriate treatment and to establish the prognosis. Also, genetic counseling for the family was offered.

**Comments:** 49,XXXXY syndrome is a rare chromosomal condition characterized by severe hypogonadism, variable low IQ levels and progressive cognitive impairment with age, growth retardation that can appear shortly after birth or even *in utero*, congenital defects of the heart, skeletal anomalies or deformities (radioulnar synostosis, epiphyseal dysplasia, coxa valga, kyphoscoliosis, joint hypermobility), other dysmorphic features that are usually mild (hypertelorism, epicanthal folds, broad flat nasal bridge or tip, folded pinna, short neck), and in some cases added compromise of other systems/organs (hypoplasia of the corpus callosum, renal hypoplasia, visual acuity disorders). As most cases appear sporadically, recurrence risk for the family is low. Patients suffering from this condition need a multi-disciplinary approach for monitoring every possible outcome, despite a relatively normal life expectancy.



## PgmNr 1373: Long-sized AOH regions reveal IBD haplotypes that drive Mendelian disease through mutational burden at a locus.

### Authors:

Z. Coban Akdemir<sup>1,12</sup>; X. Song<sup>1,12</sup>; D. Pehlivan<sup>1,2</sup>; E. Karaca<sup>1,3</sup>; Y. Bayram<sup>1,4</sup>; T. Gambin<sup>5</sup>; S.N. Jhangiani<sup>6</sup>; D. Muzny<sup>6</sup>; R. Lewis<sup>1</sup>; P. Liu<sup>1</sup>; E. Boerwinkle<sup>6,7</sup>; A. Hamosh<sup>8</sup>; R.A. Gibbs<sup>1,6</sup>; N. Sobreira<sup>8</sup>; C.M.B. Carvalho<sup>1</sup>; J.E. Posey<sup>1</sup>; C.A. Shaw<sup>1,9</sup>; D. Valle<sup>8</sup>; J.R. Lupski<sup>1,6,10,11</sup>; Baylor Hopkins Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, Texas.; 2) Section of Neurology, Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030; 3) Department of Genetics, University of Alabama at Birmingham, Birmingham, AL 35294; 4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 5) Institute of Computer Science, Warsaw University of Technology, Warsaw 00-665, Poland; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 7) Human Genetics Center, University of Texas Health Science Center at Houston, TX 77030, USA; 8) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205, USA; 9) Baylor Genetics, Houston, TX 77021, USA; 10) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 11) Texas Children's Hospital, Houston, TX 77030, USA; 12) These two authors contributed equally

---

In classical human and medical genetics, we learn that population structure is relevant to disease because of individual pathogenic alleles confined and enriched within specific populations. However, the impact of population structure on rare – even private – variants, or combinations of alleles at a locus, to disease traits remains unexplored. We hypothesized that within small or consanguineous populations long-sized absence of heterozygosity (AOH) regions are more likely be formed on recently configured haplotypes that could be enriched for combinations of rare deleterious variants per locus because time would not be sufficient for recombination and selection to eliminate these haplotypes from the population. To test this hypothesis, we studied a large Turkish (TK) cohort with high levels of consanguinity and admixture. In this cohort, we investigated the genetic architecture of disease by exome sequencing (ES) and family based genomics in 556 individuals affected with diverse disease phenotypes. As anticipated, we observed a larger amount of genomic intervals with AOH in TK affected (162.1 Mb) and unaffected (78.8 Mb), compared to non-TK individuals (72 Mb). This difference was more striking in long-sized AOH regions enriched for rare homozygous deleterious variants in the individual genomes of TK affecteds compared to TK-unaffecteds and non-TK affecteds ( $P=5.4^{-10}$  and  $< 2.2e^{-16}$ ). A systematic analysis of genotype-phenotype correlations using the complete haplotype in long-sized ( $>1.606$  Mb), but not short-sized (40-515 Kb) or medium-sized (0.515-1.606 Mb) AOH, is the most explanatory for the observed disease traits by systematic analysis of HPO (Human Phenotype Ontology) terms and controlling for the number of genes or size of the block using a permutation approach. Unexpectedly, and an interesting genomic observation with relevance to transmission of genetic traits, not just an increased coefficient of consanguinity in a population but rather identity-by-descent (IBD) in clans enabled by long-sized AOH regions is associated with a

higher deleterious rare variant burden at a locus contributing to the disease traits. In summary, our results provide compelling evidence that IBD is correlated with a higher number of long-sized AOH blocks allowing an elevated mutational burden contributing to disease traits in the TK probands. Moreover, our findings further support the Clan Genomics hypothesis and also allow insights into the La Reunion Paradox in human genetics.

## PgmNr 1374: Extensive allelic and phenotypic heterogeneity revealed by point-of-care exome sequencing of Mendelian disorders in Qatar.

### Authors:

K.A. Fakhro <sup>1,2,3</sup>; A. Robay <sup>2</sup>; J.L. Rodriguez-Flores <sup>2</sup>; J.G. Mezey <sup>2</sup>; A. Al-Shakaki <sup>2</sup>; O. Chidiac <sup>2</sup>; I. Janahi <sup>1,4</sup>; A. Imam <sup>1,4</sup>; R. Ali <sup>4</sup>; J... Malek <sup>2</sup>; N. Numeiri <sup>4</sup>; A. Azzam <sup>2</sup>; F. Butt <sup>4</sup>; T. Ben Omran <sup>4</sup>; R.G. Crystal <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Sidra Medicine, Doha, Qatar; 2) Department of Genetic Medicine, Weill Cornell Medical College, Doha, Qatar and New York, NY; 3) College of Health Sciences, Hamad Bin Khalifa University, Doha, Qatar; 4) Hamad Medical Corporation, Doha, Qatar

---

The effectiveness of next generation sequencing at solving genetic disease has motivated its rapid adoption into clinical practice. In this study, we use whole exome sequencing (WES) to assess 48 patients with Mendelian disease from 30 serial families collected at the point of care, as part of the Qatar Mendelian Disease Program – a coordinated multi-center effort to build capacity and clinical expertise in genetic medicine in Qatar. By enrolling whole families (parents plus all available siblings), we demonstrate significantly improved power to discriminate candidate disease variants *versus* only sequencing trios; this held true for both *de novo* and recessive inheritance patterns. For the same index cases, we further demonstrate that even in the absence of families, variant prioritization is improved up to 8-fold when a modest set of population-matched controls is used vs large public databases (1,346 Qatari controls vs >250,000 alleles in public datasets), stressing the very poor representation of Middle Eastern alleles in presently available databases. Altogether, our in-house best practices pipeline identified candidate disease variants in 27 of 30 families (90%), 23 of which (85%) harbor novel pathogenic variants, some of which affected known disease genes, pointing to significant allelic heterogeneity and founder mutations underlying Mendelian disease in the Middle East. For 6 of these families, the clinical presentation was only partially explained by the candidate gene, suggesting phenotypic expansion of known syndromes. Our pilot study therefore demonstrates the utility of WES for Middle Eastern populations, the dramatic improvement in variant prioritization conferred by enrolling population-matched controls and/or enrolling additional unaffected siblings at the point-of-care, and 25 novel disease-causing alleles, relevant to newborn and premarital screening panels in populations of Arab ancestry. Moreover, this study lays down the groundwork, pipeline and local capacity for future whole-genome sequencing programs targeting rare diseases in Qatar and the Middle East.

## PgmNr 1375: High percentage of novel pathogenic alleles and loci found in more than 1500 parent child trio exomes from India in combination with a large diagnostics Indian database.

### Authors:

E. Stawiski <sup>1</sup>; R. Gupta <sup>2</sup>; A. Verma <sup>2</sup>; S. Phalke <sup>3</sup>; P. Raj <sup>3</sup>; J. Wall <sup>4</sup>; R. VL <sup>3</sup>; A. Peterson <sup>1</sup>; I.C. Verma <sup>5</sup>; S. Seshagiri <sup>6,7</sup>; FGDS Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Medicines Division, MedGenome, Foster City, CA; 2) Genomic Medicines Division, MedGenome, Bangalore, India; 3) Diagnostics Division, MedGenome, Bangalore, India; 4) Inst Human Gen, UCSF, San Francisco, CA; 5) Center of Medical Genetics, Sir GangaRam Hospital, New Delhi, India; 6) Molecular Biology, Genentech, S San Francisco, CA; 7) SciGenom Research Foundation, Bangalore, India

---

**Background:** Familial disease has largely been studied in populations of European ancestry. India presents a very different context with its multitude of diverse population groups, many of which have high rates of homozygosity and higher incidence of consanguineous unions. There are many genetically differentiated population groups when compared to Europe with genetic variation specific to the region.

**Methods:** To date we have performed exome sequencing on 4914 samples (median depth 77x) from 1501 families with suspected mendelian disorders representing 14 broad disease areas. Of the probands diagnosed, 42% were of consanguineous unions, the average age at presentation was 7 years and 63% were male. In addition, we cross checked candidate segregated variants to a large diagnostics database in India (MedVarDB) composed of 31,736 samples where clinical exome sequencing had been performed and phenotypic information was available.

### Results:

To date we have identified 200 alleles not previously reported in OMIM/ClinVar/HGMD. The majority of these alleles (n=126) are not found in the ExAC, or 1000G databases with the remaining alleles having an average ExAC allele frequency of <0.00006. For many of these novel alleles (n=138) we were able to match genotypes to MedVarDB that could be used to estimate allele frequencies in the Indian sub-continent and confirm phenotypes. The majority of these impact diseases related to neurology, metabolic disorders, ENT, immunology and other rare diseases. One specific example is RAB27A where we found 3 novel alleles in 6 families (4 of consanguineous unions, all homozygous) linked to Griscelli syndrome. All have not previously been reported in public databases however 2 are present at low allele frequencies within the Indian population (up to 0.0002 % in MedVarDB). In addition, we found a novel gene associated with Hemophagocytic lymphohistiocytosis. Of those cases reported, consanguineous unions showed a higher degree of diagnosis rates compared to non-consanguineous unions (80% vs 59%) as well as a higher rate of homozygous reported variants (79% vs 32%). We found higher degrees of homozygosity within our dataset for both the consanguineous

and non-consanguineous unions than would be expected compared to Europeans.

**Conclusion:** Expanding the genetic diversity in large scale familial studies will improve diagnostics worldwide as well as link new biological insights and treatment of mendelian disorders.

## **PgmNr 1376: Outcomes of selective use of rapid clinical exome sequencing for critically ill patients in a tertiary care center.**

### **Authors:**

A. Kim; J. Britton; C. Applegate; A. Hamosh

[View Session](#) [Add to Schedule](#)

**Affiliation:** McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore

---

We describe our ongoing experience with the selective use of rapid clinical exome sequencing (rES) for critically ill patients at The Johns Hopkins Hospital. This involves review by a multidisciplinary committee that assesses the likelihood that this expensive testing will affect medical management.

We have evaluated 22 inpatients with rES since 2015, the majority of whom were in the neonatal and pediatric intensive care units. Eleven had prior genetic testing. The primary indication for rES was neurologic (e.g., hypotonia) in 12 cases, gastrointestinal or hepatobiliary (e.g., acute liver failure) in 5 cases, and allergic or immunologic (e.g., immunodeficiency) in 3 cases. rES yielded a positive result in 9 (41%) cases, a negative result in 4 cases, and a non-informative variant of uncertain significance (VUS) in 9 cases. There were no dual Mendelian diagnoses. Of the 9 positive results, 6 were from rES ordered for neurologic primary indications. Newly-established diagnoses included combined oxidative phosphorylation deficiency 23 (MIM 616198), infantile liver failure syndrome 1 (MIM 615438), and RAS-associated autoimmune leukoproliferative disorder (MIM 614470). As expected, all positive results affected medical management. For example, we initiated treatment for a neonate with glucose/galactose malabsorption (MIM 606824), resolving life-threatening diarrhea. When the newly-established diagnosis was untreatable, as was the case for an infant with TK2-related mitochondrial DNA depletion syndrome 2 (MIM 609560), the result helped clarify the goals of care and concluded the diagnostic odyssey. Non-diagnostic negative results and VUSs also helped clarify goals of care and streamline differential diagnoses, and thereby affected medical management. Specific changes include transitions to comfort care, tracheostomy and gastrostomy, and prioritization of functional testing (e.g., neurotransmitter metabolite testing, global metabolomics, cytokine profiling, flow cytometry). Three inpatients died prior to return of results.

Our ongoing experience demonstrates that rES can decisively influence medical management. Both diagnostic and non-diagnostic results facilitated decision-making and clarified the goals of care. This allowed medical teams and families to prioritize treatment plans over additional diagnostic studies with diminishing returns, and avoid prolongation of extreme life-sustaining measures in some cases.

## PgmNr 1377: Potential involvement of the novel gene *ArfGAP2* in inherited retinal degeneration.

### Authors:

R. Ayyagari <sup>1</sup>; P. Biswas <sup>1</sup>; A. Chekuri <sup>1</sup>; J. Ear <sup>2</sup>; S. Borooah <sup>1</sup>; H. Matsui <sup>3</sup>; S.Y. Khan <sup>4</sup>; M. Ali <sup>4</sup>; M.A. Naeem <sup>5</sup>; M. D'Antonio <sup>3</sup>; S. Srivatsan <sup>6</sup>; K.A. Frazer <sup>3,7</sup>; S. Riazuddin <sup>5,8</sup>; S.A. Riazuddin <sup>4</sup>; P. Ghosh <sup>2</sup>; J.F. Hejtmancik <sup>9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ophthalmology, Univ California San Diego, La Jolla, California.; 2) Departments of Medicine and Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA USA; 3) Institute for Genomic Medicine, University of California, San Diego, La Jolla, CA, USA;; 4) The Wilmer Eye Institute, Johns Hopkins University School of Medicine, Baltimore, MD USA; 5) National Centre Excellence in Molecular Biology, University of the Punjab, Lahore Pakistan; 6) Department of Genetics & Molecular Biology, Vision Research Foundation, Sankaranethralaya, Chennai, India; 7) Department of Pediatrics, Rady Children's Hospital, Division of Genome Information Sciences, San Diego, CA, USA; 8) Allama Iqbal Medical College, University of Health Sciences Lahore, Lahore, Pakistan;; 9) Ophthalmic Genetics and Visual Function Branch, National Eye Institute, NIH, Bethesda, MD, USA.

**Purpose:** To identify the molecular basis of inherited retinal degeneration (IRD) in a large consanguineous pedigree with multiple affected members using linkage, haplotype, whole exome and whole genome analyses and functional evaluation of the impact of the potential causative variant segregating with disease.

**Methods:** A large consanguineous pedigree with autosomal recessive retinal degeneration underwent complete ophthalmic evaluation including funduscopy and electroretinography. Linkage and haplotype analyses were performed using microsatellite markers. Whole exome and whole genome analyses were carried out to identify potential disease causative variants. Mapping and variant calling was performed using BWA and GATK and structural variants were called using Genome STRiP and LUMPY. Variants were prioritized for further analysis using the pipeline we had developed. The effect of the potential mutation in the *ADP Ribosylation Factor GTPase activating protein 2* (*ArfGAP2*) gene was studied by evaluating transfected cells and mouse retinal tissue using qRT-PCR, immunostaining, electron microscopy (EM), immunoblotting and GTP activating protein (GAP) assay.

**Results:** Linkage and haplotype analysis mapped the disease locus to a 19.9Mb interval on chromosome-11. Whole exome and genome sequence analysis identified a variant, p.Arg255Cys in *ArfGAP2*, the only pathogenic and rare variant residing within the linkage interval, co-segregating with recessive IRD. A protein interaction network analysis of *ArfGAP2* showed significant association with IRD implicated genes. In the retina, *ArfGAP2* localized to the photoreceptors, inner-nuclear, and ganglion cells. Cells expressing the mutant-*ArfGAP2* demonstrated trimerization of the protein, fragmentation of Golgi, and high Arf1-activity, indicative of impaired GAP function. The mutant -255Cys was predicted to generate novel disulphide bonds enabling pathogenic trimerization of *ArfGAP2* on the Golgi membrane resulting in abnormal vesicular traffic.

**Conclusion:** A homozygous p.Arg255Cys variant in the *ArfGAP2* gene segregated with IRD. This mutation results in impaired GAP function leading to hyper-activation of Arf1 and abnormal vesicular traffic in cells expressing the mutant *ArfGAP2*.

## PgmNr 1378: *De novo* missense variants in the *LMBRD2* gene in five unrelated clinical cases suggest association with developmental delay, intellectual disability, and dysmorphic features.

### Authors:

A. Malhotra<sup>1</sup>; A. Ziegler<sup>2</sup>; R. Perrier<sup>3</sup>; S. Li<sup>4</sup>; E. Wohler<sup>5</sup>; N. Sobreira<sup>5</sup>; X. Mao<sup>4</sup>; D. Bonneau<sup>2</sup>; D. Perry<sup>1</sup>; A. Vanderver<sup>6</sup>; R. Taft<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Illumina Clinical Services Laboratory, Illumina Inc., San Diego, CA; 2) Department of Biochemistry and Genetics, University Hospital of Angers, France; 3) Alberta Children's Hospital, Alberta, Canada; 4) Maternal and Child Health Hospital of Hunan Province, Changsha, China; 5) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Children's Hospital of Philadelphia, Philadelphia, PA

---

LMBRD2 (LMBR1 domain-containing 2) is a poorly understood but widely expressed transmembrane protein with no known disease association. Intriguingly, however, both LMBR1 (limb development membrane protein 1) and LMBRD1 (LMBR1 domain-containing 1) are associated with polydactyly and methylmalonic aciduria and homocystinuria, respectively. In addition, recent reports have shown that LMBRD1 is essential for gastrulation. The functions of these paralogues suggest a possible role for LMBRD2 in fundamental developmental processes.

Here we describe five unrelated cases with novel *de novo* missense variants in *LMBRD2*, resulting from a GeneMatcher collaboration. Clinical or research-based whole exome sequencing (WES) or clinical whole genome sequencing (cWGS) revealed *c.577T>C* (*p.Trp193Arg*) and *c.976C>G* (*p.Gln326Glu*) variants in one case each, and a recurrent *c.1448G>A* (*p.Arg483His*) variant in the remaining three cases. All variants are absent from population allele frequency databases including the Genome Aggregation Database and are predicted to be damaging by multiple algorithms including PrimateAI, SIFT, and PolyPhen. No additional variants were identified in four of the five cases that could potentially explain the patient's clinical presentations.

The patients ranged in age from one year to 14 years of age. Phenotypic overlap was observed among cases, which included developmental delay in all patients, intellectual disability (three patients), thin corpus callosum (three patients), microcephaly (two patients), hypotonia (two patients), and dysmorphic features (two patients). Variability in clinical presentation was also noted, with additional phenotypes of spastic paraplegia, clonus, facial features characteristic of Kabuki syndrome, patent foramen ovale, hypomyelination of white matter, and seizures being observed once.

Taken together, these findings suggest that *de novo* variants in *LMBRD2* can lead to a pronounced neurological phenotype. This finding is broadly consistent with the observation that widely expressed and so-called house-keeping genes are associated with a wide range of central nervous system disorders, including primary white and gray matter diseases. Further functional studies of *LMBRD2* are



warranted to better understand its role in neurological development.

## PgmNr 1379: Biallelic *GALM* pathogenic variants cause type IV galactosemia.

### Authors:

Y. Wada<sup>1</sup>; A. Kikuchi<sup>1</sup>; N. Arai-Ichinoi<sup>1</sup>; O. Sakamoto<sup>1</sup>; Y. Takezawa<sup>1</sup>; S. Iwasawa<sup>1</sup>; T. Niihori<sup>2</sup>; H. Nyuzuki<sup>3</sup>; Y. Nakajima<sup>4</sup>; E. Ogawa<sup>5</sup>; M. Ishige<sup>5</sup>; H. Hirai<sup>6</sup>; H. Sasai<sup>7</sup>; R. Fujiki<sup>8</sup>; M. Shirota<sup>9</sup>; R. Funayama<sup>10</sup>; M. Yamamoto<sup>11</sup>; T. Ito<sup>9</sup>; O. Ohara<sup>8</sup>; K. Nakayama<sup>10</sup>; Y. Aoki<sup>2</sup>; S. Koshiba<sup>11</sup>; T. Fukao<sup>7</sup>; S. Kure<sup>1,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan; 2) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan; 3) Department of Pediatrics, Niigata University School of Medicine, Niigata, Japan; 4) Department of Pediatrics, Fujita Health University School of Medicine, Toyoake, Japan; 5) Department of Pediatrics and Child Health, Nihon University School of Medicine, Tokyo, Japan; 6) Department of Pediatrics, Ehime Prefectural Central Hospital, Matsuyama, Japan; 7) Department of Pediatrics, Graduate School of Medicine, Gifu University, Gifu, Japan; 8) Department of Technology Development, Kazusa DNA Research Institute, Kisarazu, Japan; 9) Division of Interdisciplinary Medical Sciences, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 10) Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 11) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan.

### BACKGROUND:

Galactosemia is caused by metabolic disturbances of galactose; types I, II, and III galactosemia, which are attributed to congenital enzymatic deficiencies in the Leloir pathway, portosystemic shunts, cholestasis and other metabolic disorders. A typing of galactosemia is crucial for appropriate care. Nevertheless, the causes of galactosemia has not been identified in a subset of the patients.

### METHODS:

To explore the etiology, eight patients with unexplained congenital galactosemia were enrolled. Trio-based whole-exome sequencing was performed in two unrelated families. Additional Sanger sequencing was performed in other six patients.

### RESULTS:

We identified biallelic pathogenic variants in the *GALM* gene encoding aldose 1-epimerase which catalyzes epimerization between  $\beta$ - and  $\alpha$ -D-galactose in the first step of the Leloir pathway by the whole-exome sequencing. Biallelic *GALM* pathogenic variants (p.Trp82\*, p.Ile99Leufs\*46, p.Gly142Arg, p.Arg267Gly, and p.Trp311\*) were identified in all eight patients. Bilateral cataracts were observed in two patients. Another two patients had also transient portosystemic shunts. All eight patients did not have any other complications or sequelae. *GALM* enzyme activities were undetectable in lymphoblastoid cell lines from two patients. The immunoblot analysis showed the absence of the *GALM* protein in peripheral blood mononuclear cells from three patients. In vitro *GALM* expression and protein stability assays showed decreased stability of the mutant *GALM* proteins.

#### DISCUSSION:

In summary, biallelic *GALM* pathogenic variants cause “type IV” galactosemia. The phenotypic spectrum of *GALM* deficiency may be similar to *GALK* deficiency.

## PgmNr 1380: Evolution of the facial features in a patient with the rare Siderius type XLID syndrome caused by hemizygous loss of PHF8.

### Authors:

O. Petryna <sup>1</sup>; A. Sobering <sup>1</sup>; Z. James <sup>1</sup>; T. Donald <sup>2</sup>; B. Nelson <sup>2</sup>; J. Beighley <sup>3</sup>; J. Gerdtz <sup>3</sup>; E. Bhoj <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) St. George's University School of Medicine, True Blu, Grenada; 2) Pediatrics Ward, Grenada General Hospital, Grenada; 3) Department of Psychiatry and Behavioral Sciences, University of Washington, USA; 4) Children's Hospital of Philadelphia, Philadelphia, PA, USA

---

Most resource-limited countries are unable to offer medical genetic consultation and diagnosis; this is in sharp contrast to the standard of care that is expected in the developed countries of North America, the UK, and Western Europe. Collaboration with a US-based laboratory allowed us to identify an Afro-Caribbean person with Siderius type X-linked intellectual disability syndrome (*PHF8*; OMIM 300560). The family of this patient previously had limited access to genetic medicine. There are only 4 reports describing this ultra-rare disorder in the medical literature. It is caused by hemizygous loss of *PHF8*, which is a dual function histone methylation reader and lysine demethylase enzyme. Exome sequencing showed that our patient had a protein truncation variant, c.1996delG, that causes a frameshift and a stop codon 163 amino acids following the mutation. We present the evolution of the facial features of our patient including low-set and posteriorly rotated ears, hypertelorism with epicanthus, broad nose, prominent chin, and a narrow palate. Our patient carries a clinical diagnosis of autism and attention deficit hyperactivity disorder. Understanding the genetic cause of the disorder in this patient has ended his diagnostic odyssey. His parents report a positive impact on family life and the management of his condition, and they are grateful to have a medical reason to explain his condition. Our work has enabled us to provide recurrence risk counseling to the parents as we have identified that the mother and the sister of the proband are carriers. We are also helping to develop autism awareness in the community and to dispel myths about causation of genetic disorders.

## PgmNr 1381: Diagnostic success across 300 heterogeneous rare disease cases from the BBMRI-LPC call analysed using the RD-Connect Genome-Phenome Analysis Platform.

### Authors:

S. Laurie <sup>1</sup>; L. Matalonga <sup>1</sup>; S. Beltran <sup>1,2</sup>; D. Piscia <sup>1</sup>; M. Bayes <sup>1</sup>; B. Fusté <sup>1</sup>; M. Gut <sup>1,2</sup>; J. Dawson <sup>3</sup>; R. Thompson <sup>3</sup>; E. López-Martin <sup>4,5</sup>; M. Posada <sup>4,5</sup>; C.M. Wang <sup>6</sup>; L. Monaco <sup>6</sup>; G.B. van Ommen <sup>7</sup>; S. Sims <sup>8</sup>; E. Zeggini <sup>9</sup>; H. Lochmüller <sup>1,10</sup>; I. Gut <sup>1,2</sup>; BBMRI-LPC Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centro Nacional de Análisis Genómico (CNAG-CRG), Center for Genomic Regulation, Barcelona Institute of Science and Technology (BIST), Barcelona, Catalunya, Spain; 2) Universitat Pompeu Fabra (UPF), Barcelona, Catalunya, Spain; 3) Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, United Kingdom; 4) Institute of Rare Diseases Research, IIER-ISCIII, Madrid, Spain; 5) Centre for Biomedical Network Research on Rare Diseases, CIBERER, Madrid, Spain; 6) Fondazione Telethon, Milan, Italy; 7) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 8) Twist Bioscience, Cambridge, UK; 9) Helmholtz Zentrum München, Munich, Germany; 10) Children's Hospital of Eastern Ontario, Toronto, Ontario, Canada

---

The BBMRI-LPC WES Call was a transnational collaboration between the Biobanking and BioMolecular Resources Research Infrastructure-Large Prospective Cohorts (<http://www.bbmri-lpc-biobanks.eu/>), RD-Connect (<https://rd-connect.eu/>) and EuroBioBank (<http://www.Eurobiobank.org/>). It offered free-of-charge exome sequencing to selected collaborative projects involving PIs from at least two different countries, focusing on distinct rare diseases. A requisite of the call was that a biological specimen from each case to be sequenced be deposited at a EuroBioBank network member, to be available to the wider rare disease research community. Furthermore, a deep phenotypic descriptions for each affected case was required, and collated using Human Phenotype Ontology (HPO), OMIM and Orpha codes to facilitate case resolution.

Seventeen projects were selected for sequencing. All samples were processed using the RD-Connect standard analysis pipeline (Laurie et al, Human Mutation, 2016), and variant filtration and prioritisation undertaken using the RD-Connect Genome-Phenome Analysis Platform (GPAP, <https://platform.rd-connect.eu>). Advanced GPAP features such as on-the-fly gene-lists based upon gene-HPO relationships, homozygosity mapping, pathway associations, and patient matching through Matchmaker Exchange (<https://www.matchmakerexchange.org/>) expedited the identification of causative variants. Analysis and interpretation of cases was undertaken in collaboration between a clinical genomics specialist and relevant disease-domain experts.

Here we report the outcome of the call to date, for which the molecular diagnostic rate is approximately 40% across the project, despite the heterogeneous nature of the different phenotypes under investigation, ranging from neuromuscular disorders, through inborn errors of metabolism, to albinism. This is in line with similar studies which have focussed on a more restricted range of phenotypes. Of note, ten novel gene-disease relationships are currently undergoing functional

validation. Some of the cases that remain unsolved are being taken forward in the SolveRD project (<http://solve-rd.eu/>), where they will be included in more advanced analyses, and we anticipate the resolution rate will increase further.

## PgmNr 1382: Association of $\alpha$ -thalassemia and glucose-6-phosphate dehydrogenase polymorphisms with high-risk transcranial doppler velocity and stroke in Nigerian children with sickle cell disease.

### Authors:

O.O. Ojewunmi <sup>1</sup>; A.I. Oyetunji <sup>1</sup>; Y. Benn <sup>1</sup>; M.G. Ekpo <sup>2</sup>; T.A. Adeyemo <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Diagnostics, Sickle Cell Foundation Nigeria, Lagos, Surulere, Nigeria. P.O. Box 3463, Surulere, Lagos, Nigeria; 2) Department of Biotechnology, Afe babalola University, Ado-Ekiti, Ekiti, Nigeria; 3) Department of Haematology and Blood transfusion, College of Medicine, University of Lagos, P.M.B. 120003, Idi-Araba, Lagos

---

Stroke is a major devastating complication of sickle cell disease (SCD) and contributes significantly to morbidity and mortality of SCD. It can be predicted through abnormally high cerebral velocities using transcranial Doppler (TCD). Co-inheritance of  $\alpha^{3.7\text{kb}}$ -thalassemia deletion and glucose-6-phosphate dehydrogenase (*G6PD*) deficiency in SCD (*HBB*: c.20A>T) is associated with haemolysis but their role in the development of stroke in children with SCD has been controversial. This study aimed at investigating the influence of alpha-thalassemia and G6PD polymorphism (202 G>A and 376 A>G) on Nigerian children with clinical history of stroke or high-risk TCD velocities. Of 155 SCD children (2 – 16 years) presented in this study, 73 had normal TCD risk (time-averaged mean of the maximum velocity < 170 cm/s), 69 had abnormally high TCD risk (time-averaged mean of the maximum velocity > 200 cm/s) while 13 patients had stroke history. Presence of  $\alpha$ -thalassemia 3.7 deletion was found in 65 of 155 patients (41.9 %) and was associated with a decreased risk of abnormal TCD velocity [odds ratio (OR) 0.40, 95 % confidence interval (CI) 0.20 - 0.80,  $p = 0.009$ ]. The prevalence of *G6PD* A<sup>-</sup> variant was 9.03% and was not associated with abnormal TCD or stroke. Hence, our results show the modifying outcome of  $\alpha$ -thalassemia deletion in reducing the risk of abnormal TCD velocity in Nigerian SCD children.

## PgmNr 1383: A new global *Slc7a7* knockout mouse models human lysinuric protein intolerance.

### Authors:

B.M. Stroup<sup>1</sup>; R. Marom<sup>1</sup>; J.D. Heaney<sup>1</sup>; I. Grafe<sup>1</sup>; X. Li<sup>1</sup>; B. Dawson<sup>1</sup>; Y. Chen<sup>1</sup>; J.R. Green<sup>1</sup>; A.E. Christiansen<sup>2</sup>; M.E. Dickinson<sup>2</sup>; J.R. Seavitt<sup>1</sup>; A.L. Beaudet<sup>1</sup>; B.H. Lee<sup>1</sup>; L.C. Burrage<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 2) Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX.

Lysinuric protein intolerance (LPI) is an inborn error of metabolism, characterized by urea cycle dysfunction, growth failure, renal disease, immune dysfunction and osteoporosis. LPI is caused by biallelic pathogenic variants in *SLC7A7*, which encodes the light subunit of the y<sup>+</sup>L amino acid transporter 1 (y<sup>+</sup>LAT1) required for intestinal absorption and renal reabsorption of arginine, lysine and ornithine (cationic amino acids). Perinatal lethality of the only *Slc7a7*<sup>-/-</sup> mouse model has hindered *in vivo* studies investigating the etiology of LPI complications.

Our objective was to develop a new mouse model of LPI for mechanistic studies of LPI pathology. We developed a *Slc7a7*<sup>-/-</sup> mouse model using CRISPR/Cas9 technology (mixed 129/C57BL/6 F2 background), which has a deletion of exons 3 and 4 resulting in a frameshift mutation. Due to the reduced survival of male and female *Slc7a7*<sup>-/-</sup> mice compared to wild type (WT) littermates, tissues were harvested at 14-18 days of age. At this time point, *Slc7a7*<sup>-/-</sup> mice demonstrated growth failure (body weight (g), means±SD; WT, 13.0±1.2; *Slc7a7*<sup>-/-</sup>, 3.6±1.2; *p*<0.0001). In addition, assessment of plasma and urine amino acid concentrations showed that *Slc7a7*<sup>-/-</sup> mice had reduced plasma concentrations and increased urinary excretion of the cationic amino acids compared to WT mice, which is consistent with the biochemical phenotype of LPI. Furthermore, urine amino acid analysis revealed generalized aminoaciduria in *Slc7a7*<sup>-/-</sup> mice. This generalized aminoaciduria in combination with a loss of brush border and increased lipid vacuolation and secondary lysosomes in the proximal tubules, based on electron microscopy and hematoxylin and eosin (H&E) staining, suggests proximal tubular dysfunction in *Slc7a7*<sup>-/-</sup> mice. H&E stained spleen sections showed disorganization of splenic structures, including expansion of the red pulp, in *Slc7a7*<sup>-/-</sup> mice. Moreover, micro-CT analysis in the L<sub>4</sub> vertebrae showed a 36% reduction in bone mass (trabecular bone volume/total volume) in *Slc7a7*<sup>-/-</sup> versus WT mice.

In summary, the *Slc7a7*<sup>-/-</sup> mouse model had the biochemical phenotype of LPI and demonstrated other characteristic phenotypes, including growth failure, renal disease and osteoporosis. Although this mouse model of LPI facilitates investigation of disease mechanisms during development, an adult mouse model of LPI is needed for mechanistic studies in adulthood and investigations of potential therapies.



## PgmNr 1384: A *Drosophila* model for Coffin-Siris syndrome.

**Authors:**

R.A. MacPherson<sup>1,2</sup>; R.R.H. Anholt<sup>1,2</sup>; T.F.C. Mackay<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Center for Human Genetics, Clemson University, 114 Gregor Mendel Circle, Greenwood, South Carolina, 29646.; 2) Department of Genetics and Biochemistry, Clemson University, 114 Gregor Mendel Circle, Greenwood, South Carolina, 29646

---

Coffin-Siris syndrome is a rare genetic disorder which may include developmental delays, intellectual disability, hypotonia, facial dysmorphology and digit abnormalities. Several genes that have been associated with Coffin-Siris syndrome play a role in chromatin modification. The fruit fly *Drosophila melanogaster* presents a powerful genetic model system for investigating human genetic diseases that result from disruption of evolutionarily conserved processes. Through the use of mutants in *Drosophila* orthologues of human candidate disease genes and transgenic expression in flies of human disease associated alleles, we can identify co-regulated transcriptional networks associated with elements of chromatin modification that are relevant to Coffin-Siris syndrome. We model Coffin-Siris syndrome through targeted RNAi knockdown using the binary *GAL4-UAS* expression system in controlled genetic backgrounds. Subsequent whole fly RNA-sequencing on the Illumina Novaseq platform combined with chromatin accessibility analyses using ATAC-seq will provide new insights in the pathogenesis of Coffin-Siris Syndrome.

## PgmNr 1385: Validating patient-centered phenotyping tools for use in rare disease diagnostics.

### Authors:

I.A. Holm<sup>1</sup>; C.A. Brownstein<sup>1</sup>; K. Shefchek<sup>2</sup>; C. Lee<sup>1</sup>; H. Paterson<sup>1</sup>; J.A. McMurry<sup>2</sup>; S. Chennuri<sup>3</sup>; N. Gujral<sup>3</sup>; S.I. Ziniel<sup>4</sup>; J.S. Brownstein<sup>3</sup>; M.A. Haendel<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts.; 2) Monarch Initiative, Oregon State University, Corvallis, Oregon; 3) Innovation and Digital Health Accelerator, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts.; 4) Children's Hospital Colorado, University of Colorado Denver School of Medicine, Aurora, Colorado

---

The Human Phenotype Ontology (HPO) encodes phenotypes for rare disease diagnostics. Patients have a wealth of knowledge about their condition, including features not captured clinically. Our hypothesis is that patient self-phenotyping will provide an accurate source of HPO terms. Two approaches allow patients to generate HPO terms. The first is a traditional survey developed by GenomeConnect, the ClinGen patient registry. The second involves direct translation of the HPO terms into “layperson” terms. We developed an anatomy based-term selection application, Phenotypr, to allow patients to record the layperson HPO terms. Here, we report on synthetic validation and rare disease patient evaluation using the two self-phenotyping tools.

We generated synthetic profiles for 7,336 known rare diseases using each of the tools, permuted them with noise and omissions, and compared against gold standard annotations using the Monarch Initiative’s algorithms. We found that roughly half of the 7,701 layperson profiles performed exactly as well for diagnosis as their gold-standard counterparts, representing the theoretical maximum utility achievable by profiles created by the average patient. This number dropped to 448 when we omitted profile phenotypes that did not have a corresponding question in the GenomeConnect survey.

We then tested both tools in patients with known genetic diseases to determine which tool resulted in phenotype profiles that were more comprehensive and specific to determine diagnostic efficacy. About 250 adult participants diagnosed with, or who were the parent/guardian of a child diagnosed with, a known rare genetic disease enrolled and were randomly assigned to the GenomeConnect or Phenotypr tool; a subset completed both. Participants were from the Boston Children’s Hospital Genetics clinic, Down Syndrome program, or Manton Center for Orphan Disease Research; GenomeConnect; or a 16p13.11 research registry. We also conducted qualitative interviews of a subset of participants to determine which modality was preferable. Early results suggest that while patients preferred the GenomeConnect survey, better phenotype profiles were generated with Phenotypr. This suggests that a hybrid approach that provides familiar tooling but access to richer HPO terms may be warranted. We conclude that such tools could be used to improve and accelerate diagnostic pipelines and promote collaboration and patient engagement with clinical caregivers and diagnosticians.

## PgmNr 1386: New content and tools in the PhenX Toolkit.

### Authors:

M. Phillips<sup>1</sup>; C.C. Morton<sup>2</sup>; B. Peter<sup>3</sup>; M.L. Rice<sup>4</sup>; M.L. Marazita<sup>5</sup>; B. Konkle<sup>6</sup>; S. Pipe<sup>7</sup>; M. Bosquet Enlow<sup>8</sup>; B. Entwisle<sup>9</sup>; A. Fernandez<sup>10</sup>; J. Schoden<sup>1</sup>; L. Cox<sup>1</sup>; J. Beverly<sup>1</sup>; W. Huggins<sup>1</sup>; L. Gridley<sup>1</sup>; T. Hendershot<sup>1</sup>; D. Maiese<sup>1</sup>; A. Riley<sup>1</sup>; H. Pan<sup>1</sup>; M.C. Krzyzanowski<sup>1</sup>; S. Hwang<sup>1</sup>; N. Pino<sup>11</sup>; E.M. Ramos<sup>11</sup>; C.M. Hamilton<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) RTI International, Research Triangle Park, NC; 2) Brigham and Women's Hospital and Harvard Medical School, Boston, MA; and University of Manchester, UK; 3) Arizona State University, Tempe, AZ; 4) University of Kansas, Lawrence, KS; 5) University of Pittsburgh, Pittsburgh, PA; 6) Bloodworks Northwest, Seattle, WA; 7) University of Michigan, Ann Arbor, MI; 8) Boston Children's Hospital and Harvard Medical School, Boston, MA; 9) University of North Carolina at Chapel Hill, Chapel Hill, NC; 10) University of California, San Francisco, San Francisco, CA; 11) National Human Genome Research Institute, Bethesda, MD

---

The PhenX (**Ph**enotypes and **ex**posures) Toolkit <https://www.phenxtoolkit.org/> is a web-based catalog of recommended measurement protocols and associated bioinformatics tools to facilitate cross-study analyses. These protocols are recommended by domain experts using a consensus-based process which relies on community input. The PhenX Toolkit currently includes more than 700 measurement protocols from 25 research domains and 20 collections and has more than 2,600 registered users. An established resource that continues to expand content for investigators, the PhenX Toolkit includes a new pediatric development research domain and updates to protocols in the speech, language, and hearing domain. Depth was added in the areas of hemophilia inhibitors and social determinants of health. One Working Group was assembled to identify 15 priority measures to address Pediatric Development and to enhance existing PhenX measures with content for children. Speech, language, and hearing experts reviewed the measurement protocols and recommended additions or changes to 11 of the 22 protocols. Additions included tablet/phone applications for hearing and language assessment and new voice technologies. The Hemophilia Inhibitor Research Working Group recommended 17 measures and 38 protocols that determine levels of Factor VIII and Factor IX, quantify inhibitors to Factors VIII and IX, measure the biological response to therapeutics, and capture key outcomes. A Social Determinants of Health Working Group recommended measures to capture conditions of where people live, work, and play as they affect a wide range of health outcomes. Separately, as a pilot study, PhenX protocols and variables were analyzed for their cross-compatibility with other data dictionaries and the interoperability of data collected using PhenX protocols. New tutorials help investigators search, browse and manage their toolkits of chosen measures and the PhenX Toolkit's ontology-based search tools and browsing allows users to find related measures across domains and links to other ontologies such as Human Phenotype Ontology (HPO).

## PgmNr 1387: Severe hemophilia A and inhibitor development in Brazilian admixed patients.

### Authors:

L. Zuccherato <sup>1,2</sup>; S. M.E. Santos <sup>1</sup>; L. Jardim <sup>3</sup>; R. Camelo <sup>3</sup>; L. Moura <sup>4</sup>; D. Chaves <sup>4</sup>; S. Rezende <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Complementary Propaedeutic Department, Faculty of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; 2) Centre of Research and Education, Instituto Mario Penna, Belo Horizonte, Minas Gerais, Brazil; 3) Department of Internal Medicine, Faculty of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; 4) Fundação Hemominas, Belo Horizonte, Minas Gerais, Brazil

---

Hemophilia A (HA) is an X-linked disorder due to deleterious mutations in the factor VIII (FVIII) gene (*F8*). A recurrent inversion, originated by a homologous recombination event within the *F8* intron 22 and its extragenic inversely oriented copies (Inv22-1 and Inv22-2), accounts for about 45% of severe HA worldwide. The inversion of intron 1 (Inv1) is the second recurrent *F8* disruption, with prevalence of 4.5% among severe patients. The most clinically relevant complication of hemophilia A is the development of neutralizing alloantibodies (inhibitors) against FVIII, occurring in up to 30% of the severe patients. Environmental and genetic components are known risk factors for inhibitor development, including the molecular architecture of *F8* (large deletions, inversions and nonsense variants), as well as ethnicity, family history of inhibitors and, maybe, the type of the infused FVIII concentrate.

To explore the molecular features related to inhibitor development, we performed a customized Illumina NGS for *F8* exome sequencing and genotyping of two *CTLA4* risk variants against FVIII alloantibodies (rs23177 and rs3087243) in an admixed Brazilian cohort of 116 patients with severe (FVIII:C < 1%) HA. Inv22 and Inv1 were detected with inverse shifting and allelic specific PCRs. Patients were participants from two different studies - the HEMFIL Study (n = 67) and the BrazIT Study (n = 49). Patients were sorted as inhibitor positive (Inh+) based on the detection of inhibitor (at least two inhibitor tests with titer >0.6 Bethesda Units [BU]/mL) and negative (Inh-) when inhibitor titers were <0.6 (BU)/mL after 75 exposure days (ED) to FVIII. Most patients (73/116; 62.9%) were Inh+. The *F8* exome sequencing revealed 52 deleterious mutations, including previously reported and new variants. As expected, there was an association between *F8* mutation and inhibitor status (p = 0.022), with missense and frameshift mutations increased in Inh- and Inh+ groups, respectively (p = 0.001). Conversely, allelic frequency analysis showed no difference of rs23177 and rs3087243 of between Inh+ (n=44) and Inh- (n=43) patients (p = 0.500 and 0.806, respectively). To our knowledge, this is most complete overview of *F8* mutation of unrelated HA patients in Latin American. Considering the multidimensional features of inhibitor development in HA, we highlight the relevance of studying diverse/admixed populations at genetic level and when predicting biological markers.

## PgmNr 1388: GWAS arrays as potential sources of secondary genomic findings: Implications for return of results.

### Authors:

A.R. Bentley <sup>1</sup>; J.J. Johnston <sup>2</sup>; A. Adeyemo <sup>1</sup>; L.G. Biesecker <sup>2</sup>; C.N. Rotimi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Research in Genomics and Global Health, Natl Human Genome Research Inst., Bethesda, Maryland.; 2) Medical Genomics and Metabolic Genetics Branch, Natl Human Genome Research Inst, Bethesda, Maryland

---

In clinical sequencing, informing participants of secondary genomic findings (SGF) is the standard of care and the American College of Medical Genetics and Genomics (ACMG) has a list of 59 genes for return of SGF. There is less consensus regarding SGF in the research setting. In the context of large, epidemiological cohorts, GWAS is a widely used analysis tool, for which genome-wide genotyping array data are generally used. However, because modern GWAS arrays include exonic content, they may interrogate variants in the ACMG genes. Given the larger number of research participants that have undergone genotyping than sequencing, researchers using genotyping data may have access to a significant source of SGF. We evaluated variants included on the Illumina Multi-Ethnic Genotyping Array (MEGA) that may confer high risk for SGF-associated diseases. The manifest of the MEGA was intersected with the ACMG list of 59 genes returning 4,310 variants. Variants were annotated using Alamut Batch (Interactive Biosoftware) and filtered for MAF in gnomAD, coding effect, and pathogenicity classification in ClinVar. A total of 1,608 coding variants passed our frequency filters and were either assessed as pathogenic or likely pathogenic by at least one ClinVar submitter or were putative loss of function alleles in genes where loss of function is a known mechanism of disease. While, in the context of clinical sequencing, the low population frequency of these variants ( $\leq 0.005$  in all gnomAD populations) supports their potential pathogenicity, in the context of GWAS, for which the MEGA Array data is often used, these variants are likely to be filtered out over concerns of low power. Thus, evaluating the presence of these variants must use data before applying frequency filters. These data suggest that ACMG variants can be identified from array-based genotyping that could have significant health-care implications and potential benefit for research participants. Therefore, the widely held notion that GWAS SNP array results are irrelevant to SGF is no longer true, though positive predictive value must be evaluated. Researchers and IRBs should consider whether individuals participating in SNP array-based studies and found to be carrying such variants should be identified and what follow up steps (clinical validation, return of results, and/or referral) are indicated.

## PgmNr 1389: Optical mapping of the schizophrenia-associated 3q29 deletion reveals new features of genomic architecture.

### Authors:

T. Mosley <sup>1</sup>; D.J. Cutler <sup>1</sup>; M.E. Zwick <sup>1</sup>; R. Johnston <sup>1</sup>; J.G. Mulle <sup>1,2</sup>; The Emory 3q29 Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 2) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA

---

**Introduction:** 3q29 deletion syndrome is a rare genomic disorder (~1 in 30,000) caused by a typically *de novo* deletion spanning ~1.6 Mb and 21 genes. It is associated with neurodevelopmental and neuropsychiatric phenotypes, including a 40-fold increase in risk for schizophrenia. The 3q29 locus contains three low-copy repeats (LCRs). Two flank the canonical 3q29 deletion, LCR B (centromeric side) and LCR C (telomeric side). The third, LCR A, is 140 kb centromeric to LCR B. The presence of these LCRs suggests the region undergoes non-allelic homologous recombination (NAHR) to form the 3q29 deletion and reciprocal 3q29 duplication. However, our understanding of the local architecture at the 3q29 locus is limited.

**Methods:** We performed Bionano<sup>®</sup> Saphyr<sup>™</sup> genome mapping on 12 3q29 deletion trios and 3 additional deletion probands. We evaluated the genomic architecture at the 3q29 locus of probands and parents, characterized the breakpoints of the deletion in probands, and identified additional structural variants in the region, including inversions, in all samples.

**Results:** Deletion sizes range between 1.68-1.7 Mb. In ~60% of probands, deletion breakpoints are within LCRs B and C. For the remaining 40%, breakpoints lie on the telomeric side of LCR C. The interval in 2 probands span into *RUBCN*, a gene not previously implicated in 3q29 deletion syndrome. In parental genomes, we observe expansions at LCRs A and B ranging between 60-130 kb in size: 57% of alleles at LCR A and 54% of alleles at LCR B have expansions compared to the hg38 reference genome. We have identified common inversions at the locus, particularly between LCR A and B, but these do not appear to be related to deletion formation.

**Conclusion:** This is the first fine scale mapping of the 3q29 deletion interval in multiple trios. The observation that the deletion includes *RUBCN* in some cases may impact the phenotypes of 3q29 individuals. Expansions in LCRs A and B indicate variation at the 3q29 locus not captured by human genome build 38. The expansions may be critical factors that underlie the formation of the 3q29 deletion.

**Implications:** Fine scale mapping of the genomic architecture and identifying risk factors contributing to structural instability in the 3q29 region can be extrapolated to other genomic disorder loci. Subtle variation in deletion size and deletion of genes outside the canonical interval are not captured by clinical microarrays and may explain heterogeneity of the 3q29 deletion phenotype.

## PgmNr 1390: Mondo Disease Ontology: Harmonizing disease concepts across the world.

### Authors:

J.A. McMurry<sup>1,3</sup>; N. Vasilevsky<sup>2,3</sup>; S. Essaid<sup>2,3</sup>; N.L. Harris<sup>3,4</sup>; M.A. Haendel<sup>1,2,3</sup>; P.N. Robinson<sup>3,5</sup>; C.J. Mungall<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oregon State University; 2) Oregon Health and Science University; 3) Monarch Initiative; 4) Lawrence Berkeley National Laboratory; 5) Jackson Laboratory

---

HGVS nomenclature is the standard for reporting variants, but there is no comparable standard for reporting diseases. Existing sources of disease definitions and data models include HPO, OMIM, SNOMED CT, ICD, PhenoDB, MedDRA, MedGen, ORDO, DO, GARD etc; however, these sources partially overlap and sometimes conflict, making it difficult to know definitively how they relate to each other. This has resulted in a proliferation of mappings between disease entries in different resources; however mappings are problematic: collectively, they are expensive to create and maintain. Most importantly, the mappings lack completeness, accuracy, precision; as a result, mapping calls are often inconsistent between resources. The UMLS provides intermediate concepts through which other resources can be mapped, but these mappings suffer from the same challenges: they are not guaranteed to be one-to-one, especially in areas with evolving disease concepts such as rare disease.

In order to address the lack of a unified disease terminology that provides precise equivalences between disease concepts, we created the Mondo Disease Ontology that provides a logic-based structure for unifying multiple disease resources. Its development is coordinated with the Human Phenotype Ontology (HPO), which describes the individual phenotypic features that define a disease concept. Like the HPO, Mondo provides a hierarchical structure which can be used to annotate data at different levels of precision. Mondo provides mappings to other disease resources, but in contrast to other mapping sets, Mondo precisely annotates each mapping using strict semantics, so that we know when two diseases are precisely equivalent or merely closely related.

Incorporating Mondo into variant prioritization tools such as Exomiser could improve their efficiency and accuracy, by reducing redundancy and inconsistencies between the many disease terminologies in use.

Mondo is used to accommodate a global alignment of disease concepts - e.g., subsumption of classification and axioms by The Experimental Factor Ontology (EFO) to assist annotation of disease information where EFO is used elsewhere - e.g., EBI resources, Open Targets, and Euro-Bioimaging. In addition, Mondo is being utilized in diverse applications and resources such as ClinGen, and Gabriella Miller Kids First. We invite the community to contribute to Mondo; visit [github.com/monarch-initiative/mondo](https://github.com/monarch-initiative/mondo) for details.



## **PgmNr 1391: Implementing clinically validated automated genomic variant prioritisation for whole genomes, exomes and virtual panels, with diagnostic performance that equals human experts.**

### **Authors:**

L. Burnett <sup>1,2,3,4</sup>; T. Ohnesorg <sup>2</sup>; M. Hinchcliffe <sup>2</sup>; L. Wedd <sup>2</sup>; P. Ajuyah <sup>2</sup>; K. Ross <sup>2</sup>; M. Hobbs <sup>1,2</sup>; A. Mallawaarachchi <sup>1,2</sup>; E. Richardson <sup>2</sup>; E. Lee <sup>2,4</sup>; N. Schonrock <sup>2</sup>; A. Statham <sup>4</sup>; B. Lundie <sup>4</sup>; A. Poulet <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Darlinghurst, Sydney, New South Wales, Australia; 2) Genome.One, Australian Clinical Laboratories, Darlinghurst, Sydney, New South Wales, Australia; 3) St. Vincent's Clinical School, UNSW Sydney, Darlinghurst, Sydney, New South Wales, Australia; 4) Genome.One, Garvan Institute of Medical Research, Darlinghurst, Sydney, New South Wales, Australia; 5) Institut des Sciences Pharmaceutiques et Biologiques - Faculté de Pharmacie de Lyon, Lyon France

---

**Background:** Analysis and interpretation of genomes (and exomes) are major challenges requiring bioinformatics and professional resources in limited supply. We have developed an automated variant prioritisation framework, whose performance equals genomic diagnosis using human experts.

**Methods:** Genome.One operates a NATA-accredited (ISO15189) whole genome diagnostic service in Sydney, Australia. A reference set of ~500 clinical referrals underwent bioinformatics using a precisionFDA Award-winning pipeline (Kinghorn Centre for Clinical Genomics). Tertiary bioinformatics used published tools (Seave), with every case analysed by Doctoral genetics professionals. Additionally, a team of genetics experts performed variant review of selected non-trivial cases. We evaluated two automated pipelines: *Moon* (diploid.com), *Eclipse* (derived from Seave), and a combination of the two. Reference set concordance was required for all ACMG Class 4 and 5 variants, and desirable (but not required) for Class 3 variants. Discordant results were adjudicated by the variant review expert team. Following initial validation, prospective parallel analyses was undertaken by both manual and automated protocols.

**Results:** For whole exome analysis performed by whole genome sequencing, automated pipelines performed nearly as well as human experts, identifying at least 88-98% of Class 4/5 and 60-75% of Class 3 variants. Applying both in combination equalled human experts, with 100% concordance for all Class 4/5 reportable variants and 75-81% concordance for Class 3 variants. The new approach resulted in order-of-magnitude reduction (median reduced from 225 to 6 variants for whole genomes) in interpretation time. We have now extended this approach to include virtual panel analysis.

**Conclusion:** An automated variant prioritisation protocol has been developed, whose performance equals that of credentialed genetic diagnostic professionals. The pipeline is accredited for clinical diagnosis, for both whole genome analysis as well as now for virtual panel analysis.



## PgmNr 1392: Pathogenic variants of *EPHB4* cause lymphatic anomalies through over-activation of mTOR.

### Authors:

M.R. Battig<sup>1</sup>; M.E. March<sup>1</sup>; L.S. Matsuoka<sup>1</sup>; L. Finoti<sup>1</sup>; C. Seiler<sup>2</sup>; J.T. Bennett<sup>3,4</sup>; I.A. Glass<sup>4</sup>; T.L. Wenger<sup>5</sup>; D. Li<sup>1</sup>; H. Hakonarson<sup>1,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Zebrafish Core Facility, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; 4) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, WA; 5) Division of Craniofacial Medicine, Seattle Children's Hospital, Seattle, WA; 6) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 7) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Central conducting lymphatic anomaly (CCLA) is a rare and progressive disorder in which lymph is inadequately cleared as a result of dilated lymphatic channels, dysmotility, and/or distal lymphatic obstruction. We identified a number of pathogenic mutations in *EPHB4*, a gene encoding ephrin type-B receptor 4 and involved in vascular development, through whole exome sequencing of patients diagnosed with lymphatic anomalies, including two missense mutations (c.2288G>A:p.R763Q and c.2654A>G:p.K885R) and a splicing mutation that leads to the use of a cryptic splice donor and the retention of the intervening 12-bp intron sequence and an in-frame insertion of four amino acids (c.2334+1G>C:p.L778\_G779insLMLG). Functional characterization using an *in vitro* spheroid-sprouting assay in human lymphatic endothelial cells showed these mutations cause a loss of function of *EPHB4*, and lead to unregulated lymphangiogenic sprouting and development. The functional consequences of these mutations were confirmed in a zebrafish model. Injection of *ephb4a* morpholino into zebrafish resulted in lymphatic vessel misbranching and developmental deformities that mimicked the lymphatic presentation observed in the patients. Further analyses revealed that over-activation of mTOR, as a consequence of *EPHB4* loss-of-function, led to the observed phenotype. Importantly, the phenotype could be rescued in both the spheroid and zebrafish models upon treatment with rapamycin and the mTOR dual inhibitor OSI-027. These findings demonstrate that these loss-of-function variants of *EPHB4* are implicated in CCLA, and that mTOR inhibitors may have therapeutic benefits in patients with lymphatic anomalies and other vascular disorders resulting from mutations that induce mTOR pathway upregulation.

## PgmNr 1393: A genocentric, optimization-based approach to discovery of Mendelian disease genes.

### Authors:

A.W. Hansen <sup>1,2</sup>; M. Murugan <sup>1</sup>; H. Li <sup>1</sup>; M.M. Khayat <sup>1,2</sup>; L. Wang <sup>1</sup>; J. Rosenfeld <sup>2</sup>; B.K. Andrews <sup>1</sup>; S. Jhangiani <sup>1</sup>; Z. Coban Akdemir <sup>2</sup>; F. Sedlazeck <sup>1,2</sup>; D. Muzny <sup>1,2</sup>; E.E. Davis <sup>6</sup>; N. Katsanis <sup>6</sup>; A. Sabo <sup>1,2</sup>; J.E. Posey <sup>2</sup>; Y. Yang <sup>2</sup>; M.F. Wangler <sup>2</sup>; C.M. Eng <sup>2</sup>; V.R. Sutton <sup>2,4</sup>; J.R. Lupski <sup>1,2,3,4</sup>; E. Boerwinkle <sup>2,5</sup>; R.A. Gibbs <sup>1,2</sup>; Task Force for Neonatal Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas.; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030; 4) Texas Children's Hospital, Houston, TX 77030; 5) School of Public Health, UTHealth, Houston, TX 77030; 6) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC 27701

The advent of inexpensive, clinical exome sequencing (ES) has led to the accumulation of genetic data from thousands of samples from individuals affected with a wide range of diseases, but for whom the underlying genetic and molecular etiology of their clinical phenotype remains unknown. In many cases, detailed phenotypes are unavailable or poorly recorded and there is little family history to guide study. To accelerate discovery, we integrated ES data from 18,696 individuals referred for suspected Mendelian disease, together with relatives, in an Apache Hadoop data lake and implemented a genocentric analysis which rapidly identified 154 candidate Mendelian disease genes.

The approach did not rely on case-specific phenotypic classifications, but iteratively queried the data with variable gene-level (i.e. pLI) and variant-level (i.e. REVEL) filtering criteria revealing genes where at least five cases harbor ultra-rare (MAF < 1/10k), potentially pathogenic variants absent from all controls. Optimal filtering values were identified which maximize the ratio of output genes with recently-reported Mendelian disease-gene associations. Past discovery within a given gene set was consistently shown to correlate with continued discovery, supporting this optimization-based strategy. In the months following our analysis, Mendelian disease associations for 19 of the 154 candidates were independently reported to OMIM (expected = 2.29;  $p < 0.00001$ ;  $n=100,000$  permutations): *ATP1A1*, *CACNA1E*, *CHD3*, *CLTC*, *DOCK3*, *FBXO11*, *IRF2BPL*, *KDM5B*, *KIAA1109*, *LINGO1*, *MACF1*, *MAST1*, *MYO9A*, *PDE1C*, *SCN3A*, *SET*, *TBX2*, *TCF20* and *VPS13D*.

We sought replication by intersecting the candidate disease genes with a set of genes harboring *de novo* mutations in an independent cohort of ES trios with a wide range of congenital anomalies. Six genes overlapped between the set of 154 candidate disease genes and the 216 genes harboring *de novo* nonsynonymous variants (expected = 1.68;  $p = 0.005$ ;  $n = 10,000$  permutations): *AATK*, *CELSR1*, *IRF2BPL*, *MYO5C*, *ROCK1*, and *UBC*. No candidates overlapped with the set of 93 genes harboring noncoding variants, supporting the model that *de novo* nonsynonymous variants are much more likely to be pathogenic than other *de novo* variants.

Thus, we demonstrate the utility of genocentric and optimization-based approaches toward

accelerating Mendelian discovery, and provide a high-quality set of novel candidate disease genes (and variants) for ongoing discovery efforts.

## PgmNr 1394: Extensive allelic, genetic, and phenotypic heterogeneity identified in exome study of primary lymphedema.

### Authors:

E.C. Lawrence <sup>1</sup>; E.E. Blue <sup>2</sup>; D.A. Nickerson <sup>3</sup>; M.J. Bamshad <sup>4</sup>; R.E. Ferrell <sup>1</sup>; D.N. Finegold <sup>1</sup>; The University of Washington Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, WA

The Pittsburgh Lymphedema Family Study (PLFS) is a cohort of over 300 families with hereditary lymphedema phenotypes, including a subset of individuals diagnosed with lymphedema-distichiasis syndrome (LDS). Causal variants in *FOXC2* were previously identified in this subgroup, though not all LDS probands revealed a *FOXC2* mutation.

For 27 members of 20 multiplex families previously screened [using early whole exome sequencing (WES) technology without a molecular diagnosis], we performed WES using an updated exome capture kit, 75bp paired-end reads, high coverage (mean = 60 to 80X), and an expanded analytical pipeline. Rare variants were extracted with a maximum alternate allele frequency of 0.5% predicted to have moderate to high impact consistent with the mode of inheritance predicted by patient pedigree data. We prioritized conserved variants, those predicted to be pathogenic, and those in genes previously implicated in lymphedema in humans or related phenotypes in animal models. The observed phenotype in 2/20 LDS families could be explained by frameshift variants in *FOXC2* not detected in the original analysis. To identify novel genotype-phenotype relationships, we expanded our WES analysis to include 94 probands from PLFS families along with co-segregation studies within families by Sanger sequencing. Among 22/114 families (19%), the phenotype may be explained by variants in genes causing dominant forms of lymphedema: *CELSR1* (4/114), *FLT4* (12/114), *FOXC2* (4/114), *GATA2* (1/114), and *GJC2* (1/114). An additional 10 families may be explained by variants in genes previously associated with lymphedema in humans with strong functional or animal model support: *FN1* (1/114), *HGF* (3/114), *KDR* (2/114), *NRP2* (1/114), *SOX17* (3/114), and *VCAM1* (1/114). Including genes necessary for lymphangiogenesis or animal models exhibiting lymphedema, we identify potentially causal variants in an additional 18/114 (16%) of families: *CALCR* (1/114), *CLEC1B* (1/114), *EMILIN1* (4/114), *ITGA9* (1/114), *PLXNA3* (3/114), *PROX1* (1/114), *SPRED2* (1/114), *TIE1* (5/114), and *VEZFI* (1/114).

We expand the known genotype - phenotype relationship in several genes and add support for novel genes in primary lymphedema. Current next-generation sequencing technology applied to previously analyzed phenotypes yields a wealth of novel variation. Resequencing of older samples yields important relevant new variation. These results add to the extensive genetic and allelic heterogeneity in primary lymphedema.

## **PgmNr 1395: The NHGRI Centers for Mendelian Genomics: Accelerating novel discovery in Mendelian disease.**

### **Authors:**

L. Chadwick; T. Stephan; C. Wellington

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Human Genome Research Institute, Bethesda, MD.

---

The National Human Genome Research Institute (NHGRI)'s Centers for Mendelian Genomics (CMG) aim to discover the genetic basis of as many Mendelian traits as possible. The CMGs accelerate discovery by dissemination of tools, and approaches for Mendelian genomic analysis, collaboration with disease-focused investigators, and providing a range of educational opportunities. This poster will highlight the Centers for Mendelian Genomics contributions to the field, describe how data and findings from the CMG are shared with the community, and discuss some of the gaps and challenges that remain to be addressed.

## PgmNr 1396: A pipeline for genome-wide genotyping, filtering, and prioritization of tandem repeats.

### Authors:

N. Mousavi<sup>1</sup>; M. Maksimov<sup>1</sup>; N. Pusarla<sup>1</sup>; R. Yanicky<sup>1</sup>; S. Shleizer-Burko<sup>1</sup>; D. Dimmock<sup>3</sup>; M. Bainbridge<sup>3</sup>; J. Friedman<sup>1,2</sup>; J. Gleeson<sup>1,2</sup>; M. Gymrek<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of California San Diego, La Jolla, CA; 2) Rady Children's Hospital San Diego (RCHSD), San Diego, CA; 3) Rady Children's Institute for Genomic Medicine (RCIGM), San Diego, CA

---

The application of clinical next generation sequencing (NGS) has resulted in dozens of successful diagnoses for a wide range of disorders. However, several cases of clinical NGS remain unsolved. One potential explanation is that many genetic disorders may be due to complex mutations such as repeats not captured by standard analysis pipelines. Tandem Repeats (TRs) are regions comprised of short motifs of 1-20bp repeated in tandem. TRs have been associated with dozens of Mendelian diseases, such as Huntington's disease and Fragile X Syndrome. Beyond these well-known loci, additional pathogenic TRs continue to be discovered. Despite their established role in disease, repeat variants are typically missing from clinical NGS pipelines due to technical challenges of analyzing TRs from short reads. We present an end-to-end pipeline for genome-wide identification of pathogenic TR expansions involved in Mendelian disease. Our pipeline leverages GangSTR, which incorporates multiple sources of information from paired-end reads into a joint likelihood model capable of genotyping both short and expanded TRs from short reads. Our pipeline consists of the following steps: (1) genome-wide TR genotyping (GangSTR); (2) quality control and filtering (dumpSTR); (3) prioritization of candidate TR expansions based on gene annotation and comparison to unaffected control populations and available unaffected family members (postmaSTR). We validated our pipeline by analyzing 30x PCR-free whole genome sequencing from individuals with known pathogenic repeat expansions at the Huntington's Disease locus. For 12/13 individuals, genome-wide analysis correctly identified the premutation or pathogenic CAG repeat expansion. We then applied our pipeline to 41 undiagnosed pediatric cases from the Undiagnosed Neurological Disease Program at Rady Children's Institute for Genomic Medicine. Our pipeline was completed in an average of 3 hours per sample running in a cloud environment using virtual machines with 8 cores each. We identified an average of 100 candidate TRs genome-wide and 10 candidate TRs in coding or UTR regions per family. The top candidate list for several families included variants in or near a gene previously implicated in a neuro-developmental disorder. Our pipeline for genome-wide genotyping, filtering, and prioritization of TRs is publicly available on Github or as a Docker container and can be easily deployed on cloud environments such as DNA Nexus or Amazon Web Services.

## PgmNr 1397: Large-scale replication and discovery of genetic associations for rare diseases with self-reported medical data.

### Authors:

S. Shringarpure <sup>1</sup>; W. Wang <sup>1</sup>; B. Cameron <sup>1</sup>; D. Dhamija <sup>1</sup>; A. Auton <sup>1</sup>; R. Gentleman <sup>2</sup>; 23andMe Research Team

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Research, 23andMe, Mountain View, California.; 2) Therapeutics, 23andMe, Mountain View, California.

---

Though nearly 7000 diseases affecting fewer than 0.05% of the population are known, a large number of rare diseases still remain unsolved. State-of-the-art approaches for discovering the genetic basis of rare diseases rely on sequencing cases or case families for a single disease, followed by automated and manual curation to discover likely causal variants. This severely limits the rate at which the genetics of rare disease can be studied and understood.

Here we demonstrate the use of SNP array data combined with self-reported phenotype data to study rare diseases at a large scale and identify genetic associations for rare diseases through genome-wide association studies (GWAS). Using web-based questionnaires, we gathered self-reported data on rare diseases from a cohort of over 3.6 million research-consented DTC-genotyped individuals. We ran GWAS on over 250 rare disease phenotypes with a minimum of 30 cases in individuals of European ancestry.

We replicate known associations at a genome-wide significance level ( $p\text{-value} < 5e-8$ ) with a diverse range of minor allele frequencies (minimum MAF=0.0012, maximum MAF=0.4989) and effect sizes (minimum OR=1.16, maximum OR=207.02). We find common variant associations for Hirschsprung's disease (*RET*), cleft palate (*GSDMC*), Fuch's corneal dystrophy (*TCF4*), hemochromatosis (*HFE*), craniosynostosis (*BMP2*), factor V deficiency (*F5*). In addition, we find rare variant associations for Huntington's disease (*HTT*), cystic fibrosis (*CFTR*), alpha-1 antitrypsin deficiency (*SERPINA1*). In some rare diseases, we find associations tagging known causal genes, for example, Lynch syndrome (2p16.3, tagging *MSH6*) and epidermolysis bullosa simplex (12q13.13, tagging *KRT5*). For polycythemia vera, we replicate known associations in the *JAK2*, *TET2*, and *TERT* genes. For Von Willebrand disease, we replicate known associations in the *VWF* and *ABO* genes. We also find a novel association for Von Willebrand disease in the *USP7* gene, which has been previously associated with platelet count.

Our results show that large-scale online collection of self-reported data is a viable method for large-scale discovery and replication of disease gene associations for rare diseases. Further analyses will enable the discovery of more novel genetic associations for increasingly rare diseases.

## PgmNr 1398: Aromatic amino acid decarboxylase deficiency in 17 Mainland China patients: Clinical phenotype, molecular spectrum and therapy overview.

### Authors:

D. Qian

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pediatric Endocrinology and Genetic Metabolism, Shanghai Institute for Pediatric Research, Xinhua Hospital, School of Medicine, Shanghai Jiao Tong U, Shanghai, China, Weiqian, Dai, MS

---

**Objectives:** The prevalence of aromatic amino acid decarboxylase deficiency (AADCD) is relatively high in certain Asian populations, due to a founder effect. However, only three cases have been reported in Mainland China. The purpose of this study is to broaden the clinical, molecular and treatment spectrum of AADC deficiency. **Methods:** We recruited 14 previously undescribed patients, who received a diagnosis of AADC deficiency, together with three reportedly cases in Mainland China. Full clinical information was collected, and disease-causing mutations in the DDC gene were detected by High-throughput sequencing. **Results:** The common clinical manifestation of patients, including intermittent oculogyric crises, retarded movement development, and autonomic symptoms. A patient showed bone-density loss which have not been reported so far. There are twelve of the 14 patients who showed severe phenotype besides two mildly affected children. The majority of patients showed a poor response to drug therapy, but two mildly phenotype patients improved psychomotor function after being prescribed medication. The most common genotype of Mainland Chinese AADCD is the splice-site pathogenic variant (IVS6 + 4A > T; c.714 + 4A > T), which accounts for 58.8% of all DDC mutations, followed by c.1234C > T (p.R412W; 17.6%). Three novel compound heterozygous mutations, c.565G>T, c.170T>C, and c.1021+1G>A, was firstly reported. Besides, we discovered that patients may presented with mild or moderate if found to be compound heterozygote or homozygote for one of the following missense or nonsense mutations: c.478C>G, c.853C>T, c.1123C>T, c.387G>A, c.665T>C. **Discussion:** The clinical data of the cohort of 17 patients in Mainland China broaden the clinical, molecular and treatment spectrum of AADC deficiency.



## PgmNr 1399: Prediction of lethality and severity of osteogenesis imperfecta variants in the triple-helix regions of COL1A1 and COL1A2.

### Authors:

V. Pejaver <sup>1,2</sup>; C. Bock <sup>3,4</sup>; D. Beltran <sup>5</sup>; S.D. Mooney <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA; 2) The eScience Institute, University of Washington, Seattle, WA, USA; 3) Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland; 4) SIB Swiss Institute of Bioinformatics, Switzerland; 5) Invitae Corporation, San Francisco, CA, USA

---

**Background:** Osteogenesis imperfecta (OI) is a group of connective tissue syndromes primarily characterized by a lifetime liability to fractures. The classical form of OI is typically caused by autosomal dominant mutations in type I collagen-coding genes (*COL1A1* and *COL1A2*), particularly the substitution of glycine residues in the triple helical (TH) domain. OI is phenotypically heterogeneous with four common types: mild (Type I), perinatally lethal (Type II), progressively deforming (Type III), and moderately deforming (Type IV). However, a straightforward link between variant characteristics and OI types has proven elusive, complicating the interpretation of novel variants. Here, we address the assignment of OI types to missense variants in the TH domain by developing machine learning models for two related tasks: lethality prediction (Type II vs. the rest) and severity prediction (Types II and III vs. the rest). **Methods:** We investigated whether existing pathogenicity and functional effect predictors could be repurposed for these tasks. We also evaluated the performance of two methods more suited to this application: an OI-specific lethality predictor and a gene/disease-independent severity predictor. We then trained multi-layer neural network models on 891 variants obtained from multiple data sources, using features related to the variants' sequence context and position. Predictive performance was assessed through cross-validation experiments using standard evaluation metrics. **Results:** Pathogenicity, functional effect and severity predictors yielded poor (near-random) performance for both tasks. While the existing OI-specific lethality predictor fared better, it was outperformed by our models. Interestingly, although the "lethality" and "severity" models were generally correlated with each other, the scores from the former correlated better with the severity of the four OI types, which is a task that it was not explicitly trained for. **Conclusions:** Despite the use of a basic set of features, our COL1A-specific model performed better than more sophisticated methods when predicting OI severity. Our comparison of the lethality and severity prediction tasks also has implications for the development of methods that go beyond simplistic notions of "pathogenic" or "benign" by demonstrating that consistent definitions of phenotype severity and problem formulations across different genes/diseases are important for the development of generalizable methods.

## PgmNr 1400: Clinically associated genomic mutations database for heritable epidermolysis bullosa disease.

### Authors:

G. Kumar <sup>1</sup>; A.H. Saeidian <sup>2,3</sup>; A. Ertel <sup>1</sup>; L. Youssefian <sup>2,3</sup>; H. Vahidnezhad <sup>2</sup>; J. Uitto <sup>1,4</sup>; P. Fortina <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sidney Kimmel Cancer Center, Philadelphia, Pennsylvania.; 2) Department of Dermatology and Cutaneous Biology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, USA; 3) Genetics, Genomics and Cancer Biology PhD Program, Thomas Jefferson University, Philadelphia, PA, USA; 4) Jefferson Institute of Molecular Medicine, Thomas Jefferson University, Philadelphia, PA, USA

---

Genetic variant classification is critical for an accurate genetic diagnosis. However, for rare inherited disorders, it poses a major challenge since post-genomic era the number of genetic variants and complexity associated with assigned clinical impact has been increased exponentially. To address this challenge in epidermolysis bullosa (EB), a clinically and genetically heterogeneous group of rare inherited skin fragility disorders, we developed the Database of Epidermolysis Bullosa (EBDB). This comprehensive resource integrates all genetic variants in 21 currently known EB-associated genes from multiple data sources, including 1000 Genome, Exome Sequencing Project, Exome Aggregation Consortium, ClinVar, dbSNP, GME, Iranome and GNOMAD. All the genetic variants are re-categorized into single classification according to ACMG guidelines first computationally (using a custom built variant categorization pipeline) and then validated using manual curation. EBDB clearly improves the knowledge of existing clinically associated variant classification in the EB disease. As an example, *COL7A1*- a gene associated with dystrophic EB has a total of 7704 coding variants, where ClinVar classifies only 3% of the variants into different categories of variant effect including Pathogenic (P), Likely Pathogenic (LP), Variant of Uncertain Significance (VUS), Likely Benign (LB), or Benign (B). The remaining 97% genetic variants are not classified in any category. Using our custom built variant categorization pipeline which utilizes minor allele frequency (MAF) and functional prediction algorithm scores to re-categorize the variants, we successfully re-classified the *COL7A1* variants into 0.9% - P, 0.1% - LP, 7% - LB, 29% - B, and 63% - VUS. This re-classification of each variant was confirmed using manual curation. Using this approach, EBDB is successfully resulting in an improved classification of more than 95% of coding variants where pathogenicity is driven by minor allele frequency thresholds, variant effect and protein domain. In conclusion, we have implemented a computational variant classification algorithm to develop a database encompassing genomic variants re-categorized according to the ACMG standards to facilitate the application of comprehensive genetic testing to the clinical care of EB patients.

## PgmNr 1401: Variation in alpha diversity of the oral microbiome based on periodontal disease status, NHANES 2009-2012.

### Authors:

J. McLean <sup>1</sup>; S. Lukacs <sup>1</sup>; E. Vogtmann <sup>2</sup>; A. Chaturvedi <sup>3</sup>; N. Freedman <sup>2</sup>; C. Abnet <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Health and Nutrition Examination Surveys, National Center for Health Statistics, Centers for Disease Control and Prevention, Hyattsville, Maryland.; 2) Metabolic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 3) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD

---

The National Health and Nutrition Examination Survey (NHANES) is a multistage nationally representative probability sample of the multi-ethnic, civilian, non-institutionalized US population. NHANES assesses the health and nutritional status of adults and children annually through personal interview, standardized examinations, and laboratory analysis of collected biospecimens. The v4 region of the 16S rRNA gene was sequenced from oral rinse samples collected from NHANES participants ages 14-69 years in NHANES cycles 2009-2010 and 2011-2012 (n =10,466) to characterize the oral microbiome. The resulting sequence data was processed via two pipelines: closed reference operational taxonomic unit (OTU) clustering and DADA2 for sequence variant generation. Resulting data include the following: alpha-diversity metrics (i.e., richness and evenness of microbes), beta-diversity matrices, principal coordinates of the beta-diversity matrices, and relative/absolute abundance tables for each taxonomic level from phylum to genus for each pipeline. Alpha diversity was calculated for data rarefied to 10,000 reads using three metrics: observed OTUs, Faith's Phylogenetic Diversity, and the Shannon-Weiner index. Participants with data from the oral health periodontal (30 years and older) and oral microbiome components were utilized (n=5,519). Periodontal disease (PD) status (PD vs. no PD) was calculated using CDC/American Academy of Periodontology case definitions. We examined the association between PD status and alpha diversity data by sex, age group, poverty status, self-reported race/ethnicity, education, smoking status, and marital status. We used pairwise *t*-tests to compare alpha diversity estimates by PD status overall and within all subgroups. Participants with PD had significantly higher alpha diversity than those without PD (p-value < 2 x10<sup>-11</sup>). This was observed across all three metrics in each pipeline. For demographic characteristics all subgroups had statistically higher alpha diversity for those with PD vs those without PD, the greatest differences were in the following: age group 35-49 (p-value < 1 x10<sup>-12</sup>) and never smoker (p-value < 6 x10<sup>-10</sup>) across all metrics in each pipeline. Further analysis of beta diversity and the taxonomic abundances will allow characterization of the community and individual microbial differences based on PD status at the US population level.

## PgmNr 1402: Federating association analysis in type 2 diabetes to protect participants' privacy.

### Authors:

J. Spalding <sup>1</sup>; N. Burt <sup>2</sup>; M. Costanzo <sup>2</sup>; M. McCarthy <sup>3</sup>; T. Keane <sup>1</sup>; AMP-T2D Knowledge Base and Portal Team

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) European Genome-phenome Archive, EMBL EBI, CAMBRIDGE, United Kingdom; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom

---

By increasing the number of genomes from subjects with type 2 diabetes (T2D) for analysis, we can increase the number of detected T2D genetic associations, leading to the discovery of functional pathways and development of novel therapeutic targets. The Accelerating Medicines Partnership in T2D (AMP-T2D) is a multi-sector, pre-competitive partnership among government, industry, and nonprofit organizations. AMP-T2D has built an infrastructure consisting of nodes for data storage and analysis (knowledge base), and presentation (knowledge portal). Many countries are now developing personalised health programmes, resulting in an increasing number of genomes being made available for secondary research purposes. Personal health information is subject to national laws, and may not be available for analysis in a global context. By separating the knowledge base from the knowledge portal a network of knowledge bases can be developed, which allows analyses to be configured on the knowledge portal, and distributed to nodes located within the jurisdiction where the data was generated. Analyses can be run on the federated nodes, and return anonymised results to the knowledge portal for interrogation and display. The AMP-T2D Knowledge Portal (<http://www.type2diabetesgenetics.org>) is an open access resource allowing analysis of data federated between the USA and UK.

Data are stored in a Data Coordinating Center (DCC) at the Broad Institute and also at the European Bioinformatics Institute (EBI). As of June 2019 the EBI knowledge base (KB) was hosting 8 datasets, which can be queried directly from the knowledge portal along with the datasets hosted by the DCC. The knowledge portal supports custom online genetic association analysis with dynamic sample filtering and covariate specification. A genetic risk score module has been added allowing calculation of the genetic risk score for a range of phenotypes based on a set of variants with association to T2D, and interactive manhattan plots are available to display genetic associations across the genome. To facilitate federated analysis the analysis engine is run inside Docker containers, which are accessible via a REST API. Phenotypes and sample attributes are harmonized across the network by utilising a registry. Access to the REST API is restricted to authorised users, with the authentication and authorization procedure compatible with Global Alliance for Genomics and Health (GA4GH) standards to maximise interoperability.

## PgmNr 1403: Addressing the missing data issue in multi-phenotype genome-wide association studies.

### Authors:

M. Anasanti <sup>1</sup>; M. Kaakinen <sup>1,2</sup>; M.R. Jarvelin <sup>3,4,5,6,7</sup>; I. Prokopenko <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genomics and Common Disease, Imperial College London, London, United Kingdom; 2) Department of Clinical and Experimental Medicine, University of Surrey, Guildford, United Kingdom; 3) Department of Epidemiology and Biostatistics, Imperial College London, United Kingdom; 4) Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, United Kingdom; 5) Center for Life Course Health Research, University of Oulu, Oulu, Finland; 6) Unit of Primary Care, Oulu University Hospital, Oulu, Finland; 7) Biocenter Oulu, University of Oulu, Oulu, Finland

---

Joint analysis of multiple phenotypes in genome-wide association studies (MP-GWAS) increases power for locus discovery but suffers from missingness in phenotype values. We investigated properties of missing data imputation methods within MP-GWAS, focussing on single and multiple-imputation (SI/MI) using Bayesian approach and expectation-maximisation bootstrapping (EMB), k-nearest neighbour (kNN), left-censored imputation method (QRILC) and random forest (RF). We simulated genetic data for 5,000/50,000/500,000 individuals using Hapgen2, and highly ( $r=0.64$ ) and moderately correlated ( $r=0.33$ ) phenotypes (3/9/30/120) for them. We randomly selected common, low-frequency and rare variants to be significantly ( $P < 5 \times 10^{-8}$ ) associated with the simulated phenotypes. We considered several proportions of missing data (1/5/20/50%) under missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). We used the Root Mean Squared Error (RMSE) for the evaluations with complete cases (CC) vs. full data RMSE as the reference level. RF and MI-EMB diverged the most from the reference under MCAR ( $P_{RF}=7.18 \times 10^{-4}$ ,  $P_{MI-EMB}=6.27 \times 10^{-4}$ ). RF also outperformed under MAR ( $P=6.22 \times 10^{-4}$ ), whereas QRILC outperformed under MNAR ( $P=7.45 \times 10^{-5}$ ). RF was applied to the Northern Finland Birth Cohorts (NFBC) 1966 and 1986 (4,955 and 2,687 individuals, respectively) for the imputation of anthropometric and glycaemic measurements and 149 serum metabolite levels. MP-GWAS of 31 amino acids showed a novel association at *ADAMTS* after imputation with RF ( $P=2.61 \times 10^{-11}$  vs.  $P=5.68 \times 10^{-7}$  in CC) and improved power at *FCGR3B* ( $P=1.86 \times 10^{-9}$  vs.  $P=1.72 \times 10^{-8}$ ). We propose improved solutions for phenotype imputation in high-dimensional omics data-analyses and have implemented these into a user-friendly and computationally efficient imputeSCOPA software tool.

## PgmNr 1404: In silico analysis of coding/non-coding SNPs of human *RETN* gene and characterization of their impact on resistin stability and structure.

### Authors:

I. Morjane <sup>1,2</sup>; L. Elkhatabi <sup>1,2</sup>; H. Charoute <sup>2</sup>; S. Amghar <sup>2</sup>; H. Bouafi <sup>1,2</sup>; Z. Elkarhat <sup>2</sup>; R. Saile <sup>1</sup>; H. Rouba <sup>2</sup>; A. Barakat <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratoire De Biologie et Santé, Faculté des Sciences Ben M'Sik, Université Hassan II, Casablanca, Morocco; 2) Laboratoire de Génomique et Génétique Humaine, Institut Pasteur du Maroc, Casablanca, Morocco

---

Resistin (*RETN*) is a gene coding for pro-inflammatory adipokine called resistin secreted by macrophages in humans. Single nucleotide polymorphisms (SNP) in *RETN* are linked to obesity and insulin resistance in various populations.

Using dbSNP, 78 non synonymous SNPs (nsSNPs) were retrieved and tested on PredictSNP 1.0 megaserver. Among these, 15 nsSNPs were predicted as highly deleterious and thus subjected to further analyses, such as conservation, post-transcriptional modifications and stability.

The 3D structure of human resistin was generated by homology modeling using Swiss Model. Root Mean Square Deviation (RMSD), hydrogen bonds (h-bonds) and interactions were estimated.

Furthermore, UTRscan served to identify UTR functional SNPs.

Among the 15 most deleterious nsSNPs, 13 were predicted to be highly conserved including variants in post-translational modification sites. Stability analysis predicted 9 nsSNPs (I32S, C51Y, G58E, G58R, C78S, G79C, W98C, C103G and C104Y) which can decrease protein stability with at least three out of the four algorithms used in this study.

These nsSNPs were chosen for structural analysis. Both variants C51Y and C104Y showed the highest RMS deviations (1.137 Å and 1.308 Å respectively) which were confirmed by the important decrease in total h-bonds. The analysis of hydrophobic and hydrophilic interactions showed important differences between the native protein and the 9 mutants, particularly I32S, G79C and C104Y. Six SNPs in 3'UTR (rs920569876, rs74176247, rs1447199134, rs943234785, rs76346269 and rs78048640) were predicted to be implicated in polyadenylation signal.

This study revealed 9 highly deleterious SNPs located in human *RETN* gene coding region and 6 SNPs within 3'UTR that may alter the protein structure. Interestingly, these SNPs are worth to be analysed in functional studies to further elucidate their effect on metabolic phenotypes occurrence.

## PgmNr 1405: Data-driven analysis of omics and electronic health records from the Estonian Biobank identifies early changes associated with incident type 2 diabetes.

### Authors:

J. Kronberg-Guzman <sup>1</sup>; K. Läll <sup>1,2</sup>; N. Taba <sup>1</sup>; A. Kalnapenkis <sup>1</sup>; T. Haller <sup>1</sup>; K. Fischer <sup>1,2</sup>; A. Metspalu <sup>1</sup>; T. Esko <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia; 2) Institute of Mathematics and Statistics, University of Tartu, Estonia; 3) Medical and Population Genetics, Broad Institute, USA

---

High-throughput data has been used extensively to study the differences between healthy and diseased individuals. However, it is even more important to understand the earliest changes preceding the development of a disease, paving the way for possible prevention. Genetic risk scores can be used to summarise the genetic component, but risk also depends on other factors, such as age, lifestyle and external exposures.

We use NMR metabolomics data from the Estonian Biobank, coupled with the analysis of electronic health records. From 152,000 biobank participants, 10,840 have <sup>1</sup>H-NMR metabolomics data available. From these, 328 had developed incident type 2 diabetes by the time of linking of health records. Machine learning identified 33 metabolic markers predictive of incident type 2 diabetes. We used systems biology for the interpretation of these predictive metabolites, specifically, we created a similarity network between all metabolites. Additionally, Olink proteomics data and genetic risk score for type 2 diabetes was integrated with the same network.

Network analysis shows that 67.6% of the first-neighbour proteins of the predictive metabolites have already been shown to be involved in type 2 diabetes or related phenotypes. Specifically, we find that the set of first neighbours is enriched in cytokine signalling. When using STRING to see whether these proteins have any interactions, we identify a highly connected network of 9 proteins, including 6 ligands for chemokine receptors. The genetic risk score for type 2 diabetes is directly connected with glucose levels in the network, and in close proximity of  $\beta$ -hydroxybutyrate, lactate and acetate. Our data-driven analysis has shown the potential to use omics data from a population-based biobank and electronic health records for gaining insight into a complex disease. More specifically, by using only data to inform the analysis, we have recovered information already known about some aspects of the early changes in the development of type 2 diabetes.

Our next steps will include the develop of a predictive model using age, gender, predictive metabolites and genetic risk score, and validation of the findings in an independent cohort.



## PgmNr 1406: Explanatory knowledge graph analysis of GWAS and MWAS associations with obesity reveals biological interpretations of statistical results.

### Authors:

C. Bizon <sup>1</sup>; K. Bradford <sup>1</sup>; E. Morris <sup>1</sup>; M. Graff <sup>2</sup>; Y. Wang <sup>2</sup>; K.E. North <sup>2,3</sup>; P. Gordon-Larsen <sup>4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC

GWAS analysis discovers statistical associations between genetic variation and phenotypes; these associations require interpretation to determine the mechanistic role of the identified variation. One approach to probing biological underpinnings is to simultaneously collect a broader spectrum of data, such as RNA expression, methylation, or metabolomics, and conduct data driven integrative analyses. A second approach is to integrate derived associations with prior knowledge including e.g. gene functions, pathways, and metabolic reactions. To address the paucity of methods in the field, we have developed a Knowledge Graph framework approach to uncover the mechanistic role of genetic variants that combines both data driven and known pathways.

The ROBOKOP Knowledge Graph (RKG), developed as part of the NCATS Data Translator program and hosted at <http://robokopkg.renci.org>, is a collection of half a million nodes representing entities such as diseases, phenotypes, genes, chemicals, and sequence variants. The over 12 million relationships between these entities represent the integration of many open knowledge sources and bio-ontologies. We have augmented RKG through the integration of the GWAS catalog, GTEx eQTLs, and calculated Linkage Disequilibrium (LD), to enhance its use as a source of information in an integrative analysis.

We tested our platform using genetic and metabolomic data from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), a multi-center epidemiologic study of the factors playing a protective or harmful role in the health of Hispanics/Latinos. Using GWAS- and MWAS-associations with phenotypes, we used RKG to identify associations and pathways relative to literature-based prior knowledge. Explanatory queries can be issued against this combined graph, such as “find all GWAS variants near genes that metabolize MWAS-significant blood metabolites”, which implicates genes such as *AHR*, *TAT*, and *CASP1*. These queries produce scientifically coherent sub-networks containing a combination of new results and prior knowledge, and potentially explaining the disparate statistical results.

This work was funded by UNC-Chapel Hill’s Obesity Creativity Hub and the National Center for Advancing Translational Sciences, National Institutes of Health, through the Biomedical Data Translator Program (award number OT2TR002514).



# PgmNr 1407: Machine learning based histology phenotyping to investigate epidemiologic and genetic basis of adipocyte morphology and cardiometabolic traits.

## Authors:

C.A. Glastonbury<sup>1,2</sup>; S.L. Pulit<sup>1</sup>; J. Honecker<sup>3</sup>; J.C. Censin<sup>4</sup>; S. Laber<sup>1</sup>; H. Yaghootkar<sup>5,13</sup>; N. Rahmioglu<sup>4,8</sup>; E. Pastel<sup>5</sup>; K. Kos<sup>5</sup>; A. Pitt<sup>6</sup>; M. Hudson<sup>6</sup>; C. Nellåker<sup>1</sup>; N.L. Beer<sup>7</sup>; H. Hauner<sup>12</sup>; C.M. Becker<sup>8</sup>; K.T. Zondervan<sup>4,8</sup>; T.M. Frayling<sup>5,10</sup>; M. Claussnitzer<sup>3,9,11</sup>; C.M. Lindgren<sup>1,4,11</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Big Data Institute, University of Oxford, Oxford, United Kingdom; 2) BenevolentAI, London; 3) University of Hohenheim, Germany; 4) Wellcome Trust Centre for Human Genetics (WTCHG), Oxford, UK; 5) Genetics of Complex Traits, University of Exeter Medical School, Royal Devon & Exeter Hospital, Exeter, EX2 5DW, UK.; 6) Research Centre for Optimal Health, School of Life Sciences, University of Westminster, London, UK; 7) Novo Nordisk Research Centre Oxford (NNRCO), Oxford UK; 8) Endometriosis CaRe Centre Oxford, Nuffield Department of Women's and Reproductive Health, University of Oxford, UK; 9) Beth Israel Deaconess Medical Center, Harvard Medical School, Boston MA; 10) NIHR Clinical research facility, University of Exeter Medical School, University of Exeter, Exeter, UK; 11) Broad Institute of MIT and Harvard, Cambridge MA; 12) Chair of Nutritional Medicine, Else Kröner-Fresenius-Center for Nutritional Medicine, TUM School of Life Sciences Weihenstephan, Technische Universität München; 13) Research Centre for Optimal Health, School of Life Sciences, University of Westminster, London, UK.

**INTRODUCTION:** Although obesity is a known cardiometabolic risk factor, little is known about how genetic variation influences human adipocyte size across adipose depots or how such variability may confer risk to obesity and other cardiometabolic outcomes.

**OBJECTIVE:** To establish a link between genetic variation and adipocyte cell size by performing the first large-scale study of automatic adipocyte phenotyping with a novel method using both histology and genetic data.

**METHOD:** We developed an ML method, the Adipocyte U-Net, to rapidly estimate adipocyte area estimates from histology. The Adipocyte U-net is a Convolutional Neural Network (CNN) based on the U-net architecture. Here, we use a combination of a fine-tuned InceptionV3 network to classify histology tiles and our Adipocyte U-net to perform counting and binary segmentation of adipocytes. We validate our method using three SOTA approaches (Adiposoft, CellProfiler and floating adipocytes), run blindly on 2 external cohorts. We observe strong concordance between our method and traditional approaches (Adipocyte U-net vs CellProfiler:  $R^2_{\text{visceral}}=0.94$ ,  $P<2.2\times10^{-16}$ ,  $R^2_{\text{subcutaneous}}=0.91$ ,  $P<2.2\times10^{-16}$ ), and tractable run times (10k images: 6mins vs 3.5hrs).

**RESULTS:** We applied our Adipocyte U-Net to 4 cohorts with histology, genetic, and phenotypic data (N=820). After meta-analysis, we find that adipocyte area positively correlated with Body Mass Index ( $P_{\text{subcutaneous}}=8.13\times10^{-69}$ ;  $P_{\text{visceral}}=2.5\times10^{-55}$ ; average  $R^2=0.49$ ) and recapitulate the finding that adipocytes in subcutaneous depots are larger than their visceral counterparts ( $P_{\text{meta}}=9.8\times10^{-7}$ ). A random effects meta-analysis demonstrated no significant association between adipocyte size and T2D, independent of BMI, age and sex ( $P_{\text{subcutaneous}}=0.11$ ,  $P_{\text{visceral}}=0.37$ ), suggesting BMI is the primary

modifier of adipocyte size. Next, we observe sexually-dimorphic adipocyte size between adipocyte depots, where female visceral adipocytes are smaller on average than male visceral adipocytes ( $P_{\text{meta}}=3\times 10^{-7}$ ). Lastly, we performed the first GWAS and subsequent meta-analysis of adipocyte area and intra-individual adipocyte variation.

**CONCLUSION:** Despite having 2× the number of samples than any similar study, we find no genome-wide significant associations. This suggests that both larger and more homogenous collections of adipose depots derived from the same anatomical location will be required to identify robust genetic associations that validate previous findings.

## PgmNr 1408: Computational software for peak alignment and quantification of comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry metabolomics data.

### Authors:

M.Z. Kos <sup>1</sup>; A. Blackburn <sup>2</sup>; P. Stevens <sup>1</sup>; K. Truax <sup>1</sup>; L. Blondell <sup>1</sup>; H.H.H. Goring <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, The University of Texas Rio Grande Valley School of Medicine, San Antonio, Texas.; 2) Blackburn Statistics, LLC, San Antonio, Texas.

---

High-dimensional metabolomics profiling is an informative approach to characterize the metabolic state of tissues and bodily fluids. Comprehensive two dimensional gas chromatography time-of-flight mass spectrometry (GC×GC-TOFMS) is a sophisticated and sensitive technology that is well suited for separating complex biological samples. Here we present new software for processing and aligning peaks from large GC×GC-TOFMS data sets from the LECO Pegasus 4D instrument. The pipeline: (1) calculates and applies retention time adjustments based on spiked-in internal standards for all deconvoluted peaks; (2) compares mass spectra between proximate peaks within a sliding retention time window by performing pairwise correlation analysis; (3) aligns peaks within and across samples using a chosen correlation coefficient threshold; and (4) characterizes aligned peaks by outputting total ion count areas, among other statistics. To evaluate the software, we conducted a total of 75 runs (3 vials, 25 runs each at two different concentrations) of standards from EPA Method 8270 and processed the resulting metabolomics data from ChromaTOF with our software. After alignment (at Pearson  $r > 0.9$ ), 268 peaks were identified across all runs, of which 66 were observed in more than 95% of the runs and 146 were singletons. Of the 76 compounds in the standard mix, 69 (91%) were detected among the 66 robustly observed peaks; missed compounds included semi-volatiles with reactive nitro groups and both small and large compounds that fell outside the data acquisition window (470 to 3540 s). After standardization against total peak areas and log transformation, the robust peaks showed an average coefficient of variation of 5% (range from 1% to 18%), with increased variability observed at low 1D retention times (<600 s). Pairwise correlations yielded high  $r$  values between runs (0.63 to 0.995), which was reflected by tight clustering in PCA results. REML modeling attributed the vast majority of the total variation to variation between aligned peaks (93%), with non-significant amounts attributed to differences arising from the independent runs and vial source (both <1%). Thus, our software appears well suited for processing complex GC×GC metabolomics data, exhibiting limited variation in aligned peak areas of known compounds across runs. The software is now being used in the analysis of metabolomics data generated on >1800 blood plasma samples for identifying diabetes and obesity biomarkers.

## PgmNr 1409: A taxonomy-aware method for background noise correction increases power to detect associations.

### Authors:

L. Briscoe <sup>1,2</sup>; B. Balliu <sup>3</sup>; L. Shenhav <sup>4</sup>; S. Sankararaman <sup>1,2,3,4</sup>; E. Halperin <sup>1,2,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA.;  
 2) Department of Human Genetics, University of California Los Angeles, Los Angeles, CA; 3)  
 Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA; 4)  
 Department of Computer Science, University of California Los Angeles, Los Angeles, CA; 5)  
 Department of Anesthesiology, University of California Los Angeles, Los Angeles, CA

---

In recent years, several studies have demonstrated associations between the human microbiome and diseases, such as obesity and diabetes. Yet, due in part to small sample sizes, results from individual studies are often inconsistent and replication among studies is poor. Combining datasets from different studies can overcome the lack of power but is complicated by the introduction of technical variability, i.e. batch effects, between datasets due to differences in sequencing protocols, experimental design, time of collection, etc. Ignoring these effects can lead to both false positive associations as well as loss of power in downstream analyses. Quantile normalization is the current method specifically designed for batch correction in microbiome data even though this method has been shown in other omics fields, e.g. epigenomics and transcriptomics, to remove biological variation and reduce power to detect associations. We propose an empirical Bayes framework for adjusting microbiome data for batch effects which takes into account the count nature and taxonomic hierarchy present in microbiome data. Our method extends ComBat, a method for batch correction in microarray gene expression studies, by partitioning the method of moments estimators based on taxonomy and learning taxa-specific batch parameters. Using real 16S data from six microbiome studies of obesity, we establish that technical differences between datasets have varying effects on different phyla justifying the need for a taxonomy-aware method. We find that the observed abundance of individual taxa as a function of coverage increases faster for some studies than others, resulting in considerable bias. We show that our method adjusts for batch effects, outperforming existing batch correction methods as measured using the KBET score and by fitting a linear mixed model to estimate sources of variation due to study label compared to disease status. Interestingly, the study label or batch of samples explains up to 0.45 of the variance in the uncorrected data, and 0.01 in the data corrected by our method. The ratio of variance explained by disease status to variance explained by study label also increases after applying our method from 0.10 in the uncorrected data to 1.94. Lastly, we show that our improved batch correction finds stronger associations between obesity status and composition, based on both testing of the entire microbiome composition and testing of individual taxa.

# PgmNr 1410: A deep learning model for accurate variant calling in congenital adrenal hyperplasia.

## Authors:

S. Hong <sup>1</sup>; A. Ganesh <sup>2</sup>; K.A. Beauchamp <sup>1</sup>; D. Muzzey <sup>1</sup>; K.R. Haas <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Myriad Genetics, South San Francisco, CA.; 2) Stanford University, Stanford, CA

---

## Introduction

Congenital adrenal hyperplasia (CAH) is an autosomal recessive disease that impairs steroidogenesis. Mutations in *CYP21A2* account for a large fraction of CAH cases. *CYP21A2* and a pseudogene *CYP21A1P* have high sequence identity. Variant calling in *CYP21A2* is technically challenging due to frequent and complex gene rearrangements with *CYP21A1P*. We developed an enhanced deep learning model ("deepCAH") which introduces additional features and class labels to improve CAH variant calling.

## Methods and Results

Our original model for *CYP21A2* variant calling utilized two sets of features: i) allele-balance measurements summarizing the fraction and count of genic and pseudogenic alleles of interest, and ii) copy-number measurements consisting of normalized sequencing depths at differentiable loci on *CYP21A2* and *CYP21A1P*. Allele balance for *CYP21A2* variants is the aggregate of benign and deleterious reads counts paired end-specific to *CYP21A2* and *CYP21A1P*. Due to frequent sequence rearrangement, the fidelity of read contribution is often assessed during manual call review by examining paired end-specific read contribution. To provide the model with all information available to human review, we extended the model to include novel features---namely, paired end-specific read contributions. In addition to 11 variants, the extended model included a new variant indicating a deleterious mutation (Q319X) is in cis with gene duplication. A cohort of >50,000 samples was split into training (80%) and validation sets (20%). To improve sensitivity on rare variants, additional positive samples were generated by data augmentation. We employed a multi-layer recurrent neural network comprised of long short-term memory (LSTM) cells and a weighted cross-entropy loss function. The model was implemented in TensorFlow and trained using the Adam optimizer. Considering the human-reviewed calls as ground truth, the extended LSTM showed a validation error rate of < 1 in 500.

## Conclusions

The enhanced deep learning model, deepCAH, achieved high accuracy (>99.9%) for technically challenging CAH variant calling. The deepCAH caller is expected to significantly reduce call review burden as it can substitute secondary confirmation by another call reviewer.

## PgmNr 1411: Multi-omic analysis of discordant and concordant sib-pairs with inflammatory bowel disease.

### Authors:

A. Stiemke <sup>1</sup>; S. Brandt <sup>2,3</sup>; C. Simpson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Genomics and Informatics, University of Tennessee Health and Science Center, Memphis, Tennessee.; 2) Department of Medicine, Department of Genetics and the Human Genetics Institute of New Jersey, Piscataway, New Jersey.; 3) Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, School of Medicine and Department of Epidemiology, School of Public Health, Johns Hopkins University, Baltimore, Maryland.

---

Inflammatory bowel disease (IBD) is an immune-mediated chronic intestinal disorder that is typically divided into two distinct types; ulcerative colitis (UC) and Crohn's disease (CD). Over 240 IBD susceptibility loci have been identified, however, much of the etiology remains unexplained and, even within families, the disease can have a heterogeneous clinical presentation. Here we present a study attempting to discern the cause of that heterogeneity in discordant and concordant sib-pairs by comparing whole exome sequencing, epigenetic differences using DNA methylation analysis, and gene expression using next generation RNA sequencing (RNAseq).

A total of 96 discordant and concordant sib-pairs were included in the study. Genomic DNA and RNA were extracted from peripheral blood mononuclear cells that were isolated using cell sorting from whole blood. DNA methylation values were measured using the Illumina EPIC array. Whole exome sequencing was performed using the Illumina TruSeq Exome kit, and RNAseq was performed using the Illumina RNA Kit v2. Of 5000 initial variants, five genes— Notch Receptor 1 (NOTCH1); Serine/Arginine Repetitive Matrix 2 (SRRM2); Zinc Finger Protein 276 (ZNF276); GATA Zinc Finger Domain Containing 2A (GATAD2A); and Diphthamide Biosynthesis 1 (DPH1)—remain after cross-matching preliminary exome and methylation analyses. RNAseq analysis is currently underway and results will be presented.

Combining data from multiple technologies is an important next step for interpreting the results of genome-wide association studies (GWAS) and there is currently no agreement on the best approach for integrating these data. Given the known genetic and clinical heterogeneity in IBD, using discordant and concordant sib-pairs attempts to leverage the expected sharing between the sibs as supporting information to inform future studies.

## PgmNr 1412: Machine learning-based urinary screening for lysosomal disease.

### Authors:

O. Migita <sup>1,2</sup>; H. Uryu <sup>2,3</sup>; H. Yamamoto <sup>1</sup>; N. Shimizu <sup>1</sup>; K. Hata <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, St. Marianna University, School of Medicine, Kawasaki, Kanagawa, Japan;  
2) Department of Maternal-Fetal biology, National Center Institute for Child Health and Development, Tokyo, Japan; 3) Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill School of Medicine, NC, USA

---

Fabry disease, one of the most frequently observed lysosomal diseases, is an inborn error of metabolism with severe symptoms. Enzyme replacement therapy and pharmacological chaperones can stabilize the progression in Fabry disease but cannot reverse organ damage. However, time delays in the diagnosis and treatment often occur, because non-specific symptoms including pain can be the early symptoms of Fabry diseases. Physicians can find low enzyme activities, high biomarker level and specific Fabry mutations to confirm the diagnosis, but the prior the those biochemical tests, they needs some clue for diagnosis. Recently, it has been reported that the prevalence of Fabry disease is considerably higher than believed.

For addressing the diagnostic delay problem, we collected urine sediment images from Fabry patients and non-Fabry patients as training data sets for machine learning. The developed our models successfully distinguished the Fabry urinary images from the other patients' images. The AUCs on the dataset were 0.994 (0.987–0.998) for the existence of mulberry cells in each segment, with high-sensitivity and high-specificity. We expect this screening tools for Fabry disease to contribute to early disease diagnosis and intervention, near future.

## PgmNr 1413: BD-CoCoLasso: Alternating block descent penalized regression adapted for the presence of a block of inaccurate features.

### Authors:

C. Escribe <sup>1,4</sup>; J. Keller-Baruch <sup>1</sup>; J.B. Richards <sup>1,3</sup>; K. Oualkacha <sup>2</sup>; C.M.T. Greenwood <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 2) Department of Mathematics, Université du Québec à Montréal, Montreal, Quebec, Canada; 3) McGill University, Montreal, Quebec, Canada; 4) Ecole Polytechnique, Palaiseau, France

---

**Introduction:** Sparse regression has achieved great success when modelling phenotypes as a function of high-dimensional feature sets. Although the standard formulation of a prediction problem assumes clean data, real world data is often corrupted—either partially missing or measured with error. Datta and Zou (2017) developed the CoCoLasso algorithm, an adaptation of the LASSO approach capable of handling a general class of corrupted datasets and achieving better prediction performance. Here, we adapt CoCoLasso to a class of problems where there are both corrupted (e.g. exposure information) and uncorrupted covariates (e.g. SNP data). The motivation of our work is to reduce the computational needs of this new penalization method when only a small percentage of the features are corrupted.

**Methods:** CoCoLASSO requires a computationally demanding positive semi-definite projection of the covariance matrix for a high dimensional feature set. In our context when there are corrupted and uncorrupted covariates, we take advantage of the block descent minimization trick to develop a more efficient algorithm. In an alternating block minimization algorithm, the CoCoLasso corrections are used when updating corrupted coefficient vectors, and a simple LASSO is used for the uncorrupted coefficient vectors. Both subproblems are convex and hence a global solution can be obtained, even though adaption of the cross-validation step requires care in this setting where there are products of corrupted and uncorrupted matrices.

**Results:** In simulations, our method proved efficient both when we generated additive errors, and when we varied proportions of missing data. The algorithm was tested with either sample size (n) or feature set dimension (p) as large as 1e4, assuming only a small proportion of the features were corrupted. Our algorithm decreases the bias on the coefficient vector compared to a simple LASSO algorithm, while it limits the number of false positives. Computational times are reduced relative to full CoCoLasso.

**Discussion:** This work has promising applications. When trying to predict complex traits such as osteoporosis, it may be interesting to mix genetic features with lifestyle features. Although the former, after imputation, can be considered as almost without error, the latter often suffers from missing data. Applying this block alternating CoCoLasso algorithm to such data could allow improvements of explained variance and take advantage of missing and noisy data.



## **PgmNr 1414: Clinical surface measurements derived from 3D imaging on a mobile device: Semiautomated workflow and results suggest utility for high-throughput phenotyping.**

### **Authors:**

A. Kumar; E. Kravets; J.A. Bernstein

[View Session](#) [Add to Schedule](#)

**Affiliation:** Stanford University School of Medicine, Stanford, California.

---

While dysmorphology plays an essential role in delineation and diagnosis of many human syndromes, the acceleration in use of molecular testing has outpaced capacity to perform comprehensive clinical evaluation. Recent approaches to capture facial geometry using 2D facial images have demonstrated promise for high-throughput phenotyping but are affected by artifacts related to projection distortion that affect measurement accuracy. Capturing 3D images can circumvent these limitations but current devices (3dMD, Atlanta GA) are less widely available due to cost or space limitations.

Infrared based sensors used for facial recognition in mobile phones can capture 3D images for use in clinical measurements. The portability and ubiquity of these sensors may overcome some barriers to the widespread use of 3D clinical surface imaging. Here, we describe and validate a methodology and workflow to obtain clinically relevant dysmorphology measurements in a semiautomated fashion employing a mobile application and custom software.

We developed software to facilitate a workflow that comprises 1) capture of 3D images using infrared sensors from the iPhone X (Apple Inc, Cupertino CA), 2) subsequent semiautomated image analysis to identify facial landmarks based on publicly available models for facial recognition 3) calculation of point-point distances including palpebral fissure length and inner/outer canthal distance. We apply this approach as well as a conventional 3D imaging platform (3dMD, Atlanta GA) and compare measurements for 10 individuals. Measurements were obtained in triplicate for both modalities. Across individuals average differences in linear measurements from 3dMD and the iPhoneX ranged from 0.09-3.3mm for inner canthal distance and 0.33-2.5mm for outer canthal distance. Validation data from a pediatric population will also be presented. Potential sources of error positioning of landmarks during image analysis. As mobile sensors and computing continue to improve they may represent an increasingly important method of capturing clinical phenotypes to facilitate the diagnosis and delineation of syndrome.

## PgmNr 1415: HLA allele imputation with deep convolutional neural network.

### Authors:

C. Chi; L.F. Barcellos

[View Session](#) [Add to Schedule](#)

**Affiliation:** UC Berkeley, Berkeley, California

---

Human leukocyte antigen (HLA) genes in the major histocompatibility complex encode antigen-presenting proteins within the host immune system. HLA alleles are highly polymorphic and many have large effect sizes in autoimmune and infectious diseases, but direct HLA typing is expensive. Existing methods such as SNP2HLA have been developed to perform HLA allele imputation from genotype data. However, while imputation performs well for common alleles, quality drops for less frequent alleles. We investigate whether deep convolutional neural networks (CNN) can improve imputation performance compared to existing methods.

Our dataset is comprised of 5,225 individuals of European ancestry from the Type 1 Diabetes Genetics Consortium (T1DGC), each with 5,698 SNPs genotyped with the Illumina 550K array. HLA alleles were directly typed for *HLA\*A*, *HLA\*B*, *HLA\*C*, *HLA\*DPA1*, *HLA\*DPB1*, *HLA\*DQA1*, *HLA\*DQB1*, and *HLA\*DRB1* at 2-field resolution, with a total of 296 unique HLA alleles. For each HLA locus, we use flanking SNPs within 250 kb to predict the HLA allele. SNPs are first transformed into 5-grams before input into the CNN. Our CNN model consists of an embedding layer to train a feature vector representation for each 5-gram, followed by two 1D convolution layers, a dense layer, and a final softmax output layer for prediction. Performance is evaluated with 5-fold cross validation accuracy, with bootstraps ( $B = 1,000$ ) on the hold-out dataset to establish approximate 90% quantile intervals on accuracy. Our method is compared against SNP2HLA, which is considered the state-of-the-art for HLA allele imputation.

We report evidence that CNN outperforms SNP2HLA. SNP2HLA achieved a 5-fold cross validation accuracy of 96.05%, while CNN achieved 97.20% (97.10%, 97.29%). For each HLA locus, the CNN imputes as least as accurately as SNP2HLA, and significantly more accurately for *HLA\*B*, *HLA\*DRB1*, *HLA\*A*, and *HLA\*DPA1*. An occlusion sensitivity analysis, which blocks out SNPs in the neighborhood of a tag SNP to see how much the probability of the true allele drops, shows evidence that the CNN is indeed predicting HLA alleles using tag SNPs. These results show our application of CNN improves upon existing imputation methodology for HLA alleles.

## PgmNr 1416: Accurate and rapid prediction of first-line tuberculosis drug resistance using machine learning method.

### Authors:

Kuang; M. Garvin; F. Wang; Z. Zhang; K. Hernandez; G. Kuffel; R. Grossman

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Translational Data Science, University of Chicago, Chicago, IL

---

Antimicrobial resistant (AMR) infection is one of the main threats to human health. Accurate and rapid in silico AMR prediction approaches are essential for physicians to prescribe the correct antibiotic to patients in a timely and effective manner. Existing knowledge-based AMR prediction tools, such as Mykrobe and ARIBA, process genomic sequencing data of bacterial isolates and achieve varying results: high accuracy on some antibiotics but relatively low accuracy on others. Here we present a machine learning method that predicts AMR to first-line Mycobacterium tuberculosis drugs. This method uses microbial lineage and detected AMR-associated genes and single nucleotide polymorphisms from raw sequencing data as input and completes AMR prediction of a target mycobacterium isolate in 6 minutes requiring only the computational resources of a standard laptop. Using an independent validation data set, in terms of accuracy, we found that this method results in significant performance increase in drugs that previously had relatively poor prediction from the state-of-the-art

knowledge-based method Mykrobe (76.3 % to 89.3% for ethambutol, and 92.5 % to 94.5% for rifampicin), while it performs as well as Mykrobe for isoniazid (96.2%) and pyrazinamide (87.3%).

In addition, the same method we describe here can be directly applied to AMR prediction of different antibiotics for other bacteria, if sequencing data, drug resistance phenotype, and lineage data are available for a large number of strains that cover the existing diversity.

## PgmNr 1417: Evaluation of multi-omic methods applied to immunology.

**Authors:**

R. Mathur; E. Earley; N. Vandergift

[View Session](#) [Add to Schedule](#)

**Affiliation:** Biostatistics & Epidemiology, RTI International, Research Triangle Park, North Carolina.

---

Future breakthroughs in the field of human genetics will increasingly come from integration of high dimensional datasets. Datasets with sequencing, metabolomics, methylation arrays, and others are now routine; however, these only measure one component of a complex system. Common human diseases appear to be caused by an accumulation of many low-effect variants, many of which exist outside of current genome annotations. Thus, elucidating the causes and therapeutic approaches to common diseases will require a more integrated approach that studies traits at multiple levels of organization.

Traditionally, integrating multiple 'omic datasets has been a serial process where experimental signals from one 'omic assay (e.g., GWAS) are then fed into another 'omic assay for validation or characterization (e.g., RNAseq). This approach prevents the discovery of synergistic effects which may exist between the 'omic datasets. In addition, using the first 'omic dataset to inform analysis of the subsequent 'omics datasets can prevent discovery of novel findings hidden in the latter datasets.

We have compared recently published multi-omic integration methods which analyze datasets simultaneously, including DIABLO, MCIA, and LASSO. Each tool approaches biomarker discovery with related but distinct statistical methods. We have used publicly available data (ImmPort, NIAID) from two different multi-omics datasets, each using a different combination of 'omic datasets. Correlated molecular signals (or a network of them) from transcriptomic, methylation, and metabolomics for two datasets, an influenza vaccine response dataset (n=150) and a newborn development dataset from Gambia and Papua New Guinea (n=50), were discovered and compared between all methods.

## PgmNr 1418: Gene scoring algorithms for identifying disease candidate genes.

### Authors:

Y. Nam <sup>1</sup>; D. Kim <sup>2,3</sup>; H. Shin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Industrial Engineering, Ajou University, Republic of Korea; 2) Department of Biostatistics, Epidemiology & Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 3) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA

---

A major understanding of the underlying pathophysiology of autoimmune diseases has been evaluated in genome-wide association scans, which have identified a degree of genetic sharing among autoimmune diseases such as rheumatoid arthritis and multiple sclerosis, among others. Based on the potential commonality among immune genes, we design a graph-based gene scoring algorithm. It initially constructs a backbone network based on protein interactions. Then, patient gene expression networks are added into the network. In the add-on process, disease-specific networks were obtained from weighted gene co-expression network analysis (WGCNA) from patients with a specific immune disease. Any WGCNA network that passes the screening procedure can be added on. With the disease-specific network, it employs the semi-supervised learning for gene scoring. The proposed method was applied to 27 immune diseases. Based on the scoring results, gene scoring identified potential key genes in immune diseases. In terms of prediction performance, an average AUROC of 0.82 was achieved, lifted from Avg. 0.76. Highly ranked genes are evidenced through retrieval and review out of PubMed literatures.

## **PgmNr 1419: The investigation of cell separation-induced gene expression via a penalized deconvolution approach.**

### **Authors:**

A. Tai <sup>1</sup>; W. Hsieh <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute of Statistics, National Chiao Tung University, Hsinchu County, Taiwan; 2) Institute of Statistics, National Tsing Hua University, Hsinchu County, Taiwan

---

Recently, there has been increasing interest in studying genomics and transcriptomics at the single cell level. One of the keys to single-cell study is the success of developing cell sorting technology that is used to separate cells according to their type. However, the process of cell isolation changes the cell microenvironment that affects the gene activity, and thus, the gene expression change induced by cell separation can bias the conclusion of the single-cell study. To address the perturbation of the gene expression caused by cell sorting, we proposed the statistical deconvolution approach, named PEACH, to decompose the cell type-specific expression from bulk tissue and at the same time identify the genes that are expressed differentially to the pure cell samples that were affected by the cell sorting procedure. In this study, PEACH uses cell type-specific samples as the reference to estimate the underlying cell proportions of bulk samples with immune-related diseases and to identify the genes changing in its expression. As a result, the genes we identified are related to biological functions such as stress-responsive transcription factors, cellular metabolism, ribosomal proteins, and human leukocyte antigen. Our study illustrates that genes sensitive to the cell sorting process are biological meaningfully and non-negligible, and it may provide new insight into single cell study for transcriptomic analysis.

## PgmNr 1420: Navigating the treacherous waters of HLA imputation with the SHLARC (SNP-HLA Reference Consortium).

### Authors:

N. Vince <sup>1,2</sup>; V. Douillard <sup>1,2</sup>; S. Limou <sup>1,2,3</sup>; P.A. Gourraud <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre de Recherche en Transplantation et Immunologie UMR 1064, INSERM, Université de Nantes, Nantes, France; 2) Institut de Transplantation Urologie Néphrologie (ITUN), CHU Nantes, Nantes, France; 3) Ecole Centrale de Nantes, Nantes, France

---

The HLA region plays a crucial role in numerous pathologies as it accounts for 25% of known associations from the GWAS catalog, especially with immune-related diseases. Although SNP association studies importantly expanded in the last decade, direct HLA allele association has been hindered by the complexity of typing. HLA imputation offers a statistical alternative to current HLA typing, cutting costs and time alike. Customized machine learning models using attribute bagging, available within the R package HIBAG, can be generated from individuals with known SNP+HLA data to allow prediction of HLA alleles from SNPs. The composition of these models (or reference panels) highly affects imputation accuracy as target populations differ in both SNP and HLA allele diversity and frequency. To assess parameters potentially impacting this accuracy, we used close to 2,000 individuals from the 1000 Genomes Project and an African-American population. Our results demonstrated the effect of: 1) sample size, with a 2x increase (average of 45% to 86%) in accuracy from 10 to 100 individuals; 2) SNP number, accuracy improved from 75% to 86% with 100 to 5,000 SNPs. Additionally, we showed that custom models (i.e. taking only matching SNPs between reference panels and individuals to impute: reducing 2-15x the number of SNPs) can significantly reduce computation time and limit the prior need for SNP imputation. Our results contribute to define gold standards on how to dive in the HLA imputation waters, and call for further collective expeditions to grasp the role of pre-modelling, with population matching and identification of essential SNPs. We have therefore launched an IHI workshop component named 'SHLARC' ([bit.ly/2WGw8hr](https://bit.ly/2WGw8hr)) to gather immunogeneticists on a unique platform 1) to build and share large public reference panels with anonymized data, or 2) to directly impute HLA from SNPs, using our on-demand computation power. Moreover, we started to assemble scientists with complementary expertise (immunology, bioinformatics, informatics, intensive computing, artificial intelligence, inflammation diseases, population genetics, genomics, kidney diseases, immunogenetics, transplantation) from 18 labs in 10 countries and 4 continents aiming to collect data but also to improve methods and facilitate access to HLA imputation models. Altogether, we believe sharing SNP+HLA data in a global consortium will be beneficial to all and place HLA association at the forefront of immunogenomics.

## PgmNr 1421: HLA-poll, surveying tools for multiple HLA-prediction computational programs from WES/WGS data.

### Authors:

Zhuo <sup>1</sup>; K. Parr <sup>2</sup>; R. Stewart <sup>1</sup>; K. Quann <sup>2</sup>; W. Shlomchik <sup>2</sup>; S. Ito <sup>2</sup>; A. Berndt <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UPMC Genome Center, University of Pittsburgh Medical Center, Pittsburgh PA; 2) Hematopoietic Stem Cell Transplant and Cell Therapy, Division of Hematology-Oncology Department of Medicine, University of Pittsburgh, Pittsburgh PA

---

**Background:** Human leukocyte antigen (HLA) typing is crucial for donor-recipient matching in solid organ and stem cell transplantation as well as for a selection of suitable subjects in certain immune therapy. The traditional HLA typing methods are generally time and labor consuming. As sequencing costs drop exponentially, whole exome or genome sequencing (WES/WGS) has become a widely acceptable method-of-choice for genetic diagnosis and precision medicine, which pushes the demand in silico HLA typing on WES/WGS data.

**Methods:** We performed WES and WGS for GIAB NA12878 trio DNA sample and compared to a truth set on three Class I and three Class II HLA genes. WES with 80x coverage was also performed for 16 samples of donor-recipient pairs from stem cell transplantation and we compared the result to the clinically certified HLA-typing result using PCR method with sequence-specific oligonucleotide (SSO) probes. We built a capsulated pipeline HLA-poll, which surveys the HLA typing result from six computational algorithms; HLA-HD, HLAMiner, HLAScan, HLA-VBseq, Kourami, and Polysolver and generate a combined calling report.

**Results:** For the GIAB trio WGS, HLA-HD, Kourami, and HLA-scan all have >95% positive predictive value (PPV) and <5% false discovery rate (FDR). For the GIAB trio WES, HLA-HD demonstrated the best performance in PPV and FDR in comparison to five other algorithms. For the 16 clinical samples, both HLA-HD and Kourami have >95% PPV and <5% FDR. Our HLA-poll passes the >95% PPV and <5% FDR thresholds in all three cohorts.

**Conclusion:** Although various algorithms are available for in silico HLA typing on WES/WGS data, each algorithm has strengths and weaknesses. Most of the algorithms performed better in WGS than in WES. Class I HLAs are generally typed with better accuracy in comparison to Class II HLAs. By combining six algorithms, our HLA-poll showed convincing performance across various examined HLA classes in tested WES and WGS samples. Our method can be also applied to several Class III genes which needed to be validated with a diagonal truth set data.



## PgmNr 1422: Machine learning feature selection integrated with co-expression and Bayesian network analyses further elucidates inflammatory bowel disease in a large multi-omics cohort.

### Authors:

P.H. Comella <sup>1,2</sup>; N.D. Beckmann <sup>1,2</sup>; S. Gopinath <sup>1,2</sup>; S. Tyler <sup>1,2</sup>; A.T. Cohain <sup>1,2</sup>; R. Savic <sup>1,2</sup>; R. Kosoy <sup>1,2</sup>; W.-M. Song <sup>1,2</sup>; R. Huang <sup>3</sup>; W. Wang <sup>1,2</sup>; M. Suarez-Farinas <sup>1,3</sup>; B. Losic <sup>1,2</sup>; J. Zhu <sup>1,2,4</sup>; A. Kasarskis <sup>1,2</sup>; C.A. Argmann <sup>1,2</sup>; L.A. Peters <sup>1,2</sup>; E.E. Schadt <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Data Science, Icahn School of Medicine at Mount Sinai, New York, New York.; 2) Icahn Institute for Genomics and Multiscale Biology, New York City, NY; 3) Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York City, NY; 4) Sema4 Genomics, 333 Ludlow Street, South Tower, 3rd floor Stamford, CT 06902

---

In the Big Data era of biology, -Omics technologies provide a wealth of information, but obtaining useful knowledge from this abundance of data can pose a challenge for researchers. Machine learning classifiers are a useful tool for exploring large datasets, but their performance can be stunted in cases where relatively few features (e.g., genes) are important for prediction, and those features that are chosen to maximize prediction accuracy may not relate directly to biological processes. To improve this, we have utilized a machine learning feature selection and classification strategy which relies on iterative feature ranking that is robust to these limitations. Application of our approach to RNA sequencing data generated on intestinal biopsies isolated from an Inflammatory Bowel Disease (IBD) cohort comprised of Ulcerative Colitis (UC) and Crohn's Disease (CD) patients and controls, resulted in the identification of a focused, stable, and predictive classifier that could well distinguish IBD cases from controls based upon a set of predictive genes identified by our feature selection strategy. The predictive set of genes identified were enriched in highly interconnected modules derived from the RNA-sequencing data, indicating mechanistic links of the machine learning classifiers to the underlying biological processes and disease phenotypes. Bayesian networks were created from these machine learning features and co-expression modules to uncover how these genes interact with each other and how they differ within disease and non-disease states. The synergy of the classifiers, co-expression, and Bayesian networks are being used to further elucidate differences between not only IBD and healthy controls, but also differences between UC and CD in various intestinal biopsy tissues. This feature selection strategy effectively reduces the dimensionality of a large RNA-seq Inflammatory Bowel Disease dataset (~100 genes per classifier with Receiver Operator Curve Area Under Curves: 0.7 - 0.9) while reflecting a greater predictive biological context of disease as shown in the enriched co-expression modules. Key Driver Analysis performed on the Bayesian networks uncovered genes that potentially modulate disease state. Future direction will be leveraging the classifiers, co-expression, and Bayesian networks to form a multi-omics model that incorporates the other datatypes from the cohort in an effort to better predict phenotypic outcomes of treatment effects.

## PgmNr 1423: A novel algorithm for *KIR* copy number imputation by KIBAG and consolidation of population specific *HLA* & *KIR* references in HKimpNet (*HLA* & *KIR* imputation network).

### Authors:

S. Khor<sup>1,2</sup>; X. Zheng<sup>3</sup>; A. Ishitani<sup>4</sup>; F. Azuma<sup>5</sup>; C. Pyo<sup>6</sup>; Y. Omae<sup>1,2</sup>; T. Yabe<sup>5</sup>; D. Geraghty<sup>6</sup>; M. Carrington<sup>7,8</sup>; K. Tokunaga<sup>1,2</sup>; HLA & KIR imputation network (HKimpnet)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Medical Science Project, National Center for Global Health and Medicine, Tokyo, Tokyo, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Bunkyo-ku, Japan; 3) Department of Biostatistics, University of Washington, Seattle, WA, USA; 4) Nara Medical University, Kashihara, Nara 634, Japan; 5) Japanese Red Cross Kanto-Koshinetsu Block Blood Center, Tokyo, Japan; 6) Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 7) Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, MD, United States; 8) Ragon Institute of the Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Boston, MA, United States

---

Natural killer (NK) cells are essential in innate immunity as well as playing a role in antigen-specific responses. NK cell activity is genetically modulated by differential expression of inhibitory and activating killer cell immunoglobulin-like receptors (*KIR*) that recognize *HLA* class I ligands. The high level of both allelic and gene copy number polymorphism of *KIR* may have resulted from selection pressure driven by exposure to a wide variety of diseases, possibly in parallel with or causally related to selection pressure underlying *HLA* polymorphism. Indeed, coordinated *KIR*-*HLA* associations have been detected with HIV infection, hepatitis C infection, human papilloma virus (HPV) induced cervical cancer, psoriatic arthritis and type 1 diabetes. With the recent advancement of next generation sequencing (NGS) technology, *KIR* can now be characterized to high resolution including *KIR* gene copy number (CP), *KIR* haplotypes and *KIR* allele types. However, *KIR* gene detection is currently very limited when examining genome wide SNP data from GWAS studies using conventional computational methods, despite the high-density SNP coverage. Thus, we developed a new methodology named KIBAG (*KIR* Genotype Imputation with Attribute Bagging) which enables a rapid and inexpensive way to determine *KIR* typing from a GWAS dataset. *KIR* typing of 726 Japanese individuals and 2,383 individuals from a wide range of populations was performed using the Scisco Genetics *KIR* typing kit and the Illumina Miseq sequencing system. SNP information of these samples were determined by a combination of genotyping platform including Illumina Omni 2.5, Affymetrix Axiom Japonica V1 & V2 and Affymetrix 6.0. *KIR* reference types were generated by the attribute bagging method applied in KIBAG. Both internal validation and external validation of *KIR* CP imputation were performed and average accuracies of 97.9% and 95.7% were obtained respectively across 16 *KIR* genes. However, *KIR* with > 2 copies remain the most difficult to impute due to the insufficiency of individuals with 3 gene copy numbers of any given *KIR* gene in the reference panel. Re-examination of GWAS with *HLA* class I associations or with SNP peaks in the *KIR* region may reveal novel *HLA*-*KIR* associations. This work is part of the project of HKimpNet (*HLA* & *KIR* Imputation Network) with the aim of providing

population specific *HLA* & *KIR* imputation systems to the research community.

## PgmNr 1424: Towards HLA-VBSeq v3: An HLA calling package for HLA class I and II.

### Authors:

M. Nagasaki <sup>1</sup>; Y. Wang <sup>5</sup>; S.S. Khor <sup>2,3</sup>; O. Gervais <sup>1</sup>; Y. Kawai <sup>2,3</sup>; Y. Hitomi <sup>2,4</sup>; K. Tokunaga <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Sakyo-ku, Kyoto, Japan.; 2) The University of Tokyo, Bunkyo-ku, Tokyo, Japan.; 3) Genome Medical Science Project, National Center for Global Health and Medicine, Shinjuku-ku, Tokyo, Japan.; 4) Department of Microbiology, Hoshi University School of Pharmacy and Pharmaceutical Sciences, Shinagawa-ku, Tokyo, Japan.; 5) Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan.

---

**Introduction:** The human leukocyte antigen (HLA) system, which is the most variable gene region in the entire genome, encodes the major histocompatibility complex (MHC) proteins in humans on chromosome six and has been reported to be associated with numerous immune-mediated diseases. Various HLA calling tools from whole genome sequencing data (WGS) have been developing. Our tool HLA-VBSeq v1 (Nariai *et al.* BMC Genomics 2015) is one of the tools to estimate the most probable combinations of HLA genotypes from the multi-mapped reads to the HLA reference sequence. From the limitation of this approach, the HLA calling accuracy partially depends on the haplotype coverage of HLA reference sequence. The largest international HLA sequence database IPD-IMGT/HLA still cannot fully cover especially population specific HLA haplotypes. The situation was not exceptional to the Japanese population.

### Materials and Methods:

To solve the problem, we have recently developed a new package HLA-VBSeq v2 (Wang *et al.* Human Genome Variation 2019) that integrate the novel HLA Class I sequences constructed from PacBio sequencer that cover the full gene body regions to the major haplotypes of HLA-A, HLA-B, and HLA-C in Japanese (Mimori *et al.* Pharmacogenomics J 2019). In addition, we have also extended the core implementation of HLA-VBSeq v2 to estimate accurate haplotype combinations especially for the recent high coverage sequencing data (more than 30x) from short-read sequencer towards the release of HLA-VBSeq v3,

**Results:** We evaluated the performance of HLA-VBSeq v3 by using 418 Japanese samples as testing data, in which HLA genotypes were determined with both Luminex technology and WGS HLA typing kit. All of the tests were implemented based on 4-digit resolution. As a result, the calling performance of both HLA class I and II in Japanese population is significantly increased compared to HLA-VBSeq v1.

**Conclusions:** HLA-VBSeq v3 is the better calling tool for the former package v1 and v2. The software will be available from (<http://nagasakilab.csml.org/hla/>) as HLA-VBSeq v3.

## PgmNr 1425: A novel workflow for identifying causal variants underlying multivariate GWAS results.

### Authors:

S. Ruotsalainen<sup>1</sup>; J. Partanen<sup>1</sup>; I. Surakka<sup>2</sup>; M. Pirinen<sup>3,4</sup>; C. Benner<sup>1,4</sup>; A. Cichonska<sup>1,5,6</sup>; M. Salmi<sup>7,8</sup>; S. Jalkanen<sup>7,8</sup>; V. Salomaa<sup>9</sup>; A. Palotie<sup>1,10,11</sup>; M.J. Daly<sup>1,10,11</sup>; S. Ripatti<sup>1,3,10</sup>; J. Koskela<sup>1,10</sup>; FinnGen

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland, FIMM, Helsinki, Espoo, Finland; 2) Department of Internal Medicine, University of Michigan, Ann Arbor, MI, US; 3) Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland; 4) Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; 5) Department of Computer Science, Helsinki Institute for Information Technology HIIT, Aalto University, Espoo, Finland; 6) Department of Future Technologies, University of Turku, Turku, Finland; 7) Medicity Research Laboratory, University of Turku, Turku, Finland; 8) Department of Medical Microbiology and Immunology, University of Turku, Turku, Finland; 9) National Institute for Health and Welfare, Helsinki, Finland; 10) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, US; 11) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, US

---

Multivariate analysis of correlated traits increases power compared to univariate analysis. To date, however, multivariate genome-wide association studies (GWAS) have had some evident shortcomings: they have lacked essential tools for follow-up readily available for univariate GWAS, such as fine-mapping. Here we introduce a novel fine-mapping workflow that integrates Canonical Correlation Analysis (CCA) of correlated traits to boost statistical power of identifying causal variants from multivariate GWAS summary statistics.

We performed multivariate GWAS for 12 highly correlated inflammatory markers in the population-based FINRISK study of 6,890 Finns using metaCCA, a software implementing CCA for univariate GWAS summary statistics. In multivariate GWAS, CCA aims to find the linear combination of the phenotypes that is most highly correlated with a particular variant. We further developed metaCCA to produce these linear combination phenotypes (LCP) at a single variant level allowing us to perform LCP-GWAS detecting 11 independent genome-wide significant loci, 3 of which were not detected by univariate method. We used FINEMAP, a software aiming to identify causal variants from summary statistics using the Shotgun Stochastic Search, to identify causal variants from LCP-GWAS summary statistics.

Overall our FINEMAP results were very much in line with both multivariate GWAS and stepwise conditional analysis. FINEMAP increased the number of variants of interest from 11 lead variants from multivariate GWAS to 18 putative causal variants. In the *PCSK6* loci the top configuration included 3 variants among which rs2955967 and rs6598475 showed disease associations ( $p < 1e-4$ ) in the large Finnish FinnGen study ( $N = 135,638$ ). We also detected a Finnish specific missense variant rs200532195 in the *F5* locus with a higher probability of being causal than the lead variant in the locus. These variants would have gone undetected without LCP-GWAS and fine-mapping.

Our results demonstrate that multivariate GWAS outperforms univariate GWAS and fine-mapping multivariate GWAS results further increases the number of interesting variants. In conclusion, we introduce a novel and simple workflow that enables fine-mapping of multivariate GWAS signals, an important follow-up analysis previously only available for univariate GWAS results.

## PgmNr 1426: Algorithm for gene regulatory network inference recovers biological insights from large-scale gene expression data.

### Authors:

A. Loboda <sup>1,2</sup>; M. Artomov <sup>1,3</sup>; M. Daly <sup>1,3,4</sup>; A. Sergushichev <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Institute, Cambridge, MA, USA; 2) ITMO University, St. Petersburg, Russia; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) Institute for Molecular Medicine Finland, Helsinki, Finland

---

Gene regulatory network (GRN) is a set of causal interactions between genes underlying observed gene expression levels changes in response to external stimuli. Such networks are widely used in network analysis tools where perceiving gene regulation mechanisms help unravel hidden causes behind experimentally observed data. Currently existing methods for inference of GRNs do not allow processing the data of the real experimental size in current human studies. We developed a probability-based GRN inference method with ability to process tens of thousands of genes in reasonable time with robust and reproducible performance. We present new method based on Monte-Carlo Markov Chain (MCMC) approach. To improve quality of resulting GRN we propose an approach of careful data discretization. However, this requires additional time in network likelihood estimation. To resolve this issue and overall high complexity of the problem we provide an efficient way of storing computed probabilities along with advanced computational approach in running MCMC. We evaluated our method on simulated data to confirm its superior properties both in solution quality and resource usage compared to the known methods. Further, we used Brain Span developing brain expression data to construct regulatory networks that change significantly between pre-natal and post-natal conditions. Relating the observed gene regulation patterns we investigated relevance of gene expression regulatory genes in our network to known intellectual disability and autistic spectrum disorders. **In summary**, we present the first to date method allowing processing of the full-size human gene expression data to reconstruct regulatory interactions pattern.

## PgmNr 1427: *slivar*: Flexible filtering and annotation of genetic variants with simple expressions.

### Authors:

B.S. Pedersen; A.R. Quinlan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, Salt Lake City, Utah.

---

*slivar* is a suite of command-line tools that facilitate rare disease research. It operates directly on data in variant call format (VCF) using a simple expression language that allows users to apply customizable filters. For example a filter to find *de novo* variants could be:

```
variant.FILTER == "PASS" && INFO.gnomad_popmax_af < 0.005
&& kid.het && mom.hom_ref && dad.hom_ref
&& kid.DP > 7 && mom.DP > 7 && dad.DP > 7
```

This example query enforces the expected genotype pattern for a *de novo* mutation and requires a sequencing depth of more than 7 in all three samples. In addition, it requires that candidate mutations are also rare in gnomAD, according to an annotation in the INFO field. Finally, to ensure high quality candidates, it requires that the FILTER field of the variant would be 'PASS'. This expression would be **applied to each variant, for each trio** discovered from the VCF and a pedigree file that indicates the relationships among samples.

Because of the utility of the (gnomAD) population allele frequency in rare disease research, we have created a special format and annotator within *slivar* that reduces, for example, the whole genome gnomAD VCF from 450 gigabytes down to about 2 gigabytes for selected annotations. This reduced file allows *slivar* to annotate more than 20 thousand variants per second.

We show that combining *slivar*'s fast expression language with population allele frequency filtering reduces the number of candidate variants to about 15 for a typical rare-disease trio in a matter of seconds. This number can be further reduced by filtering on the functional status (missense/stop-gained/synonymous) of each variant. This includes variants meeting *de novo*, X-linked recessive, X-linked *de novo*, compound heterozygous and single-site autosomal recessive inheritance patterns. We have encapsulated these analyses into a simple, best-practices workflow for rare-disease trios.

*slivar* is available under the MIT license at: <https://github.com/brentp/slivar> with the rare-disease commands at: [http://bit.ly/slivar\\_rare\\_disease](http://bit.ly/slivar_rare_disease)



## PgmNr 1428: Game theoretic centrality supports link between autism spectrum disorder and the HLA complex.

### Authors:

M.W. Sun <sup>1,6</sup>; S. Moretti <sup>2</sup>; K. Paskov <sup>1</sup>; N. Stockham <sup>3</sup>; J.Y. Jung <sup>1,6</sup>; M. Varma <sup>4</sup>; B. Chrisman <sup>5</sup>; P. Washington <sup>5</sup>; D.P. Wall <sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Data Science, Stanford University, Stanford, CA; 2) CNRS, UMR 7243, Université Paris-Dauphine, PSL Research University, Paris, France; 3) Department of Neuroscience, Stanford University, Stanford, CA; 4) Department of Computer Science, Stanford University, Stanford, CA; 5) Department of Bioengineering, Stanford University, Stanford, CA; 6) Department of Pediatrics, Stanford University, Stanford, CA

---

Complex genetic diseases with etiological heterogeneity like Autism Spectrum Disorder (ASD) often pose a challenge for traditional genome-wide association study approaches in defining a clear genotype to phenotype model. We propose applying a new method based on coalitional game theory that can capture the combinatorial interaction of mutations that manifests into a disease.

Coalitional game theory is used to model interactions among players--in our case genes--when they form coalitions. This approach has been previously employed to identify candidate genes in gene expression and whole genome sequence data. We applied a game theoretic centrality measure based on *Shapley value* to rank genes by their relevance--the individual gene's synergistic influence--in a protein-protein interaction network. Shapley values measure the average marginal contribution of a player across all possible coalitions. *Game centrality* extends the notion of Shapley value to the evaluation of a gene's contribution to the overall connectivity of its corresponding node in a biological network. Each of the nodes can also be assigned a weight based on *a priori* importance of the genes. Compared to other commonly used centrality measures such as degree centrality and betweenness centrality, game centrality takes into account not only the contribution of a single gene to the network, but also the interaction with all possible groups of genes.

We implemented and applied this method to rank genes using game centrality on whole genomes from 756 multiplex autism families. Likely gene disrupting mutations in coding regions were encoded into case (ASD) and control (unaffected) binary matrices. Genes harboring these mutations were used to generate a protein-protein interaction graph with STRING. We computed game centrality on this graph both with and without *a priori* knowledge--corresponding Shapley values computed from the binary matrices. The top five percent genes with the highest game centrality in both approaches were enriched for pathways of the immune system. In particular, three of the 19 selected genes--HLA-B, HLA-G, and HLA-DRB5--are part of the human leukocyte antigen (HLA) complex, which has been previously associated with ASD. These results suggest that game centrality can identify influential, disease-associated genes within biological networks, thereby decoding the polygenic underpinnings of complex diseases like autism.

## PgmNr 1429: Transcript mediation analysis of multi-trait associations.

**Authors:**

B. Wu

[View Session](#) [Add to Schedule](#)**Affiliation:** Division of Biostatistics, University of Minnesota, Minneapolis, Minnesota.

---

Recent research has suggested that the transcript expression plays an important role in the genetic architecture of most complex human traits: they can act as a direct or indirect mediator for the genetic effects on most clinical phenotypes. Utilizing the publicly available GWAS summary data and the large-scale gene expression and genetic data from GTEx project, and further capitalizing on the widespread pleiotropy association across multiple traits, we set out to develop novel methods and tools for studying the transcript mediated multi-trait associations. Our methods can quantify the contribution of mediated association and the residual association, which on the other hand can help boost our power to better detect those genetic variants that are significantly associated with multiple correlated phenotypes.

We will demonstrate the utility of our proposed methods through rigorous numerical studies and analysis of GWAS summary data for multiple lipids traits. Our approach identified many novel loci that were not detected by the existing methods, and also provided novel insights into the underlying genetic architecture of lipids traits.

## PgmNr 1430: Searching for genetic variants matching a given multivariate target profile.

### Authors:

H. Topa <sup>1</sup>; A.S. Havulinna <sup>1,2</sup>; V. Salomaa <sup>2</sup>; M. Kähönen <sup>3,4</sup>; T. Lehtimäki <sup>5</sup>; O.T. Raitakari <sup>6</sup>; M.R. Järvelin <sup>7,8</sup>; M. Ala-Korpela <sup>9,10</sup>; P. Würtz <sup>11</sup>; M. Pirinen <sup>1,12,13</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) The Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland; 4) Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; 5) Department of Clinical Chemistry, Fimlab Laboratories, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; 6) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 7) Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK; 8) Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland; 9) Systems Epidemiology, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia; 10) Faculty of Medicine, University of Oulu & Biocenter Oulu, Oulu, Finland; 11) Nightingale Health Ltd., Helsinki, Finland; 12) Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; 13) Department of Public Health, University of Helsinki, Helsinki, Finland

---

Modern high-throughput technologies provide highly informative multivariate profiles of an exposure or an intervention, often manifested as quantitative changes in the affected traits. Finding similarities between such profiles of an intervention and genetic variants can pinpoint to important targets at the genetic level. Here, we aim to develop a statistical method to rank “candidate profiles” according to their similarities to a given “target profile”.

In our test case, we define a target profile as changes in 53 circulating metabolites associated with starting statin therapy, measured by a nuclear magnetic resonance platform, and we define the candidate profiles as effect sizes of different genetic variants on the same traits, estimated by genome-wide association studies (GWAS). Our goal is to identify the variants that lead to similar changes in the circulating metabolites as the statin therapy.

Conventional solution is to rank the variants by their correlation to the target profile. However, this approach does not account for uncertainty of the estimated effect sizes or the correlations among traits. Therefore, we adopt a Bayesian approach that incorporates these additional sources of information into the posterior distribution of the profiles. We then rank genetic variants by the Wasserstein distance between the posterior distributions of the genetic variant and the target profile. Our simulation results indicate that the ranking based on the Wasserstein distance of the posteriors (WDP) outperforms the correlation-based ranking (COR), especially when the number of traits included in the profiles is small.

We observe a similar behavior when we use GWAS data from Finnish samples (n=10,753) at 156 lipid-associated SNPs to search for genetic variants which have similar metabolic profile to starting statin

therapy. When all 53 traits are available, the top five variants selected by both methods are near the genes *LDLR*, *SORT1*, *APOC1*, *HMGCR* and *APOB*, which is consistent with known function of statins. If we use only the three standard measurements of LDL-C, HDL-C and TG, the ranks assigned to these genes become 1, 2, 3, 5, 11 with WDP and 17, 14, 31, 35, 42 with COR. To sum up, WDP seems to be a more reliable approach for ranking similarities between multivariate profiles than COR, as it is more robust to the number of traits available. Hence, we expect that WDP proves useful to identify new genetic targets that resemble profiles of successful interventions.

## PgmNr 1431: Precision genotyping of APOE from whole genome sequencing.

### Authors:

C. Zhu <sup>1</sup>; J. Farrell <sup>1</sup>; A. Kuzma <sup>6,7</sup>; G. Schellenberg <sup>6,7</sup>; L. Farrer <sup>1,2,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA; 2) Department of Neurology, Boston University School of Medicine, Boston, MA; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 4) Department of Ophthalmology, Boston University School of Medicine, Boston, MA; 5) Department of Epidemiology, Boston University School of Public Health, Boston, MA; 6) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 7) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

As APOE e4 is the strongest risk allele for Alzheimer's disease (AD), the accurate genotyping of APOE is important for both AD research and precision medicine. The genotype is based on two non-synonymous variants (rs429358 and rs7412) within 138 bp of each other. AD researchers have typically called these two variants with RFLP analysis or more recently, TaqMan SNP genotyping. GC-rich genes such as APOE have been difficult regions for whole genome sequencing for calling both SNPs using PCR+ libraries. With the availability of PCR Free (PCR-) sequencing, the GC-rich APOE region should result in better genotype quality and a higher call rate for the 2 APOE variants.

We evaluated our APOE genotyping pipeline on 4789 Alzheimer Disease Sequencing Project WGS samples with both PCR+(N=479) and PCR-(N=4310) libraries with 30x coverage aligned to GRCh38. The APOE WGS pipeline achieves greater than 99.9% accuracy for joint-genotyped whole genome sequencing samples in both PCR+ and PCR- WGS using a "gold standard" based on RFLP and TaqMan genotypes.

To achieve this accuracy, the pipeline uses GATK HaplotypeCaller (4.1.1) with a filter for bacterial reads. We had found reads from bacterial contamination had aligned near rs429358 and interfered with the HaplotypeCaller genotyping. The poor genotype quality (GQ<20) occurred in approximately 2% of the APOE genotype calls for SNP rs429358. Without the bacterial reads, the GQ scores passed (GQ ≥20). To fill-in missing SNP calls (3%) due to low coverage in the PCR+ samples, the VCF was phased with Eagle 2 and then genotyped with the APOE-Genotyper tool.

The WGS APOE calls had a concordance of 99% with the original "gold standard". We selected 18 of the 56 samples that were discordant for retesting and found 17 of 18 samples agreed with the WGS APOE calls. The single WGS ApoE call that was incorrect was due to a low level of human DNA contamination found in that sample (3%). Such human DNA contamination will be taken into account in a future version of the pipeline.

We found that highly accurate APOE genotyping can be achieved in PCR- and PCR+ WGS with the developed pipeline. With the updated "gold standard", we demonstrate that genotypes based on WGS achieved higher precision than lab-based approaches. The APOE pipeline should be useful for high

throughput precision genotyping of APOE calls in large scale sequencing projects such as TopMed and UK Biobank where readily available lab-based calls are not available.

## PgmNr 1432: Circadian phase inference in single-cell RNA-seq.

### Authors:

B. Auerbach <sup>1</sup>; M. Li <sup>2</sup>; G. FitzGerald <sup>3</sup>; R. Anafi <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomics & Computational Biology, University of Pennsylvania, Philadelphia, Pennsylvania.; 2) Biostatistics, University of Pennsylvania, Philadelphia, Pennsylvania.; 3) Pharmacology, University of Pennsylvania, Philadelphia, Pennsylvania.; 4) Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

---

Disruption of the circadian clock and transcriptional timing are common features of aging and age-related disease such as Alzheimer's disease (AD). Characterizing changes in transcriptional timing will be important for understanding the biological consequence of these changes, identifying their cause, and understanding how to prevent/reverse their deleterious effects. Previous bulk RNA-seq time courses have reported age-related damping of circadian rhythms. However, such approaches may be confounded by circadian phase heterogeneity across cells from different types, suggesting the need for single-cell transcriptional measurements, such as single-cell RNA-seq (scRNA-seq). Using scRNA-seq data to study transcriptional timing requires methods to assign internal biological times to cells, such as placement in the circadian cycle. While computational phase inference methods have been developed for bulk RNA-seq samples, no approaches have been developed for scRNA-seq. To fill this gap, we develop a neural network autoencoder based approach for single-cell circadian phase inference. To capture circadian phase information, we introduce two circular nodes in the autoencoder latent space. By constraining their values to lie along the unit circle, these two nodes can capture angular information, and hence be used to infer circadian phase. We further include other latent variables to account for the contribution of other potential confounders, such as batch effect, to gene expression variation. Although autoencoders are ideally suited to model circadian phase, neural networks are an inherently challenging tool due to their non-convex optimization. To improve parameter search for neural networks, we propose an approach that improves both autoencoder training times and reconstruction errors using existing scRNA-seq datasets and simulated data. This method will also be applied to an ongoing scRNA-seq study on AD, which includes 146,473 cells generated across 15 human brains. Disruption in circadian rhythms are common symptoms of AD. By applying our method to this phenotypically well characterized dataset, we expect to find altered patterns of gene expression timing that are associated with AD. With the growing popularity of scRNA-seq in biomedical research, we believe our method offers the community a robust tool that enables the inference of circadian phase with single-cell resolution.

## **PgmNr 1433: MSeqDR mitochondrial disease data interpretation platform: Quick-mitome, registry, genome, and literature mining.**

### **Authors:**

L. Shen <sup>1</sup>; E. McComick <sup>2</sup>; C. Muraresku <sup>2</sup>; Z. Zolkipli-Cunningham <sup>2</sup>; D. Merkurjev <sup>1</sup>; M. Falk <sup>2,3</sup>; X. Gai <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Personalized Medicine, Children's Hospital Los Angeles, Los Angeles, California.; 2) Mitochondrial Medicine Frontier Program, Division of Human Genetics, Department of Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 3) Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 4) Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.

---

MSeqDR (<https://mseqdr.org>) is a centralized genomic bioinformatics Web resource that supports genetic analyses and variant interpretation for people working in mitochondrial biology and disease. MSeqDR is a central data resource hub for the mitochondrial disease community, including MitoMap, HmtDB, LeighMap, ClinVar, ClinGen, PhenoTips, and GENOMIT resources, facilitating data sharing, cross-referencing, and active collaboration.

MSeqDR Content and Tool Update: Genome and phenotype data has been enhanced by assembling a large meta-population reference dataset for mitochondrial DNA (mtDNA) allele population frequencies, compiling ~90K mtDNA genomes from MitoMap, GeneDx, publications and sequencing initiatives, with substantial amount of data from Asian populations. MSeqDR-LSDB is a curated database of pathogenic variants for mitochondrial diseases, genes, and variants generated by data mining and systematic curation of public genomic resources and literature. It hosts 12,000+ variants from 1,607 genes related to mitochondrial function, including over 350 known disease genes with pathogenicity assessments for over 8,000 variants. We also developed a suite of bioinformatics tools to annotate and analyze mtDNA genome variants (MSeqDR-mvTool and Phy-Mer) and centralized access to a suite of tools hosted in or linked to MSeqDR, such as MToolBox from HmtDB, MitoMaster from MitoMap, MitoTIP, and LeighMap.

The mitochondrial disease data interpretation platform is built upon the above rich data content and tools to assess variant pathogenicity. Quick-Mitome automates online Exomiser analysis of whole exome/genome datasets. Then the candidate variants are deeply annotated with disease and mtDNA data. The unified report summarizes the Exomiser ranking and all relevant MSeqDR data. The literature mining tools process mitochondrial disease, variants, genes entries from Pubmed services including pubtator, and reconstruct the variant to disease associations. MSeqDR “pseudo” case registry for rare Leigh disease mines publications to compile the anonymous case level data from over 800 cases.

Overall, MSeqDR now provides extensive sets of genomic data, diverse bioinformatics tools that support mtDNA variant annotation, and haplogroup determination, exome data analysis guided by clinical feature. MSeqDR is now recognized world-wide as the genomic and data collaborative hub



supporting the mitochondrial biology and disease diagnostic research community.

## PgmNr 1434: Computational pipeline for RNA-seq based genetic diagnosis.

### Authors:

M.F. Müller<sup>1</sup>; V.A. Yépez<sup>1</sup>; C. Mertes<sup>1</sup>; D. Andrade<sup>1</sup>; H. Prokisch<sup>2</sup>; J. Gagneur<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Technical University of Munich, Munich, Bavaria, Germany; 2) Helmholtz Zentrum München, Neuherberg, Bavaria, Germany

---

RNA sequencing has become a useful companion to DNA sequencing to diagnose patients with genetic disorders [1,2]. Specifically, three patterns (aberrant expression, aberrant splicing and mono allelic expression) can be detected from RNA-Seq data and used to prioritize candidate variants and genes. Although statistical methods for detecting each of these three patterns have been developed, a pipeline that integrates them and automatizes the generation of results from raw files is lacking and very much needed to analyse cohorts of hundreds of patients. Here, we present such an end-to-end pipeline that not only reports gene candidates from RNA-Seq data, but also summarizes and annotates variants called from DNA sequencing. The pipeline is based on the workflow management tool Snakemake [3] and wBuild [4]. It is built in R in a modularized fashion, so that the user can select which of these three patterns to use. The Snakemake framework makes it possible to interrupt the pipeline and resume it at any point without having to rerun previous steps that do not need to be updated. It also offers a parallelized backend, which allows to execute same processes on different samples at the same time and make effective use of compute clusters. Moreover, the wBuild functionality renders the scripts in HTML format for convenient visualization. This pipeline makes such analyses more stable and simple for larger sample sizes. We have successfully used it to analyze more than 200 patients with rare mitochondrial disorders. Overall, we foresee that the pipeline can be easily deployed on different environments and extended to include other multi omics modules.

[1] Cummings et al. Sci Transl Med. (2017)

[2] Kremer et al., Nature Communications (2017)

[3] <https://snakemake.readthedocs.io>

[4] <https://wbuild.readthedocs.io>

## PgmNr 1435: Genetic characterization of glucose-6-phosphate dehydrogenase deficiency in Qatar.

### Authors:

R. Zaied <sup>1</sup>; S. Malik <sup>1</sup>; N. Sayed <sup>2</sup>; M. Al-Shafai <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Sciences, College of Health Sciences, QU Health, Qatar University, Doha, Qatar; 2) Sidra Medicine, Doha, Qatar

---

Glucose-6-phosphate dehydrogenase (G6PD) deficiency is one of the most common human enzymopathies in the world and is inherited in an X-linked recessive manner. Defects in the G6PD enzyme results in the destruction of premature red blood cells leading to hemolytic anemia, which is usually triggered by the presence of oxidative agents. To date, more than 200 genetic variants of *G6PD* have been identified globally, however, none were reported in the Qatari population, although the disease incidence is around 5%.

In this study, we aim to investigate the genetic basis of G6PD deficiency in Qatar using the Qatar Genome Project (QGP) data, which consists of whole genome sequencing data (WGS) of 6,000 Qatari nationals and long-term residents of Qatar generated using the Illumina HiSeq X Ten1 platform. After processing the WGS data, we identified 21 high impact, 17 moderate impact, 19 low impact, and 320 modifier impact variants, based on SnpEff annotation, that are located within or near the *G6PD* gene. We focused on moderate and high impact variants and used various variant databases and computational analysis tools to investigate their pathogenicity. Among those variants, p.S188F (*G6PD Mediterranean*), p.N126D (*G6PD Ilesha*), p.V68M (*G6PD Asahi/A-*), and p.A335T (*G6PD Chatham*) were the most common known G6PD disease-causing variants seen in Qatar. In addition, we revealed 10 novel variants in the *G6PD* gene (5 high impact and 5 missense variants). We performed molecular visualization analysis of the G6PD variants using PyMol to assess the impact of three novel missense variants and seven reported variants on the protein structure. The novel variants seem to affect G6PD structure and they merit further validation and functional investigations in order to establish genotype-phenotype correlation. This study is the first to address the genetic basis of G6PD deficiency in Qatar and it highlights the importance of investigating such understudied populations in providing novel insights about the disease's pathogenesis.

## PgmNr 1436: Mitochondrial heteroplasmy: A review of current bioinformatic assessment tools.

**Authors:**

S.S. Chill; A. Baranova

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Systems Biology, George Mason University, Fairfax, Virginia

---

Current advancements in next generation sequencing (NGS) technology has increased the resolution of mitochondrial DNA (mtDNA) generate data, allowing for the investigation of features such as heteroplasmy, with greater precision. Heteroplasmy has been identified through several recent studies, although because comprehensive bioinformatic tools have been scarce, data analysis required significant manual manipulation. As associations with complex disease patterns are becoming more evident, particularly as it relates to telomers and aging, tools to automate and standardize processing are in high demand. Four current bioinformatic methods (mtDNA-Server, MitoSeek, MitoSuite, and MToolBox) were assessed, to determine their overall functionality for heteroplasmy discovery, accuracy and consistency with calling features, and availability of quality control metrics. Publicly available data was analysed using each platform, with standard parameters and inputs. While congruency was found between samples (blood and skin) in each platform, the number of heteroplasmic calls, number of congruent pairs between samples, and congruency between heteroplasmic calls on each platform varied. Each platform allowed for automation of many of the steps required to make heteroplasmic calls, however, most still required varying amounts of user manipulation required to obtain heteroplasmy results. Additionally, users were required to curate the final results in terms of heteroplasmic matches among sample types and across platforms, in addition to integrate pathogenic variant information to the heteroplasmic findings on most platforms. Overall, while tools have advanced to allow for more automated calling, differences in platforms calls, in QC metrics, and in input type, demonstrate some of the current bioinformatic needs in the field.

## PgmNr 1437: Assessing and improving mitochondrial haplogroup assignments with low-coverage, genotype data.

### Authors:

E.L. Drummond <sup>1</sup>; J. Knight <sup>2,3,4</sup>; D.J. Clancy <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biology and Life Science, Lancaster University, Lancaster, Lancashire, United Kingdom; 2) Lancaster Medical School, Lancaster University, Lancaster, Lancashire, United Kingdom; 3) Institute of Medical Science, University of Toronto; 4) Psychiatry Department, University of Toronto

Many loci within the mitochondrial DNA are polymorphic in the general population allowing over 5400 haplogroups to be defined by these variations. This makes choosing a set of SNPs for efficient haplogrouping of mtDNA a challenge.

We employed the 1000 Genome sequence data to develop a method to assess the accuracy, precision and efficiency of individual haplogroup prediction based on SNPs typed by the UKBiobank (UKB).

Our protocol involves; (1) Haplogroup assignment (using haplogrep) of individuals in the 1000 genome dataset, (2) Extraction of variation from this dataset based only on loci typed in the UKB dataset, (3) Haplogroup assignment of individuals based on this subset of variation, (4) Comparison of the two assignments. (5) Assessment of success rate by sample and haplogroup, (6) Reduction of precision, by haplogroup, until every group member is perfectly accurately defined.

When samples assigned to haplogroups on the UKB loci variations alone were compared with the gold standard results from the full data, we saw a collapse of 1137 subgroups down to 260 but only H and M contained samples from other letter branches. For example, 687 samples were correctly assigned to L but fell into just 40 of 224 possible, gold-standard subgroups. We had to reduce subgroup specificity of 11 of the 40 to ensure every subgroup member was assigned accurately. In 9 of these 11 adjustments, we were reducing precision for a small minority of samples (N<9) to guarantee group accuracy.

Beyond our specific use, this work flow allowed us to develop a more general tool for the investigation of a batch of SNPs before committing them to a microarray chip and also determine accuracy in other datasets.

The needs of a project will vary according to their target populations and their accuracy threshold. The success of their SNP combination can be visualised on a dendrogram and we can also predict groups obscured by a lack of defining SNPs, and where they will appear in the tree.

Our work has explored the SNP choices made for the microarray chip used by UKBiobank , with a long view to extend inferences and effectively 'impute' the variation in their individuals. This will greatly improve exploration of the effects of mito-nuclear interaction on human health, looking into the

complexities introduced by mito-nuclear ecology potentially throughout the gamut of phenotypes collected by UKB.

## PgmNr 1438: The added value of RNA sequencing over WES for variant interpretation and diagnosis of patients with rare genetic disorders.

### Authors:

V. Yeppez<sup>1,2</sup>; M. Gusic<sup>3,4</sup>; R. Kopajtich<sup>3,4</sup>; C. Mertes<sup>1</sup>; M. Mueller<sup>1</sup>; Y. Liang<sup>1</sup>; S. Stenton<sup>1</sup>; J. Mayr<sup>5</sup>; S. Wortmann<sup>3,4</sup>; A. Rotig<sup>6</sup>; R. Olsen<sup>7</sup>; K. Murayama<sup>8</sup>; D. Piekutowska-Abramczuk<sup>9</sup>; T. Schwarzmayr<sup>3,4</sup>; T. Strom<sup>3,4</sup>; T. Meitinger<sup>3,4</sup>; J. Gagneur<sup>1,2</sup>; H. Prokisch<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Informatics, Technical University of Munich, Munich, Bavaria, Germany; 2) Quantitative Biosciences Munich, Department of Biochemistry, Ludwig-Maximilians Universität, Munich, Germany; 3) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 4) Institute of Human Genetics, Technical University of Munich, Munich, Germany; 5) Department of Pediatrics, Paracelsus Medical University, Salzburg, Austria; 6) Université Paris Descartes—Sorbonne Paris Cité, Institut Imagine, Paris, France; 7) Department of Clinical Medicine - Research Unit for Molecular Medicine, Aarhus, Denmark; 8) Department of Metabolism, Chiba Children's Hospital, Chiba, Japan; 9) Department of Medical Genetics, Children's Memorial Health Institute, Warsaw, Poland

---

RNA sequencing (RNA-Seq) is a promising technology to complement Whole Exome Sequencing (WES) for pinpointing the genetic cause of Mendelian diseases [1,2]. However, with improved and ever-growing databases of WES and Whole Genome Sequencing, and better variant effect predictors, the relative advantage of RNA-Seq may be questioned. Moreover, the relevance of RNA-Seq performed on non-affected tissues, which are typically more accessible for transcriptome profiling, has been debated. To address these questions, we investigated expression patterns in healthy donors from the GTEx project and performed WES and RNA-Seq of fibroblast samples for a cohort of 196 individuals affected with a rare mitochondrial disease, yielding the largest RNA-Seq compendium for rare disease diagnostics. Disease-associated genes are enriched among the genes expressed in the respective affected tissues. Nonetheless, the majority of OMIM genes are expressed in blood or skin. In particular, over 90% of mitochondrial disease genes and over 80% of OMIM diseases genes are expressed in skin, showing that RNA-seq of accessible, yet not affected, tissues is valuable. We considered three patterns used to prioritize genes: Aberrant expression [3], aberrant splicing, and mono allelic expression and developed a computational pipeline to integrate their results. Underexpression outliers were highly enriched (29% vs 0.1% in non-outliers) for rare (MAF <0.001) and likely expression-disrupting variants (stop, frameshift and splice). Similarly, we found a strong enrichment of rare splicing, intronic and missense variants in splicing outliers (18% of outliers with rare splicing variants vs 3% in non-splicing outliers). Also, monoallelic expressed genes were enriched for imprinted genes and for rare heterozygous stop and missense variants. Altogether, this analysis led to increase our diagnosis rate by 13% by (i) validating or (ii) invalidating the impact of WES variants on RNAs, and by (iii) revealing aberrant expression patterns that could not be predicted from WES data alone. Overall, our study and the computational pipeline we offer will help the rare disease research community by providing insights and tools to efficiently use RNA-Seq data for diagnostics.

- [1] Kremer et al, Nat Commun (2017)
- [2] Cummings et al, Sci Transl Med (2017)
- [3] Brechtmann et al, AJHG (2018)



## PgmNr 1439: THUNDER: A reference-free deconvolution method to infer cell type proportions from bulk HiC data.

### Authors:

B. Rowland <sup>1</sup>; R. Huh <sup>1</sup>; Y. Li <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

---

Studying spatial chromosome structure has increased our knowledge of the underlying biological function of regulatory regions in the human genome. Hi-C data provide population averaged estimates of three-dimension (3D) chromatin contacts in the human genome across different cell types in heterogeneous bulk samples. To effectively leverage insights from the analysis of Hi-C data for clinical application, we must control for the almost inevitable confounding factor of differential cell-type proportions across these heterogeneous bulk samples. Deconvolution methods have been successfully applied to similar problems in the gene expression literature because of the availability of single-cell gene expression data, but such data had not existed in the Hi-C context until recently. To take advantage of these recent improvements, we propose a nonnegative matrix factorization (NMF) based Two-step Hi-c UNSupervised DEconvolution appRoach (THUNDER), to infer cell type proportions from bulk Hi-C data. The first step of our approach selects distinguishing features for every cell type if possible, and the second step leverages these features to estimate cell type proportions. We conducted extensive real, single-cell Hi-C-data based simulations to test our method. We explored both inter- and intra-chromosomal contacts to learn how different Hi-C readouts affect profiling. THUNDER proves robust to the level of Hi-C measurement used. In addition, THUNDER, being unsupervised, allows us to estimate cell-type signatures of 3D contacts under the realistic scenario where the number of contributing cell types are known. This feature of THUNDER is significant given that purified cell-type signatures of Hi-C contacts are not yet available in practice. Our results show that with less information THUNDER outperforms alternatives, including supervised methods such as CIBERSORT. For example, when analyzing interchromosomal contacts, THUNDER achieves a 61.38% reduction in mean absolute deviation (MAD) when applied to Hela-HAP1 mixtures and a 66.52% reduction in MAD when analyzing Patski- mouse embryonic fibroblast mixtures. We believe that our method, which can accurately identify cell type proportions in bulk Hi-C data, will be a useful tool to facilitate appropriate adjustments in subsequent analyses such as exploring differential 3D chromatin organization or investigating the genetics governing chromatin contacts.

## **PgmNr 1440: Development and implementation of a computational workflow for the molecular diagnosis of constitutional syndromes using genome-wide DNA methylation analysis.**

**Authors:**

E. Aref-Eshghi; B. Sadikovic

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pathology and Laboratory Medicine, London Health Science Centre, London, Ontario, Canada

---

Constitutional syndromes frequently present with overlapping clinical features and uncertain molecular findings that result in significant challenges in obtaining a diagnosis. Many such disorders exhibit specific patterns in the epigenome (epi-signatures), utilizing which can assist their clinical diagnosis. We describe the computational workflow for mapping and clinical utilization of DNA methylation epi-signatures in ~30 different syndromes, involving feature selection, unsupervised clustering, and training and validation of supervised classification models. We illustrate the level of overlap and dissimilarities across different syndromes and highlight unusual situations when one syndrome presents with more than one epi-signature, two syndromes with contrasting DNA methylation patterns, and others with overlapping profiles. We discuss the analytical challenges posed by these instances and demonstrate the necessity of the concurrent modeling of multiple epi-signatures for disease classifications. Using a multi-class support vector machine classifier trained on ~1,500 informative CpG sites, we show that one can confidently use DNA methylation data for simultaneous classification of all of the syndromes discussed. We demonstrate the utility of this tool in resolving ambiguous clinical cases and disease screening among unresolved cohorts. Finally, we present the procedure used for the implementation of DNA methylation analysis in the clinical setting including the validation workflow, reporting framework, and early diagnostic yield of our computational tool as a clinical test across multiple clinical laboratories in Europe, Canada, and the US.

## PgmNr 1441: AMELIE 3.0: Continuous automatic reanalysis of undiagnosed patients with suspected Mendelian disease.

### Authors:

G. Bejerano <sup>1,2,3,5</sup>; J. Birgmeier <sup>1</sup>; E. Steinberg <sup>1</sup>; E. Bodle <sup>2</sup>; C.A. Deisseroth <sup>1</sup>; K.A. Jagadeesh <sup>1</sup>; P.D. Stenson <sup>4</sup>; D.N. Cooper <sup>4</sup>; J.A. Bernstein <sup>2</sup>; UDN

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, Stanford University, Stanford, California 94305, USA; 2) Department of Pediatrics, Stanford School of Medicine, Stanford, California 94305, USA; 3) Department of Developmental Biology, Stanford University, Stanford, California 94305, USA; 4) Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff, UK; 5) Department of Biomedical Data Sciences, Stanford University, Stanford, California 94305, USA

---

**Background:** An ever growing number of patients with a suspected Mendelian disease are sent to exome or genome sequencing every year, but only 30% receive a definitive diagnosis. This is partly due to our incomplete knowledge about Mendelian-disease-causing genes. Because a novel Mendelian gene-disease association is published, on average, every other day, thousands of undiagnosed patient cases could receive a diagnosis each year if the genomes were regularly compared to the latest literature. With millions of genomes expected to be sequenced for rare disease analysis by 2025, and over 25,000 peer reviewed papers appearing in PubMed every week, manually reanalyzing undiagnosed patients is not sustainable.

**Methods:** We describe an automatic reanalysis framework for patients with suspected, but undiagnosed, Mendelian disorders. The presented framework automatically parses newly published primary literature about Mendelian diseases. If an article possibly contains a new diagnosis, the system sends a notification to a clinician. We test the accuracy of the system using a “turn back the clock” experiment: we train our system on data available up to 2011 and then test its performance on a large cohort of both singleton patients (where no sequencing data of relatives is available) and trios (where sequencing data of unaffected parents is available) who gradually became diagnosable between January 2012 and the present.

**Results:** Our system is capable of alerting clinicians of newly published, likely diagnostic, articles for a large fraction of undiagnosed Mendelian patients at a practicably low false positive rate. For example given trio data matching current reality, where only ~10% of cases in the undiagnosed patients database become diagnosable per year, a full 1/3 of our alerts offer timely diagnoses, for a yield of over 3/4 of cases.

**Conclusion:** We show that continuous automatic reanalysis of patients with suspected Mendelian disease is feasible in practice and has the potential to expedite the diagnosis rate of patients with suspected Mendelian disease. The new system will be launched at <https://amelie.stanford.edu/>

## PgmNr 1442: Prediction of four novel SNPs V17M, R11H, A66T, and F57S in the SBDS gene possibly associated with formation of Shwachman-Diamond syndrome, using an in silico approach.

### Authors:

A. Sati; S. Ali; R. Abdullallah; M. Elsied; M. Hassan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Applied Bioinformatics Department, Africa City of Technology, khartoum, Sudan

---

Shwachman-Diamond syndrome SDS (MIM 260400) is an autosomal recessive disorder. Characterized by exocrine insufficiency of the pancreas, bone marrow hypoplasia resulting in cytopenias, especially neutropenia, variable degree of skeletal abnormalities, failure to thrive, and increasing risk of developing myelodysplasia or transformation to leukemia. SDS is mainly caused by Shwachman Bodian Diamond Syndrome gene (MIM ID 607444). *SBDS* gene is 14899 bp in length, located in chromosome seven in the eleventh region of the long arm, and is composed of five exons. It's a ribosomal maturation factor which encodes a highly conservative protein that has widely unknown functions despite of its abundance in the nucleolus. A total number of 53 SNPs of homo sapiens *SBDS* gene were obtained from the national center for biotechnology information (NCBI) analyzed using translational tools, 7 of which were deleterious according to **SIFT** server and were further analyzed using several software's (**Polyhen-2, SNPs&Go, I-Mutant 2.0, Mutpred2, structural analysis software's and multiple sequence alignment software**). Four SNPs (rs11557408 (V17M), rs11557409 (R11H), rs367842164 (A66T), and rs376960114 (F57S)) were predicted to be disease causing, localized at highly conservative regions of the *SBDS* protein and were not reported in any previous study. In addition, the study predicts new functions of the *SBDS* protein DNA related, and suggests explanations for Patients developing cytopenias and failure to thrive through genetic coexpression and physical interaction with *RBF1* and *EXOSC3*, genes respectively.

## PgmNr 1443: Will variant matching improve the novel disease gene discovery rate?

### Authors:

F. Schiettecatte <sup>1</sup>; E. Wohler <sup>2</sup>; R. Martin <sup>2</sup>; A. Hamosh <sup>2</sup>; J. Posey <sup>3</sup>; C. Akdemir <sup>3</sup>; Z. Hande <sup>3</sup>; S. Jhangiani <sup>3</sup>; J.R. Lupski <sup>3</sup>; D. Valle <sup>2</sup>; N.L.M. Sobreira <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) FS Consulting LLP, Seattle, WA; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

---

Regardless of the sequencing strategy, the endgame for novel disease gene discovery comes down to identifying multiple affected individuals with similar phenotype and candidate variants in the same gene. For that end, the Baylor-Hopkins Center for Mendelian Genomics (BHCMG) has created GeneMatcher ([www.genematcher.org](http://www.genematcher.org)), a freely accessible website that allows investigators to post genes of interest automatically sending reciprocal emails to investigators who post the same gene. Subsequent follow-up is at the discretion of the submitters. There is also an option to match, based upon OMIM® number, genomic location, and on phenotypic features. The matches in GeneMatcher have supported more than 207 publications on novel disease genes and novel Mendelian diseases. More recently, we have created another data sharing tool, VariantMatcher ([variantmatcher.org](http://variantmatcher.org)), intended to share the data from the sequencing performed at the BHCMG with the public at the variant level. To comply with strict patient privacy and safety codes users of the site must register and be approved by site administrators. Users upload up to 10 genomic coordinates per day to the site and are notified if there is a match or not. The query format is “chr: coordinate refAllele>AltAllele” and is available for genomic builds hg18, hg19, or hg38. Phenotypic features can also be added to the query. When there is a match, both parties will be notified by a simultaneous email that contains the email address of both parties so they can exchange further information about their cases. If a match is not made the queried coordinates can be stored for future matching. The full dataset is not searchable. VariantMatcher contains the rare (MAF <1%), coding (synonymous not included), single nucleotide variants identified in 5,797 VCF files of affected and unaffected individuals sequenced by the BHCMG. As of June 6th 2019, VariantMatcher had 34 users from 13 countries and 4 variants had been matched. Future updates to VariantMatcher will include the capability to query databases such as CaféVariome, RD-Connect, and Illumina CaseLog through the Matchmaker Exchange API that we are developing now. With VariantMatcher we expect to increase our discovery rate by increasing the specificity of our matches, but, nevertheless, incomplete penetrance, variable expressivity of the phenotype, age of onset, and zygosity are some of the factors that should be considered when the phenotypes under investigation are being compared.

## PgmNr 1444: A computational approach for prioritizing potentially syndromic patterns of birth defects involving cleft lip and/or palate.

### Authors:

M.L. Navarro Sanchez <sup>1</sup>; R.H. Benjamin <sup>1</sup>; H. Chen <sup>1,2</sup>; L.E. Mitchell <sup>1</sup>; P.H. Langlois <sup>3</sup>; M.A. Canfield <sup>3</sup>; M.D. Swartz <sup>4</sup>; A.E. Scheuerle <sup>5</sup>; D.A. Scott <sup>6,7</sup>; H. Northrup <sup>8</sup>; C.P. Schaaf <sup>6,9,10</sup>; J.W. Ray <sup>11</sup>; S.D. McLean <sup>12</sup>; P.J. Lupo <sup>13</sup>; A.J. Agopian <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Epidemiology, Human Genetics and Environmental Sciences, UTHealth School of Public Health, Houston, TX.; 2) Center for Precision Health, UTHealth School of Public Health and UTHealth School of Biomedical Informatics, Houston, Texas; 3) Birth Defects Epidemiology and Surveillance Branch, Texas Department of State Health Services, Austin, Texas; 4) Department of Biostatistics and Data Science, UTHealth School of Public Health, Houston, Texas; 5) Department of Pediatrics, Division of Genetics and Metabolism, University of Texas Southwestern Medical Center, Dallas, Texas; 6) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 7) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas; 8) Department of Pediatrics, Division of Medical Genetics, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, Texas; 9) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas; 10) Heidelberg University, Institute of Human Genetics, Heidelberg, Germany; 11) Department of Pediatrics, Division of Medical Genetics and Metabolism, University of Texas Medical Branch, Galveston, Texas; 12) Clinical Genetics Section, The Children's Hospital of San Antonio, San Antonio, Texas; 13) Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston, Texas

**Background and Objectives:** Orofacial clefts are among the most common birth defects, and they can occur as part of a variety of syndromes and birth defect associations. However, the prevalence and type of co-occurring birth defects vary between studies, and it is likely that there are additional cleft syndromes that have not yet been described. The goal of this study was to investigate patterns of birth defects that co-occur with cleft lip and/or cleft palate, using an approach capable of assessing high order combinations based on data from a population-based registry. Since other studies have focused on pairwise birth defect combinations and clinical populations, our findings could contribute towards identifying new syndromes and elucidating the genetic etiologies of these birth defects.

**Methods:** We analyzed data from the Texas Birth Defect Registry on 10,011 infants with cleft lip and/or cleft palate delivered between 1999 and 2014 and excluded cases with syndrome diagnoses. We used our new software platform, Co-Occurring Defect Analysis (CODA), to assess all possible 2-, 3-, 4-, and 5-way combinations of major birth defects among these cases by calculating an adjusted observed:expected ratio for each defect combination.

**Results:** We excluded 929 (27%) cases with cleft palate (CP) and 937 (14%) cases with cleft lip with or without cleft palate (CL/P) with syndrome diagnoses. The prevalences of non-syndromic CP and CL/P were estimated as 4.0 and 9.2 cases per 10,000 births, respectively. Approximately 29% of

infants with CP and 24% of infants with CL/P had at least one additional congenital anomaly, while 7% of cases with CP and 5% of those with CL/P had five or more birth defects. For both CP and CL/P, the co-occurring birth defects with the highest observed:expected ratios were in the central nervous and ocular systems. CP was also associated more often than expected with cardiovascular defects, such as VSD.

**Conclusion:** Our study identified patterns of multiple birth defects that co-occur with CP and CL/P more often than expected. An analysis of these high order combinations may lead to the recognition of new syndromic patterns, and a better understanding of the genetic etiologies and developmental mechanisms that underlie the respective birth defects.

## PgmNr 1445: CODA: A platform for assessing patterns of multiple congenital anomalies.

### Authors:

R.H. Benjamin<sup>1</sup>; X. Yu<sup>1,2</sup>; M.L. Navarro Sanchez<sup>1</sup>; H. Chen<sup>1,3</sup>; L.E. Mitchell<sup>1</sup>; P.H. Langlois<sup>4</sup>; M.A. Canfield<sup>4</sup>; M.D. Swartz<sup>2</sup>; A.E. Scheuerle<sup>5</sup>; D.A. Scott<sup>6,7</sup>; H. Northrup<sup>8</sup>; C.P. Schaaf<sup>6,9,10</sup>; J.W. Ray<sup>11</sup>; S.D. McLean<sup>12</sup>; P.J. Lupo<sup>13</sup>; A.J. Agopian<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Human Genetics, and Environmental Sciences, UTHealth School of Public Health, Houston, TX; 2) Department of Biostatistics and Data Science, UTHealth School of Public Health, Houston, TX; 3) Center for Precision Health, UTHealth School of Public Health and UTHealth School of Biomedical Informatics, Houston, TX; 4) Birth Defects Epidemiology and Surveillance Branch, Texas Department of State Health Services, Austin, TX; 5) Department of Pediatrics, Division of Genetics and Metabolism, University of Texas Southwestern Medical Center, Dallas, TX; 6) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 7) Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TX; 8) Department of Pediatrics, Division of Medical Genetics, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX; 9) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX; 10) Heidelberg University, Institute of Human Genetics, Heidelberg, Germany; 11) Department of Pediatrics, Division of Medical Genetics and Metabolism, University of Texas Medical Branch, Galveston, TX; 12) Clinical Genetics Section, The Children's Hospital of San Antonio, San Antonio, TX; 13) Department of Pediatrics, Section of Hematology-Oncology, Baylor College of Medicine, Houston, TX

**Introduction:** About 3% of liveborn infants in the United States are affected by a birth defect, and up to one-third of those infants will have multiple defects. Identification of syndromes, sequences, and associations has previously relied on clinicians and researchers recognizing patterns occurring across patients. We present an alternative approach that is based on statistical analysis of large-scale data from birth defect registries.

**Methods:** Our approach involves assessing combinations of birth defects that occur more frequently than expected, using adjusted observed-to-expected ratios for all possible combinations of birth defect categories. To facilitate these analyses, we developed an open source R-based platform, Co-Occurring Defect Analysis (CODA), to systematically evaluate 2-, 3-, 4-, and 5-way combinations of defects observed in a dataset. CODA calculates an adjusted observed-to-expected ratio for a combination of birth defects that accounts for nonspecific co-occurrence of anomalies. To demonstrate feasibility, we analyzed data from the Texas Birth Defects Registry (1999-2014) using a Dell OptiPlex 7040 x64-based PC (Intel Core i7-6700 CPU, 3.40 GHz, 4 core, 16 GB RAM) running Windows 10.

**Results:** Using CODA to analyze Texas Birth Defect Registry data (175 birth defects in 206,784 cases), adjusted observed-to-expected ratios were calculated for 8,455 2-way combinations, 105,412 3-way combinations, 528,287 4-way combinations, and 1,473,775 5-way combinations. CODA ran efficiently, outputting results for 5-way combinations in 18.2 hours.

**Conclusion:** CODA can facilitate large-scale analyses of co-occurrence patterns in birth defects registries. Recognition of novel patterns of multiple anomalies may indicate a shared genetic basis



that should be further investigated.

## **PgmNr 1446: CAGI SickKids challenges: Assessment of bioinformatic strategies for predicting patients' phenotypes and identifying potential pathogenic variants and disease genes from their genomes.**

### **Authors:**

M. Meyn <sup>1,2</sup>; L. Kasak <sup>3,4</sup>; J.M. Hunter <sup>5</sup>; R. Udani <sup>5</sup>; Z. Hu <sup>3</sup>; C. Bakolitsa <sup>3</sup>; S.E. Brenner <sup>3</sup>; The CAGI SickKids challenge predictors

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Human Genomics and Precision Medicine, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin.; 2) Department of Paediatrics, The Hospital for Sick Children, Toronto, Canada; 3) Department of Plant and Microbial Biology, Univ of California, Berkeley, California; 4) Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia; 5) Department of Paediatrics and Wisconsin State State Laboratory of Hygiene, University of Wisconsin, Madison, Wisconsin

---

Most patients who undergo whole genome sequencing (WGS) lack a molecular diagnosis, largely due to the large number of undiscovered disease genes and inability to assess the pathogenicity of most genomic variants. The CAGI SickKids challenges were public competitions designed to address this knowledge gap by assessing state-of-the-art methods for predicting clinical phenotypes from genomes.

CAGI4 and CAGI5 participants were provided with WGS data and clinical descriptions for 25 and 24 patients respectively who had undergone WGS but remained undiagnosed after evaluation by the SickKids Genome Clinic project (NPJ Genom Med 2016 13:12016 and Genet Med 20:435 2018). Predictors were asked to identify primary and secondary causal variants. Additionally, for CAGI5, groups attempted to match each genome to one of three classes of disease (neurologic, ophthalmologic and connective), and separately to each patient. A team of clinical and molecular geneticists assessed group submissions.

The four groups participating in CAGI4 proposed 191 diagnostic variants for 25 patients. Most variants were nominated by only one group and did not fully explain the patient's phenotype. However, two of the ten variants proposed by two groups were deemed to be diagnostic.

The CAGI5 challenge was more difficult, as genomes were not linked to patients' clinical descriptions. The eight groups participating in CAGI5 all prioritized how well a candidate gene explained a phenotype over the pathogenicity of its variants. They did no better than chance in matching genomes to the three disease categories but two groups performed significantly better than chance in matching genomes to specific patients. Importantly, patients with ophthalmologic disorders were the most likely to be matched to their genomes and had the strongest correlation between the informational content of the clinical description and their chance of matching. Despite partial success in matching genomes to patients, no candidate variants in CAGI5 were nominated by more than one

group and none were deemed to be clinically diagnostic. However, several nominated variants of unknown significance are candidates for phenotype expansion.

Despite their clinical naïveté, the bioinformaticians solved two cases and identified multiple plausible causal variants in candidate disease genes that bear further evaluation. We will discuss implications for improving *in silico* assessment of genomic variants and identifying new disease genes.

## PgmNr 1447: WebSTR: An online database of tandem repeat variants in human populations.

**Authors:**

R. Yanicky; M. Gymrek

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of California, La Jolla, CA.

---

Tandem repeats (TRs) consist of repeated motifs of 1 or more base pairs and comprise more than 2 million loci in the human genome. Due to their rapid mutation rates, TRs represent one of the largest sources of genetic variation. TRs have been shown to contribute to a variety of human diseases, including Mendelian disorders such as Huntington's Disease or Fragile X Syndrome, as well as more complex traits such as schizophrenia or cancer risk. We and others have generated population-wide catalogs of TR genotypes at more than 1.5 million loci across more than 5,000 individuals from diverse ancestry groups. We have additionally mapped associations between TR repeat number and a variety of traits including gene expression, height, and schizophrenia.

To maximize the use of these datasets, we have developed WebSTR, an online database of tandem repeat variants in human populations. WebSTR is built as a flask application that leverages an underlying sqlite database and provides interactive visualizations using the plot.ly library. It incorporates genome-wide TR genotypes from multiple cohorts, including individuals from the 1000 Genomes Project and Genotype-Tissue Expression Project. Users may browse for TRs by gene or genomic region. For each TR, WebSTR displays population-specific allele frequencies as well as summary information about phenotype associations, imputation quality, and estimated mutation rates. In future work, we are adding more cohorts and incorporating additional modules for interactively exploring TR association summary statistics.

The WebSTR platform can be expanded to include any number of cohorts and integrate results from diverse areas of genetics. This will allow researchers across many disciplines to be able to easily access and re-analyze existing data.

WebSTR can be accessed at [webstr.gymreklab.com](http://webstr.gymreklab.com)

## **PgmNr 1448: Detecting phenotype descriptors in clinical notes with ClinPhen: An assessment of sensitivity in a PhenoDB patient cohort.**

### **Authors:**

C. Taylor; M. Rouhizadeh; H. Levy; A. Baras; A. Hamosh; N. Sobreira

[View Session](#) [Add to Schedule](#)

**Affiliation:** Johns Hopkins University

---

The precise clinical and genomic diagnosis of a patient depends on the complete and accurate description of his/her features. Currently physicians assemble the patients' phenotypic features through the clinical and family history, the physical examination, laboratory and imaging results and through manual curation of prior clinical notes. The manual curation of prior notes can be a time-consuming process that is subject to inaccuracies and omissions. At the Genetics clinic of Johns Hopkins Hospital (JHH), we use PhenoDB, a secure Web-based portal, to enter and analyze the phenotypic and genomic information of our patients. In PhenoDB, the physicians described the phenotypic features of their patients using the PhenoDB terminology that includes 3,666 terms mapped to Human Phenotype Ontology (HPO), Elements of Morphology, and International Consortium of Human Phenotype Terminologies (ICHPT) terms. Based on the phenotype descriptor terms entered for each patient, PhenoDB generates a list of the 20 most likely diagnoses through an OMIM search, alerts the physicians if a gene known to cause any of these 20 possible diagnoses is found mutated in the patient's genomic data, and allows the search of other patients with similar phenotype descriptors in the database. The work here explored the use of PhenoDB to assess the sensitivity of ClinPhen to annotate the clinical notes of patients diagnosed with genetic disorders at our Genetics clinic. ClinPhen (<http://bejerano.stanford.edu/clinphen/>) is a concept recognition software that annotates free-text documents with phenotype descriptor terms from HPO. The number of notes per patient ranged from one to 119, with 32 notes on average (SD = 42). The number of unique HPO terms annotated by ClinPhen per patient ranged from 17 to 140, with 60 terms on average (SD = 40). One patient did not have any HPO terms recorded in PhenoDB and was excluded from the analysis of sensitivity. We found an average sensitivity of 71% (SD=38%) for ClinPhen to detect the HPO terms that were captured manually in PhenoDB. Our findings indicated a large variability in the per patient number of notes, number of HPO terms annotated, and sensitivity of ClinPhen to capture HPO terms for patients in PhenoDB. Further work is needed to determine the range and types of notes and patients that are optimal to enable computational phenotype extraction for diagnostic purposes.

## **PgmNr 1449: Validation of a clinical germline exome sequencing assay and bioinformatics pipeline at UPMC Genome Center.**

**Authors:**

R.O. Rosti; X. Zhuo; C. Sims; R. Stewart; L. Baumgart; A. Berndt

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genome Center, UPMC, Pittsburgh, PA

---

The University of Pittsburgh Medical Center (UPMC) Genome Center is a CLIA Certified CAP Accredited high-throughput industrial scale sequencing center providing high quality, clinical grade next-generation sequencing services to internal stakeholders and external clients. Exome sequencing for rare-disease, Mendelian disorders and other germ-line traits is experiencing increased clinical acceptance and adoption, due to faster turn-around times, and advances in cloud-based bioinformatics pipelines. Furthermore, whole-exome sequencing provides a cost-effective alternative to whole-genome sequencing when economies of scale and throughput is taken into consideration. Recently, the UPMC Genome Center, has validated an exome sequencing assay and bioinformatics pipeline to be used in newborn screening and for various Mendelian disorder panels. Here we present data from our extensive validation sets for whole exome germline sequencing assay and bioinformatics pipeline.

## **PgmNr 1450: Retrospective analysis of 634 rare and undiagnosed genetic disease cases tested by clinical whole genome sequencing identifies steps to reduce number of manually triaged variants.**

### **Authors:**

V.G. Gainullin <sup>1</sup>; A. Chawla <sup>1</sup>; V. Rajan <sup>1</sup>; S. Ajay <sup>1</sup>; C.-H. Lin <sup>1</sup>; R.T. Hagelstrom <sup>1</sup>; A. Warren <sup>1</sup>; J.D. Avecilla <sup>1</sup>; D.L. Perry <sup>1</sup>; D.R. Bentley <sup>2</sup>; R.J. Taft <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Clinical Genomics Research Dept, Illumina, Inc. San Diego, CA, USA; 2) Clinical Genomics Research Dept, Illumina Cambridge Ltd. Chesterford Research Park, Little Chesterford, UK

---

Clinical whole genome sequencing (cWGS) detects a greater range of variation than targeted panels or whole exome sequencing. As the turn-around time of sequencing and secondary analysis continues to shorten, interpretation remains a significant burden. Reducing the number of variants for manual review by a clinical laboratory case analyst is one area of focus to streamline cWGS analysis.

Here we performed a retrospective data analysis of variant triage decisions of 634 rare and undiagnosed disease cWGS cases analyzed over a two-year period at the Illumina Clinical Services Laboratory. Across the 634 cWGS cases, a total of 108,315 variants were triaged. Family structures included proband-only (n=67), duo (n=107), trio (n=436), quad (n=22), and quint (n=2). Pertinent annotations as well as variant-level and gene-level comments supporting triage decisions reflecting phenotype overlap, gene-disease association, allele frequency among others were aggregated and analyzed. We found that only 2.8% (n = 2,969, average 4.6 variants per case) were selected for follow-up and reviewed by the clinical team, while the remaining (163 variants per case) were excluded based on at least one of four major categories: variants in genes without gene-disease association (no-gda), pipeline artifact variants which escaped population-based filtering (artifacts), variants in recessive genes lacking a second allele (AR no 2nd hit), and variants in genes which did not overlap with patient phenotypes. We show that phenotype-based filtering is a powerful triage technique and allows the analyst to prioritize variants for review, however with limitations. The remaining triage decision categories (no-gda, AR no 2nd hit and artifacts) accounted for 50% of triage decisions and represent candidates for full triage automation.

Tracking case analyst input allowed us to optimize the analysis where manual triage work may be significantly reduced.

## PgmNr 1451: Phen2Gene: Rapid phenotype-driven gene prioritization for rare diseases using human phenotype ontology terms.

### Authors:

M. Zhao <sup>1</sup>; L. Fang <sup>1</sup>; Y. Chen <sup>1</sup>; C. Liu <sup>2</sup>; C. Wu <sup>3</sup>; M. Sarmady <sup>3,4</sup>; P. Botas <sup>5</sup>; J. Isla <sup>5,6</sup>; G. Lyon <sup>7</sup>; C. Weng <sup>2</sup>; K. Wang <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032, USA; 3) Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 4) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA; 5) Foundation 29, Pozuelo de Alarcon, Madrid, Spain; 6) Orphan Drug Committee, European Medicines Agency, The Netherlands; 7) Institute for Basic Research in Developmental Disabilities (IBR), Staten Island, New York, USA

---

**Background:** Incorporation of clinical phenotypes can greatly facilitate interpretation of whole genome or exome sequencing data to prioritize candidate genes for individuals with Mendelian diseases. Human Phenotype Ontology (HPO) terms are increasingly used in diagnostic settings to characterize the clinical phenotypes of patients, so methods that leverage HPO terms can improve diagnostic sequence interpretation. Although existing HPO annotations map each phenotype term to multiple candidate genes, they do not provide a means to weight/rank the genes or combine multiple phenotype terms for each patient. To address these challenges, here we first compile a knowledgebase (HPO2Gene-KB) that associates each HPO term with a list of prioritized genes, each with a weight indicating the strength of the relationships. We next present Phen2Gene, a rapid computational approach that searches and prioritizes candidate genes from a list of HPO terms.

**Methods:** Existing HPO annotations, five disease-gene relationship databases, and four gene-gene relationship databases were incorporated to compile HPO2Gene-KB. Given an individual with suspected Mendelian diseases, Phen2Gene takes a list of HPO terms as input, pre-processes the HPO terms based on ontology-based information content, and uses a weighted approach to identify and rank candidate genes for the patient.

**Results:** HPO2Gene-KB stores much more phenotype-gene relationships than existing HPO annotations, and it provides weights representing the strength of phenotype-gene relationships. We tested the performance of Phen2Gene using five cohorts of clinical cases totaling 281 individuals with a known genetic diagnosis, and compared the rankings of causal genes to existing gene prioritization tools such as Phenolyzer. Among the 281 individuals, Phen2Gene ranked causal genes within top 100 in 60 individuals, while Phenolyzer did it in 37 individuals. In summary, Phen2Gene has comparable performance to Phenolyzer, but it is ~100X faster and can generate results instantly.

**Conclusion:** HPO2Gene-KB fills a void in the community in linking standardized phenotype terms to



genes with weighted scores, and it may facilitate the development of novel computational tools that link HPO to disease genes. As a special use case of HPO2Gene-KB, Phen2Gene serves as a real-time phenotype driven gene prioritization tool to aid clinical diagnosis of rare diseases.

## PgmNr 1452: Context matters: An approachable web tool for inspecting variants in genetic disorders.

### Authors:

T. Di Sera <sup>1</sup>; A. Ward <sup>1,2</sup>; M. Velinder <sup>1</sup>; C. Miller <sup>1,2</sup>; Y. Qiao <sup>1</sup>; G. Marth <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Human Genetics, University of Utah, Salt Lake City, Utah.; 2) Frameshift Genomics, Boston MA

---

Prioritizing variants in suspected genetic disorders involves taking into account several important factors. Here we show *gene.iobio*, a web tool for variant inspection that can quickly and easily provide important context and insight during variant prioritization.

During a typical variant prioritization workflow, an analyst will consider a number of criteria for a variant, including population frequency, inheritance mode and allele segregation, variant pathogenicity assertions, and known gene-disease associations. Our *gene.iobio* tool annotates variants in real-time, allowing an analyst to consider these important criteria. Often, for well described genetic disorders these criteria are sufficient to reach a confident genetic diagnosis and *gene.iobio* can return that genetic diagnosis in seconds.

However, sometimes after these considerations have been made, a genetic diagnosis is still lacking. Here an analyst will typically explore variants of uncertain significance and inspect data quality for potential insufficiencies. Our *gene.iobio* tool allows the analyst to continue their exploration into variants of uncertain significance by providing them with additional variant information such as gene burden, amino acid conversation, additional disease/molecular associations and exons where previously reported pathogenic variants reside. Our tool also provides an analyst with interactive and intuitive visualizations to explore data quality and give important context to each variant. Using gene coverage tracks and read pileup views, an analyst can spot regions of insufficient coverage or regions of poor alignment and account for these regions when considering modes of inheritance. For instance, an analyst would likely dismiss a variant reported as de novo, where neither parent has coverage at that region. Additionally, alternate/reference read counts are often not clear 0/100, 50/50 or 100/0 splits. *Gene.iobio* provides the analyst with read count information, allowing for more accurate assessment of inheritance mode. For instance, an analyst would likely be suspicious of a recessive variant if one parent had < 10% alternate reads observed. All of these features make *gene.iobio* a powerful web tool for rapidly reaching a diagnosis for well described genetic disorders, but also provides an avenue of continued inquiry when such an obvious answer is not reached.

## PgmNr 1453: A scalable framework for collaborative rare disease diagnosis and gene discovery.

### Authors:

H. Snow<sup>1,2</sup>; M. Solomonson<sup>1,2</sup>; K. Nguyen<sup>1,2</sup>; S. Baxter<sup>1,2</sup>; A. O'Donnel-Luria<sup>1,2,3</sup>; A. Lovgren<sup>1,2</sup>; H. Arachchi<sup>1,2</sup>; H. Rehm<sup>1,2</sup>; D. MacArthur<sup>1,2</sup>; B. Weisburd<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 2) Analytic and Translational Genomics Unit (ATGU), Massachusetts General Hospital, Boston, MA, USA; 3) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA

---

Rapid advances in exome and whole-genome sequencing make it possible to identify the full spectrum of genetic variation in rare disease patients and family members. However, powerful filtering and decision support tools are needed in order to discover the causal variants amidst the noise of benign genetic variation. To address this challenge we developed seqr - an open source, interactive web-based platform that allows researchers to collaboratively search and annotate variant callsets from exome and whole genome studies involving families of rare disease patients. In our framework, variants are annotated with allele frequencies from population databases including 1000 Genomes Project, TOPMed, and gnomAD, transcript consequence predictions, clinically interpreted variant databases including ClinVar, and in silico predictor scores. Variants are filtered with these criteria as well as family-specific inheritance mode, and individual readviz data is available in the browser for all variants. Gene-level annotations and resource links such as OMIM, GTEx and UCSC genome browser provide researchers with a flexible framework for efficiently exploring the evidence for causality for candidate variants and genes. seqr is the core analysis platform for the Broad Institute's Center for Mendelian Genomics, having already been used by over 500 researchers to study potential causal variants across over 10,000 families affected by rare disease and facilitating over 1,000 diagnoses, many of them in novel disease genes.

Here, we will describe the framework underlying seqr with a focus on the design choices made to scale the system to tens of thousands of exome and genome samples. Additionally, we will describe how interactive visualization of read-level data, detailed phenotype data, external reference data, and other relevant data types were brought together in a single user-friendly interface. The ability to interactively search across terabytes of genomic sequencing data requires custom solutions and thoughtful design trade-offs, and we will present the details of how seqr dealt with these challenges.

## PgmNr 1454: Accounting for splicing effects in known missense variants improves *in silico* prediction of deleterious effect.

### Authors:

D. Qian; J. Clifford; A. Tchourbanov; Y. Tian; M.H. Black; H.M. Lu; Z. Zhu; S. Li

[View Session](#) [Add to Schedule](#)

**Affiliation:** Ambry Genetics, Aliso Viejo, California.

---

To characterize missense variation across the human genome, various *in silico* predictors exist for assessing deleterious effects from a combination of sequence, structural and evolutionary features. We previously showed that an ensemble *in silico* variant prediction (IVP) score from a Bayesian logistic regression model had high degree of prediction accuracy. In this study, an ensemble score is first estimated in a subset of 15 *in silico* predictors (including SIFT, PolyPhen2, PhyloP, PROVEAN and others), and then combined with a splicing impact prediction to form a joint score. The splicing impact prediction score was obtained from a previously developed tool, SpliceScan II, which combines *ab initio* splicing prediction with evolutionary conservation in the context of known gene structures to assess variant impact on splicing. We derived an ensemble score without accounting for potential splicing (IVPns; IVP with no splicing predictor), and the score accounting for potential splicing (IVPws; IVP with splicing predictor). We then compared the IVPws and IVPns scores in known (likely) pathogenic and (likely) benign missense variants with ClinVar classifications, defined as the most supported category among six submitters providing assertion criteria. For evaluating and reporting prediction performance, we separated the missense variants in 338 genes into two variant sets based on SpliceScan II at a cutoff 0.05 determined by the distribution of classified variants: 1) 9,761 variants with SpliceScan II  $\leq 0.05$  for no splicing impact; 2) 424 variants with SpliceScan II  $> 0.05$  for potential splicing impact. The prediction performance of IVPws and IVPns scores were compared within each of the two variant sets by area under the curve (AUC) statistic using direct cross validation. For variants in 1) with no predicted splicing impact, the AUCs of IVPws and IVPns were equivalent (0.9761 and 0.9756;  $p = 0.58$ ), showing that including splicing score in IVP model didn't impair the prediction performance. For variants in 2) with potential splicing impact, the AUC of IVPws was 4.8% higher than that of IVPns (0.971 vs. 0.927;  $p = 0.01$ ). To conclude, for missense variants with potential splicing impact, ensemble scores that incorporate splicing information improve *in silico* prediction of deleterious effects.

## PgmNr 1455: Full-length RNA-seq of Alzheimer brain sample on the PacBio Sequel II platform.

### Authors:

E. Tseng <sup>1</sup>; T. Hon <sup>1</sup>; B. Galvin <sup>1</sup>; D. Pinto <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pacific Biosciences, 1305 O'Brien Drive, Menlo Park, CA 94025, USA; 2) Psychiatry, and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, NY 10029, USA

---

Alzheimer's disease (AD) is a devastating neurodegenerative disease that affects ~44 million people worldwide, making it the most common form of dementia. It has been shown that changes in gene expression due to alternative splicing likely contribute to the initiation and progression of AD. Most transcriptome studies conducted to date have used RNA-seq short read data, which provides quantification at the gene level, but cannot resolve complex isoform structures. The PacBio Iso-Seq method provides full-length, highly accurate transcripts of 10kb or longer with no assembly required, allowing unambiguous characterization of alternative splicing events, TSS usage, and allele-specific, isoform-specific phasing information. We sequenced the whole brain of a single male with AD on the PacBio Sequel II platform. Using a single SMRT cell, we obtained 4,277,293 full-length reads, which was processed using the Iso-Seq bioinformatics pipeline followed by a quality filtering tools SQANTI2 that removes cDNA library artifacts, resulting in a final set of 160,837 unique polyadenylated isoforms covering 17,546 genes ranging from 100bp-15kb (mean: 3.4kb). Compared with Gencode v29, 67% of the isoforms detected are novel, however the majority of the novelty comes from novel combinations of known donor-acceptor sites. The majority of novel isoforms that use at least one novel donor-acceptor site, however, all still have coding potential and are likely to be bonafide novel transcripts. As such, we show that the transcriptional landscape of AD is more complex than previously appreciated, and that the PacBio Sequel II system is a cost-effective solution for sequencing whole transcriptome in human disease research.

## PgmNr 1456: Prioritizing rare variants in non-coding WGS: Experience with the ADSP family sample.

### Authors:

T.R. Day <sup>1</sup>; E.E. Blue <sup>1</sup>; C. Funk <sup>2</sup>; P. Shannon <sup>2</sup>; R. Nafikov <sup>1</sup>; A.R. Horimoto <sup>1,3</sup>; H.K. Sohi <sup>4</sup>; B. Wang <sup>5</sup>; M.O. Dorschner <sup>6</sup>; D. Tsuang <sup>7</sup>; P. Navas <sup>1</sup>; T.A. Thornton <sup>3</sup>; E.M. Wijsman <sup>1,3</sup>; ADSP

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, University of Washington, Seattle, WA.; 2) Institute for Systems Biology, Seattle, WA; 3) Department of Biostatistics, University of Washington, Seattle, WA; 4) Department of Biomedical Informatics, University of Washington, Seattle, WA; 5) Department of Statistics, University of Washington, Seattle, WA; 6) Department of Pathology, University of Washington, Seattle, WA; 7) Veteran's Administration, VA Puget Sound, Seattle WA

---

Complex genetic traits such as late-onset Alzheimer's disease (AD) are highly genetically heterogeneous. Although both family studies and GWAS have implicated multiple loci, the underlying causal variants are largely unknown. Availability of whole-genome-sequence (WGS) now allows evaluation of virtually all DNA in a region of interest, including both common and rare variants. Causal common variants can be implicated by GWAS, but rare variants may be more effectively studied through family designs. We carried out linkage analysis in 110 individual families in the AD sequencing project (ADSP), followed by family-specific association analysis and bioinformatic filtering in the regions of interest (ROIs) implicated by  $\text{lod} > 1.9$ . These variants were further filtered based on population  $\text{MAF} < 0.01$  and family-specific p-value thresholds from family-based association, resulting in 17,765 variants across 1,312 genes.

A major challenge was that most variants implicated fell in the non-coding DNA. To further prioritize variants, we developed a Comparative Tissue-Specific Expression Annotator (CTEA) to use outside databases and estimate the relative importance of a gene tagged to a variant, based on the specificity of gene expression across all tissues available and compared to all genes in a specific tissue. For our purposes, we used the NCBI Gene Expression Omnibus (GEO) and the Genotype-Tissue Expression (GTEx) datasets in order to estimate these measures, although similar measures may be developed for other datasets.

From the filtered association results, we took the top 20% for the NCBI importance in brain and the GTEx importance for three AD-relevant brain regions. We isolated genes that passed all of these filters and were also identified by TReNA, a machine learning algorithm for identifying transcriptional regulatory networks. Following this, we applied a filter to the variants using functional inference (FIRE) scores ( $\geq 0.8$ ) based on a random forest classification relating to regulatory expression. Of the 9,152 variants across 65 genes identified through in these previous filters, 45 genes have 788 variants that pass the FIRE filter. We are in the process of adding a final filter based on recently released single-cell transcriptomic results in conjunction with AD progression. These results show that there may be statistically-driven approaches to prioritize non-coding variants that make use of outside bioinformatic data.

## PgmNr 1457: Structural characterization of rare missense variants within known neurodegenerative disease proteins.

### Authors:

E.L. Palmer <sup>1,2</sup>; C. Moth <sup>3,4</sup>; N. Wheeler <sup>1,2</sup>; J. Sheehan <sup>3,4</sup>; T. Capra <sup>3,4</sup>; W. Bush <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH.; 2) Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH.; 3) Center for Structural Biology, Vanderbilt Genomics Institute, Nashville, TN; 4) Department of Biological Sciences, Vanderbilt University, Nashville, TN

---

Alzheimer's disease (AD) is the 6<sup>th</sup> leading cause of death in the US and currently remains untreatable. Recent work suggests unexplained heritability of AD may involve rare variants in genes implicated in other neurodegenerative disorders. However, existing gene-based tests have insufficient power to detect true associations when neutral variants are included. Therefore, it is critical to identify and prioritize variants with high impact to protein function before attempting gene-based statistical assessments. To that end, we employed the Phenotype Consensus ANalysis package (PCAN) in R to identify genes with Mendelian associations to neurodegenerative disorders. We examined the location of missense variants from the Alzheimer's Disease Sequencing Project (ADSP) in 3D protein structures. We compared to the locations of known variants curated from ClinVar (pathogenic variants) and ExAC (presumed benign variants) using the PathProx algorithm. Additionally, we computationally predicted the impact of each missense variant on protein folding (vs. canonical wildtype amino acid) via Rosetta's change in free energy (ΔG) algorithm. We structurally assessed protein products of the top 1% of ranked genes from PCAN (n=31). We excluded genes lacking ADSP variants (n=3) or classified as lncRNAs (n=2). The remaining genes (n=26) are presently being assessed, with preliminary findings including results from >50% of them. The number of ADSP variants per gene varied (mean 59.1, min=8, max=190). From the completed analyses, the mean number of variants per gene with PathProx scores suggestive of functional consequence was 13.6 (min=0, max=37). The mean number of variants per gene with a ΔG suggestive of a stabilizing function was 2.3 (min=0, max=16). The mean number of variants per gene with ΔGs suggestive of a destabilizing function was 10.7 (min=0, max=40). We have identified subsets of missense variants likely to impact the function of proteins associated with neurodegenerative diseases in the ADSP. These annotations can be used in multiple ways, such as producing variant sets for use in gene-based testing. Additionally, Functional characterization of rare variants using PathProx and ΔG could aid in heterogeneity testing of multifactorial diseases like AD. Finally, our methodology is phenotype-centric and applicable to other diseases with overlapping phenotypes to identify candidate genes in a variety of complex disorders.

## PgmNr 1458: Integration of interactome perturbation and cell-type-specific transcriptome implicates novel autism risk genes and cell types where they are expressed.

### Authors:

S. Chen <sup>1,2</sup>; J. Wang <sup>3</sup>; A.E. Cicek <sup>4,5</sup>; B. Devlin <sup>6</sup>; H. Yu <sup>1,2</sup>; K. Roeder <sup>3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computational Biology, Cornell University, Ithaca, NY 14853, USA; 2) Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY 14853, USA; 3) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213, USA; 4) Computer Engineering Department, Bilkent University, Ankara 06800, Turkey; 5) Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA; 6) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA

---

While large-scale sequencing has led to the discovery of more than 100 risk genes for autism spectrum disorder (ASD), fundamental questions remain. For example, regarding these risk genes, when and in what cell types are their gene products expressed? The answer would provide clues to how mutation of these genes alters typical neurodevelopment and neurophysiology. In addition, do these genes lend insight into what other genes are likely involved in development of ASD? Utilizing genomic data from a recent whole-exome sequencing study from the Autism Sequencing Consortium, we investigate the functional relevance of missense mutations through integrated interactomic and transcriptomic analyses. We use three resources for these analyses: a human interactome network compiled from eight widely used interaction databases; the BrainSpan transcriptome resource for the developing brain; and estimated cell-type-specific (CTS) expression from analysis of BrainSpan using the MIND algorithm. Interactome interface mapping reveals significantly higher interaction-disruption rate in ASD probands than in unaffected siblings (1.8-fold), and interaction disruptions in probands tend to target known ASD genes (35-fold). Using pairs of genes whose interaction is found to be disrupted in ASD subjects, we construct a PPI network. In this “disrupted network”, we find strong enrichment of ASD genes as hubs; genes closely connected to these hubs are more likely to be ASD genes themselves using the guilt-by-association principle; and the missense mutations that disrupt hub interactions are more likely to be meaningful for ASD than the larger set of non-disruptive missense mutations. Our CTS transcriptomic analysis identifies key cell types for ASD risk based on when and where genes are expressed and co-expressed. Finally, incorporating genome, transcriptome, and interactome information into a unified statistical framework implicates hundreds of genes as candidate ASD risk genes. These genes are highly intolerant to loss of function variation and ASD subjects harboring mutations of these genes exhibit more severe phenotypes. Taken together, our study facilitates translating the wealth of genomic data into biological and etiological understanding for ASD.



## PgmNr 1459: *Straglr*: Short tandem repeat genotyping using long-reads.

### Authors:

R. Chiu <sup>1</sup>; I.S. Rajan-Babu <sup>2,3</sup>; J.M. Friedman <sup>2,3</sup>; I. Birol <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada; 2) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 3) Children's and Women's Hospital, Vancouver, BC, Canada

---

Short tandem repeats (STR) are tandem arrays of 1 to 6 base pair (bp) repeat motifs. STR expansions are known to underlie the molecular pathogenesis of over 40 diseases, most of which involve the nervous system. Southern blot (SB) analysis, the gold standard molecular diagnostic test for detecting STR hyperexpansions, is labor- and time-intensive. Alternatively, short reads from next-generation sequencing (NGS) technologies with analysis by recently-developed algorithms can successfully detect expansions of some known disease STRs. However, large STR expansions that exceed sequencing read (~125-150 bp) or fragment (~350-500 bp) lengths still pose challenges for methods that use NGS data. Furthermore, GC-bias inherent in short reads presents a hurdle in genotyping GC-rich STRs, such as the *FMR1* CGG-repeat expansion mutation, which causes fragile X syndrome. Third-generation sequencing technologies, such as those from Nanopore and PacBio, are appealing alternatives due to their longer read lengths and lack of significant sequencing bias. A number of long-read STR analysis algorithms such as RepeatHMM and tandem-genotypes have demonstrated reasonable accuracy in genotyping long repeat expansion alleles in spite of the generally inferior single base accuracy of long reads relative to NGS data. However, existing tools have room for improvement in robust genotyping of STRs with different repeat motifs or lengths and in scaling for genome-wide scans for novel STR expansions.

Here we present *straglr*, a versatile software package that can be used for both identifying novel STR expansions and genotyping known disease loci. *Straglr* captures novel STR expansions by detection of sizeable insertions from minimap2 alignments of long reads and uses Tandem Repeats Finder for identification of STRs in genotyping either detected or pre-defined loci. We use Genome in a Bottle datasets to demonstrate the utility of *straglr* in performing genome-wide scans to reliably detect polymorphic expanded alleles at STR loci with a wide spectrum of repeat lengths and motifs. We also use long-read data of targeted disease loci to show *straglr* can estimate the size of expanded alleles at known STR disease loci with accuracy comparable to that of PCR or SB analysis.

## PgmNr 1460: Comparison of the gene expression between microglia and monocytes in epilepsy patients and their association with clinical variables.

### Authors:

P. Gaur <sup>1</sup>; M. Olah <sup>1</sup>; R. Sarkis <sup>2</sup>; P. Nejad <sup>3</sup>; G.R. Cosgrove <sup>2</sup>; J. Helgager <sup>2</sup>; J.A. Golden <sup>2</sup>; P.B. Pennell <sup>2</sup>; P.L. De Jager <sup>1</sup>; H.U. Klein <sup>1</sup>; E.M. Bradshaw <sup>1</sup>; V. Menon <sup>1</sup>; W. Elyaman <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Columbia University Medical Center, NY, USA; 2) Brigham and Women's Hospital: Boston Hospital & Medical Center; 3) Broad Institute of MIT and Harvard, Cambridge, MA, USA

---

**Background/Objectives:** Recent studies have implicated the role of the immune system in epilepsy, but understanding the diversity of human microglial cells (the main resident immune cell in the brain) and their association to clinical/phenotypical variables has been limited. In particular, transcriptional changes in microglia and peripheral monocytes and their association with different clinical variables such as sex, age, GTC and seizure frequency in epilepsy is not yet well understood. Here, we compared the gene expression profiles of region-specific microglia and monocytes in surgically resected brain tissue from epilepsy patients, and investigated their association with clinical and phenotypical variables.

**Methods/Overview:** We conducted bulk RNA-Seq analysis on microglia isolated from temporal cortex and amygdala/hippocampus regions of the brain and monocytes from blood to determine expression pattern alterations in 12 epilepsy patients. We fit a linear mixed model on gene expression to investigate the association of genes attributed to multiple clinical and phenotypical factors like sex, source, age of onset, GTC history and frequency etc. We included all the categorical variables as random effects in the model, whereas continuous variables like age were modeled as a fixed effect. For each gene, we were able to scrutinize the fraction of variation attributable to each factor included in the model. Using Canonical Correlation Analysis between all pairs of clinical variables, we also studied each variable's association with respect to each other.

**Results:** Our findings on 12 epilepsy patients provide insight into disease phenotypes that may be attributable to differential expression of specific genes. The largest proportion of variance was accounted for by the cell type (microglia versus monocytes) and the second strongest contributor to variance in gene expression was with age of onset of the disease. We also found a subset of genes such as SLC2A1 and BRD2 associated specifically with GTC seizure frequency and age of onset respectively. Some of these genes have previously been associated with epilepsy, but we also identified a set of less well-characterized genes that may be involved in immune-specific contributions to epilepsy subtypes.

**Conclusions:** Even with a relatively small sample size (n=12), our study recapitulates genes with known associations with epilepsy, and also identifies new candidates worth pursuing further.

## PgmNr 1461: Systematic approach for drug repositioning of anti-epileptic drugs.

### Authors:

Y... Ko <sup>1</sup>; C. Lee <sup>2,3</sup>; Y. Lee <sup>3</sup>; J. Lee <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Biomedical Engineering, Hankuk University of Foreign Studies, Gyeonggi-do, Seoul, Korea; 2) Department of Clinical Genetics, Department of Pediatrics, Yonsei University, College of Medicine, Seoul, Korea; 3) Department of Pediatrics, Yonsei University, College of Medicine, Seoul, Korea

Epilepsy is the most common neurological disorder with heterogeneous causes, affecting 1-2% of the population. The majority of patients with Epilepsy can be effectively treated, given an accurate diagnosis and appropriate treatment. Thus, it is very important to determine the accurate type of epilepsy in order to select appropriate drug therapy. Most of existing therapies for epilepsy patients have focused on the symptomatic treatment with such drugs and there are general sorted antiepileptic drug lists which have been known to use for a specific type of epilepsies. In spite of improved efficacy of new AED drugs and novel therapy, there still approximately 20~ 30% of patients, who have either intractable or uncontrolled seizures.

Recent improvement in cost and accuracy of whole-exome sequencing (WES) has enabled the accurate diagnosis of genetic diseases and it allows for identification of the causative variants of diseases. Especially, for patients with refractory epilepsy, WES is popularly used to find accurate causative genes through the trio analysis, which would be very helpful to provide the accurate diagnosis for individual epilepsy patient with unknown origin. The understanding of biological mechanisms of existing AEDs and pathology of epilepsies leads to the dramatic advance of AED development. Thus, it is feasible strategy to predict the appropriate drug based on the causative genes or perturbed biological pathways of refractory epilepsy patients, which leads to the real potential of precise medicine.

In this study, we constructed the epilepsy drug-target network (EDT) and successfully demonstrated the characteristics and effectiveness of popularly used AEDs and the pathological mechanisms of existing AEDs. Especially, we discovered that the causative genes of most of intractable patients were not the targets of existing AEDs as well as they are very far from the etiological mechanisms of existing AEDs in the functional networks. Finally, we showed the existence of new drugs which is targeting the causative genes of intractable epilepsy patients, which will be a new candidate for refractory epilepsy patients. Our systematic approach demonstrated the new possibility for the drug repositioning through combination of the drug-target network and functional network.

## PgmNr 1462: Global alteration of circRNA landscape in Alzheimer's disease.

### Authors:

F. Wang <sup>1</sup>; L. Chen <sup>2</sup>; J.Y. Chen <sup>1</sup>; E.C. Bruggeman <sup>1</sup>; J.K. Kuehner <sup>1</sup>; B. Yao <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 2) Department of Health Outcomes Research and Policy, Auburn University Harrison School of Pharmacy, Auburn, AL, USA

---

Circular RNAs (circRNAs) are a class of endogenous RNAs with a unique circularized structure derived from back-splicing of exons in precursor mRNAs, and have recently emerged as one of the pivotal regulatory RNAs in higher eukaryotes. Many circRNAs are enriched in mammalian brains and dynamically regulated during neurodevelopment. CircRNAs are relatively stable and resistant to exonuclease due to their covalently circularized structure, thus they could accumulate in the brain during aging and have been linked to age-related neurodegenerative disorders, such as Alzheimer's disease (AD). However, the molecular functions of circRNAs in AD pathogenesis remain elusive. While numerous computational tools have been developed to predict circRNA production using junction reads from RNA-seq data, algorithms to identify critical genomic features regulating circRNA biogenesis, such as repetitive elements in the introns flanking circularized exons, are still lacking. Furthermore, rigorous statistical methods to perform differential expression analysis of circRNAs remain under-developed. In this work, we first developed circMeta, a unified computational framework for genomic feature annotation and differential expression analysis of circRNAs. Second, we utilize circMeta to identify genome-wide alteration of circRNA landscape in prefrontal cortex of control and AD human postmortem brains, as well as in the well-established 5XFAD mouse model. As a result, we discovered 2,484 and 1,948 differentially expressed circRNAs associated with AD in human and mouse brains, respectively, with 516 circRNAs conserved between species. Interestingly, many circRNAs were produced from host genes that have been well-established in AD pathology, such as circAPP, circRIMS2 and circNRXN1. Mechanistically, a significant portion of differentially expressed circRNAs were produced from loci displaying aberrant alternative splicing associated with neuropathology and clinical manifestation, implying the potential contribution of circRNA biogenesis in AD-related aberrant alternative splicing. In addition, many of these circRNAs could serve as AD-related miRNA sponges to affect key miRNA-mRNA networks. One circular RNA, circKTN1, is predicted to sequester miR-516b, a key miRNA that regulates neuron differentiation and aging. Taking together, our study reveals global alteration of the circRNA landscape associated with AD and offers potential mechanistic insights of circRNA-related AD pathogenesis.

## PgmNr 1463: NeuroSeq: A platform for investigating the genetics behind neurological conditions.

**Authors:**

D.M. Evans; O. Zabaneh; M.J. Farrer

[View Session](#) [Add to Schedule](#)

**Affiliation:** Centre for Applied Neurogenetics, Univ British Columbia, Vancouver, British Columbia, Canada

---

The development of personalized medicine has far-reaching implications for clinical care, with the opportunity to assist with diagnosis, inform treatment, and give insight into clinical course. As genetic sequencing becomes more prevalent, medical associations are recommending sequencing as front-line care for an increasing number of diseases and disorders.

We are developing a comprehensive, searchable repository of genotype/phenotype correlations in neurological conditions for use by clinicians, genetic counselors and laboratory directors in their mission to provide the best available care. The repository securely holds all variant data added to it so as to protect patient privacy, while also acting as a public resource for health care, research, and education. Aggregate data is available to any user, but only appropriately educated/trained users qualify for accounts that grant additional rights: adding new genotype/phenotype data, accessing subject symptom/diagnosis details, and using analytical tools for both individuals and pedigrees.

The NeuroSeq database (<http://www.neuroseq.ca>) is available for ethically approved users: they may upload data, search for correlations, and communicate with other users about results.

## PgmNr 1464: Integrated SNV/structural variation analysis on the genome research platform GENESIS.

**Authors:**

M.C. Danzi; E. Powell; L. Abreu; D. Bis-Brewer; A. Rebelo; S. Zuchner

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dr. John T. MacDonald Foundation Department of Human Genetics, John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL

---

The diagnostic gap for many rare diseases, especially in the neurodegenerative field, still exceeds 30-50%. These patients do not receive a diagnosis and will not be able to benefit from the upcoming genetic therapy developments. We have created the GEM.app/GENESIS genomic analysis platform to facilitate research diagnosis, gene identification, and data sharing (Gonzalez, 2013, 2015). To date, more than 9,000 exomes, genomes and panels from rare disease patients have been analyzed in GENESIS and are available for gene matching. This has contributed to the identification of 65 Mendelian genes in the past 8 years. Recently, structural variant (SV) analysis of whole genomes has been added for easy querying through a user-friendly web interface. This allows for a fully integrated analysis of SNV, indels, and SV without prior bioinformatics expertise required. SVs have historically been under-studied due to the difficulty of their detection in short-read whole genome sequencing (WGS) data. We are applying the Parliament2 software (Zarate et al., 2018) annotated with population frequency and predicted loss of function data on over 400,000 SVs from gnomAD-SV along with labels of neighboring and overlapping genes and gene regulatory elements as identified by ENCODE. As we reanalyze several hundred WGS samples with this pipeline, we fully expect to diagnose unresolved patients and identify novel genes to be reported in this study in the next 6 months.

## PgmNr 1465: *In silico* characterization of repeat expansion variation in 1,116 genomes.

**Authors:**

S. Fazal; M.C. Danzi; S. Zuchner

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Human Genetics, University of Miami, Miami, Florida.

---

Rare monogenic disorders are largely caused by pathogenic variation in protein coding regions of the genome. Yet, the phenotypes of 30-50% of patients remain unexplained. Thus, we and others have speculated about non-coding sequence contribution to the pathogenic variant spectrum. One such unorthodox mechanism is the expansion of short tandem repeats (STRs) into long repetitive stretches of the wild-type or mutated motif. Only 30 such loci have been linked to diseases thus far. In the past, these repeat expansions have been found primarily through linkage analysis of families. Direct identification from short-read whole genome sequencing data has proven difficult due to the inefficiency of alignment algorithms to correctly match reads to the reference genome. However, with the advance of new tools such as ExpansionHunter, many of these challenges can be overcome. Another challenge has been the lack of an extensive structural variation database focused on repeat sequences in the general population. Here we report a thorough analysis of STR expansions in 1,116 genomes. We find that on average, each genome contains 260 large, repetitive loci; with each new genome containing 3-10 new expansions not seen in our previous analysis. Only 2% of all loci identified are shared amongst all the samples, while more than half are observed only in 1-2 samples. Certain regions in the genome show evidence of expansions in a large number of samples, which could indicate that the reference genome underestimates the number of repeats in these areas. The majority of expansions we identified are intergenic, about 28% are intronic, and less than 1% are exonic. Large expansions were overrepresented near Alu elements, appearing proximal to AluS or AluJ sites over 6 times more often than expected by random chance. We will use this data to identify pathogenic repeat expansions in rare neurological disorders such as axonal neuropathies, motor neuron diseases, and undiagnosed ataxia cases. Because of the considerable diagnostic gap in rare diseases it is likely that repeat expansions as a mutational mechanism play a larger role than currently estimated.

## PgmNr 1466: Leveraging machine learning to accelerate gene discovery in developmental brain disorders.

### Authors:

J. Scheck<sup>1</sup>; A. Timms<sup>1</sup>; G. Mirzaa<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Integrative Brain Research, Seattle Children's Research Institute, Seattle, WA.; 2) Genetic Medicine, University of Washington, Seattle, WA

---

Developmental brain disorders (DBDs) including microcephaly (MIC), megalencephaly (MEG) and cortical malformations are complex neurodevelopmental phenotypes with multiple co-morbidities. A majority of these disorders are genetic in origin caused by ultra-rare, largely *de novo*, variants. We performed exome sequencing on an ethnically diverse cohort of 154 individuals diagnosed with various types of DBDs. Using a comprehensive approach including expert review of clinical and neuroimaging features, as well as an in-house bioinformatics pipeline, we identified pathogenic variants in (44/154) 28.6% and candidate variants in (40/154) 26% of individuals. This approach, however, relies on manual curation of candidate variants and/or genes and knowledge regarding genes-variants associated with DBDs. We sought to develop a more systematic computational approach for DBDs to increase the gene discovery rate for these disorders. To address this, we developed a machine learning approach that identifies pathogenic variants in DBDs. We trained this model with clearly pathogenic variants from our cohort and annotated pathogenic variants from the Human Gene Mutation Database (HGMD), the Leiden Open Variation Database (LOVD), and ClinVar, expression profiles, supplemented by pathway analysis including KEGG, Ingenuity Pathway Analysis. This new systematic approach improved the diagnostic yield in our cohort. We believe this approach offers a valuable tool for the analysis of exome data for families with DBDs, enabling faster diagnoses and ultimately more targeted medical management.



## PgmNr 1467: Non-synonymous co-variation is crucial for SUDEP in *SCN1A* driven epilepsy.

### Authors:

P. Vandeventer; A. Ghazi; E. Chen; A. Goldman; C. Shaw

[View Session](#) [Add to Schedule](#)

**Affiliation:** Baylor College of Medicine, Houston, Texas

---

Sudden Unexpected Death in Epilepsy Patients (SUDEP) is a rare disorder that affects some individuals with *SCN1A* driven epilepsy as well as epilepsy patients with different genetic etiology. The prevailing hypothesis for *SCN1A* SUDEP risk has been that co-variation on the other homologue of *SCN1A* from the haplotype bearing the disease variant is a necessary and sufficient condition for SUDEP outcomes. We have developed an alternative genetic modifier model of *SCN1A* SUDEP. Under this model, concomitant variation – or genetic load – outside of *SCN1A* but in genes related to the SUDEP pathophysiology drive SUDEP outcome risk. Individuals with increased loads or specific patterns of co-variation in these SUDEP risk genes have higher risk, while individuals that lack such variants are at lower risk.

To pursue this hypothesis we created a data set and computational approach to evaluate this model. We identified individuals who underwent exome sequencing that have pathologic or potentially pathologic *SCN1A* variation from the Baylor Genetics laboratory who were also patients at Texas Childrens hospital. We also utilize exome data from SUDEP individuals with *SCN1A* driven epilepsy. Taken together, these data provide a set of positive control *SCN1A* SUDEP cases as well as a collection of test individuals that we can evaluate genomically and by chart review.

The computational procedure considers the rate of pathogenic or potentially pathogenic variation in genes related to the pathophysiology of SUDEP. The gene set used for load analysis was independently identified by prior expert review of the literature and publications of genetic models of SUDEP. The pathways and processes implicated include cardiac arrhythmia, respiration and spreading depression associated genes. Our preliminary results indicate significant evidence for increased load of Variants of Uncertain Significance (VUS) and potentially pathologic alleles in the candidate gene set among the SUDEP cases.

This suggests we can identify specific patterns of co-variations segregating within the pathophysiological molecular SUDEP pathways that will allow us to discriminate between high risk *SCN1A*+ SUDEP cases vs low risk for SUDEP among *SCN1A*+ live epilepsy patients.

## PgmNr 1468: Rapid genome wide tissue specific filtering in rare variant aggregation based modeling.

**Authors:**

N. Wheeler; P. Benchek; M. Warfe; W. Bush

[View Session](#) [Add to Schedule](#)

**Affiliation:** Case Western Reserve University, Cleveland, Ohio.

---

Rare variant analyses often group variants together into units that reflect a biological function, such as gene regions, to combine their small individual effects and improve statistical power. The standard practice for these aggregation based tests is to include every variant within a gene, often filtering by most damaging variant consequence, with no regard to transcript activity or tissue-specific isoform expression. This creates the possibility for variants with important effects under specific biological conditions (such as within a particular tissue) to be obscured by variants which are irrelevant to the context of interest (i.e. variants whose transcripts are actually silent within a tissue of primary interest). Resources such as the genotype-tissue expression (GTEx) project contain data describing the tissue specific activity levels of transcripts, which in turn describes the tissue specificity of the variants within them. By annotating variants with their tissue specific activity, biologically irrelevant variants can be ignored during rare variant aggregation based analyses, which may reveal biologically relevant rare variants that would otherwise be undetectable at existing sample sizes.

We conducted gene region-based rare variant analyses, using the seqMeta R package implemented on a Hadoop cluster, with tissue specific variant filtering for 53 tissues across 29,000 genes. We developed and applied a generalized framework for aggregation based analyses with variant annotations derived from the GTEx dataset, and evaluated the impact of tissue specific filtering on SKAT tests including all gene-based variants for different strata of significance cut-offs. For the most significant alpha values ( $p < 10e-6$ ), tissue specific filtering removed 1 to 3 genes from significance. At more moderate alpha values ( $10e-4 < p < 10e-6$ ), newly significant genes appear within analyses of some tissues, and in general as the significance cutoff decreases, the set of significant genes becomes more differentiated across tissues. These findings indicate that for SKAT analyses including all genic variants (without functional filtering), accounting for tissue specific expression in the analysis does not cause major shifts in p-values. We anticipate that analyses restricted to putatively functional variants (i.e. missense) may induce greater tissue-specific differences in statistical results.

## PgmNr 1469: Characterization of medically actionable variants in 2,500 publicly-available, high-depth genomes of diverse ancestry.

### Authors:

Chen <sup>1</sup>; E. Dolzhenko <sup>1</sup>; A. Gross <sup>1</sup>; B. Lajoie <sup>1</sup>; D.R. Bentley <sup>2</sup>; M.A. Eberle <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Illumina Inc., San Diego, CA; 2) Illumina Cambridge Ltd, UK

---

The creation of powerful annotation databases such as gnomAD has been enabled by the availability of population-scale whole genome sequencing (WGS). This has greatly improved clinical interpretation of simple SNVs and indels. Yet, many medically important regions and variants such as triplet repeats and homologs require specialized informatics methods and are not included in these current WGS-based databases. To this effect, population-level characterization of known clinical variants is needed to maximize the impact of population sequencing experiments. The recent availability of public sequence data such as the high depth (>30x) WGS data from >2,500 samples from the 1000 Genomes Project (1kGP) will allow scientists to develop new methods and release results for public scrutiny. As a demonstration of this concept, we have developed several tools that address difficult medical variants/regions and annotated these variants in the 1kGP samples.

For this demonstration, we identified three clinical variants that are not reported in standard secondary analysis pipelines: 1) spinal muscular atrophy (SMA) detection and carrier screening, 2) *CYP2D6* genotyping for pharmacogenomics applications and 3) detection of triplet repeat expansions. The SMA and *CYP2D6* loci pose difficulties because the regions include segmental duplications, i.e. *SMN1-SMN2* and *CYP2D6-CYP2D7* paralogs, resulting in ambiguous read alignments and unreliable variant calls. We have developed a general method that overcomes the challenges with segmental duplications. Repeat loci are challenging because they occur in low complexity regions prone to elevated error rates and are detected using ExpansionHunter. Because these are complicated regions, we confirmed our locus-specific methods by extensively validating the calls against those made by orthogonal technologies. We then applied our methods to call the *SMN1* and *SMN2* copy number, *CYP2D6* star alleles and repeat expansions in the 1kGP population and quantified differences between subpopulations. The software tools used to make these variant calls are freely available to the community under an open-source license, and the per-sample variant calls on the 1kGP samples will be made public and linked to the WGS sequence data for use as a community resource. We will present the allele frequency distributions by sub-population and ongoing perpendicular validation of these methods as validation data are generated from high-quality long reads.

# PgmNr 1470: A new machine-learning based method to accurately assess copy number variants from whole genome sequencing data and its application on the analysis of the Hirschsprung disease genome.

## Authors:

Zhuang <sup>1</sup>; R. Ye <sup>4</sup>; P.C. Sham <sup>2,4</sup>; M.M. Garcia-Barcelo <sup>1,2</sup>; C.S. Tang <sup>1,2,3</sup>; P.K. Tam <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; 2) Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; 3) Dr. Li Dak-Sum Research Centre, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; 4) Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

---

**C**opy **N**umber **V**ariants (CNVs) are defined as DNA segments larger than 50 base-pairs with copy number changes. CNVs discovery is essential for uncovering genes/risk factors for a wide range of diseases, including Hirschsprung disease(HSCR; colon aganglionosis). Incidentally, gross deletions encompassing *RET* or *EDNRB* lead to the discovery of these two main HSCR genes. Importantly, our previous CNV analysis of HSCR patients identified the association of CNVs encompassing *NRG3* with HSCR, therefore vindicating a role for CNVs in HSCR.

As CNV calling is still a challenge, we have devised and developed a new machine-learning based method CNV-JACG to accurately assess CNVs. Initially, in order to get the most comprehensive CNVs, we used four complementary CNV discovery methods namely CNVnator, Delly, Lumpy and Seeksv to call CNVs. We used 9 in-house trios to produce training dataset, 11 pairs of in-house duplicated samples for validation, and 2 trios from 1000 Genomes Project for evaluation. For deletions, after CNV-JACG assessment, the concordance between each pair of 11 duplicated samples increases from 63% to 88%, and the Mendelian inconsistent rate of the 2 trios decreases from 25% to 6%. In the benchmark sample NA12878, 84.7% of CNV-JACG predicted CNVs are consistent with previously published result.

CNV-JACG has been used for the CNV analysis of whole genome sequencing data (pair-end 150bp, ~30X) generated from 443 HSCR patients and 493 matched controls. Here we will present the performance of our new program together with the genome-wide association results.

## PgmNr 1471: A comparison of machine learning classifiers for dementia with Lewy bodies using miRNA expression data.

### Authors:

D. Shigemizu <sup>1,2,3,4</sup>; S. Akiyama <sup>1</sup>; Y. Asanomi <sup>1</sup>; K.A. Boroevich <sup>3</sup>; A. Sharma <sup>3,4,5</sup>; T. Tsunoda <sup>2,3,4</sup>; T. Sakurai <sup>1,6</sup>; K. Ozaki <sup>1,3</sup>; T. Ochiya <sup>7,8</sup>; S. Niida <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Center for Geriatrics and Gerontology, Obu, Japan; 2) 2.Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo, Japan; 3) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) CREST, JST, Tokyo, Japan; 5) School of Engineering & Physics, University of the South Pacific, Suva, Fiji; 6) 6.Department of Cognitive and Behavioral Science, Nagoya University Graduate School of Medicine, Nagoya, Japan; 7) 7.Division of Molecular and Cellular Medicine, Fundamental Innovative Oncology Core Center, National Cancer Center Research Institute, Tokyo, Japan; 8) Institute of Medical Science, Tokyo Medical University, Tokyo, Japan

---

Dementia with Lewy bodies (DLB) is the second most common subtype of neurodegenerative dementia in humans following Alzheimer's disease (AD). Present clinical diagnosis of DLB has high specificity and low sensitivity, and finding potential biomarkers of prodromal DLB is still challenging. MicroRNAs (miRNAs) have recently received a lot of attention as a source of novel biomarkers. Here, using serum miRNA expression of 478 Japanese individuals, we investigated potential miRNA biomarkers and constructed an optimal risk prediction model based on several machine learning methods: penalized regression, random forest, support vector machine, and gradient boosting decision tree. The final risk prediction model, constructed via a gradient boosting decision tree using 180 miRNAs and two clinical features, achieved an accuracy of 0.829 on an independent test set. We further predicted candidate target genes from the miRNAs. Gene set enrichment analysis of the miRNA target genes revealed 6 functional genes, including the DHA signaling pathway associated with DLB pathology. Two of them were further supported by gene-based association studies using large numbers of single nucleotide polymorphism markers (*BCL2L1*:  $P=0.012$ , *PIK3R2*:  $P=0.021$ ). Our study provides an effective tool for DLB classification, and with further improvement, such as integrative analyses of genomic and/or transcriptomic data, has the potential to contribute to practical clinical application in DLB.

## PgmNr 1472: Comprehensive identification of somatic single nucleotide variation using a multi-institute analysis of a common brain tissue sample.

### Authors:

Y. Wang; on behalf of the Brain Somatic Mosaicism Network

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan.

---

Somatic mutations are genetic differences among cells within an individual that arise from unrepaired DNA damage that occurs during DNA replication and other cellular processes. Somatic variation has been established as a causal feature in various cancers. However, while tumor cells undergo clonal expansions that often increase the frequency and subsequent detection rate of somatic mutations, it is difficult to discover somatic single nucleotide variants (SNVs) from non-tumor tissues because they are present in a low number of cells and can occur early in development, leading to their presence in both sample and control tissues. Here, we sequenced a single human brain to develop, optimize, and evaluate the performance of methods for somatic variant discovery as part of the Brain Somatic Mosaicism Network (BSMN), a multi-institutional initiative that seeks to gain a comprehensive understanding of somatic mosaicism in the human brain. We first simulated mosaic variants at various allele frequencies by mixing and sequencing DNA from four unrelated individuals. Using these data, we evaluated the performance of commonly used tools for somatic variant discovery. This experiment revealed that the naive application of common tools used for germline variant discovery from a single sample or somatic variant calling from paired samples (as typically done in cancer studies) is neither precise nor sensitive enough to confidently detect low frequency mosaicism (*i.e.*, variants present in <10% of the cells in an individual). Next, different BSMN labs produced four replicates of whole-genome sequencing and two replicates of whole-exome sequencing from bulk dorsolateral prefrontal cortex tissue and dural fibroblasts from a reference healthy human brain. We developed and applied several analytical approaches for calling and filtering somatic SNVs from these data to arrive at a set of 1298 putative SNV calls. We conducted extensive validation on a subset of 400 variants comprising both high confidence and randomly selected low confidence events, leading to the identification of 44 *bona fide* somatic variants. As part of this process, we derived a series of bioinformatics filters and consensus best practices for somatic variant discovery. This analytical resource and best practices implementation will be made available to the scientific community and provide a guide for the assessment of the role of somatic single nucleotide variation in human neuropsychiatric disease.

## PgmNr 1473: A machine learning approach to identify new epigenome wide associations in Alzheimer's disease.

### Authors:

Y. Huang <sup>1</sup>; X. Sun <sup>1</sup>; H. Jiang <sup>2</sup>; S. Yu <sup>1</sup>; C. Robins <sup>3</sup>; M. Armstrong <sup>4</sup>; Z. Mei <sup>3</sup>; W. Pittard <sup>2</sup>; P. De Jager <sup>5,6</sup>; D. Bennett <sup>7</sup>; A. Wingo <sup>8,9</sup>; T. Wingo <sup>3,4</sup>; P. Jin <sup>4</sup>; Z. Qin <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, Emory University, Atlanta, GA; 2) Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA; 3) Department of Neurology, Emory University School of Medicine, Atlanta, GA, USA; 4) Department of Human Genetics, Emory University, Atlanta, GA; 5) Cell Circuits Program, Broad Institute, Cambridge, MA, USA; 6) Center for Translational and Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY, USA; 7) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA; 8) Division of Mental Health, Atlanta VA Medical Center, Decatur, GA, USA; 9) Department of Psychiatry, Emory University School of Medicine, Atlanta, GA, USA

---

Alzheimer's disease (AD) is an age-dependent illness with increased environmental contributions at older ages of onset. Thus, understanding epigenomic changes in AD are likely to shed new light on AD disease risk. Among the most widely studied epigenomic change is DNA methylation. Multiple Epigenome-Wide Association Studies (EWASs) on AD show 100s of differentially methylated sites in different brain regions. However, current array-based technologies that these studies utilized profile only about 2% of all CpGs which is inadequate. In this study, we developed EWASplus, a novel computational method that is capable of predicting whether a CpG site is likely to be associated with a disease or trait. Build upon the results from EWAS studies conducted using array-based technologies, and using genome-wide profiling (such as ChIP-seq, DNase-seq and ATAC-seq) data as features, EWASplus uses a supervised learning strategy to train a model that can distinguish CpGs associated with the disease or not, and then apply the trained model genome-wide to predict whether an arbitrary CpG site in the genome is associated with a disease or not. Cross validation studies conducted on an existing EWAS study on seven AD-related clinical and neuropathologic outcomes show promising results (AUC of ROC curves range from 0.792 to 0.960; for CERAD and neurofibrillary tangle density, respectively). When applying EWASplus genome-wide, we identify hundreds of CpGs that are not on the array but are highly likely to be associated with AD, some of which were further validated by targeted bisulfite sequencing. The development of EWASplus will enable to identify additional epigenetic loci that could be involved in AD pathogenesis.



## **PgmNr 1474: Scalable ensemble learning frameworks on single cell transcriptome from fresh human brains accurately identify rare cell states contributing to the development of Alzheimer's disease.**

### **Authors:**

Y. Kim; J.F. John; Y. Wang; K. Allette; K. Beaumont; R.P. Sebra; P. Roussos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetics and Genomic Sciences, Icahn School of Medicine at Mt Sinai, New York, NY.

---

Better understanding of regulatory architectures and underlying disease etiology substantially enhance our ability to target effective risk variants or biological processes in complex diseases including Alzheimer's Disease (AD). Recently, one study analyzed 8,000 single cells from matched healthy and AD-transgenic mice, identified a novel microglial cell type, and further proposed generalizable but hypothetical models likely restricting the progression of neurological diseases. Transferring the insightful models into applicable therapeutics in human neurodegeneration is much more challenging, as single-cell experiments with human samples are difficult to be controlled for various covariates and their heterogeneity is presumably more complex. In this study, we used fluorescence activated cell sorting (FACS) to isolate immune cells from both biopsies and postmortem human brain tissues in 30 healthy and neurodegenerative patients including AD and Parkinson Disease and generated scRNA-seq libraries of 120,000 cells via 10x protocols. To accurately identify distinct cell-states of several but unknown immune cell-types especially in disease contexts, we also developed scalable and robust computational frameworks. The incorporation of complementary approaches from parametric generative models and nonparametric statistical analyses into iterative optimization more efficiently leverage the unwanted effects of inevitable stochastic noise, sparsity of reads, and various batch effects. With rigorous analyses of these largescale single cell assays, we observed a comparable number of immune cell populations in human to those found in mice and additionally identified distinct microglial states selectively enriched in aged controls and distinctive classes of neurodegeneration. Systematic investigation of the identified subpopulations with GWAS neuropsychiatric traits revealed that a particular subset of cell-states showed large differential regulations of genes adjacent to trait-relevant loci. Regulatory network and state transition analysis on the cell populations also corroborated the differential regulatory interplay among particular microglial states for AD developments via transcriptional binding of elective transcription factors. We hope that our scalable machine learning approach and novel biological findings assist in the identification of immunological therapies, targeting the regulatory mechanisms observed across multiple neurodegenerative diseases.



## PgmNr 1475: INFERNO2: Scalable Spark-based framework for inferring dysregulated enhancer and noncoding RNAs for WGS and GWAS data.

### Authors:

C.-Y. Lee <sup>1</sup>; P.P. Kuksa <sup>1</sup>; A. Amlie-Wolf <sup>1</sup>; E.E. Mlynarski <sup>1</sup>; Y.-F. Chou <sup>1</sup>; H.-J. Lin <sup>1</sup>; E. Greenfest-Allen <sup>1</sup>; Z. Katanic <sup>1</sup>; O. Valladares <sup>1</sup>; A. Kuzma <sup>1</sup>; A. Naj <sup>2,1</sup>; G.D. Schellenberg <sup>1</sup>; Y.Y. Leung <sup>1</sup>; L.-S. Wang <sup>1</sup>; Alzheimer's Disease Sequencing Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Biostatistics, Epidemiology, and Informatics/Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Noncoding variants and structural variants can substantially alter gene expression and are enriched at regulatory elements, highlighting the importance of noncoding variants in disease pathology. However, interpretation of these variants remains challenging in part because current tools mostly focus on protein coding genes or previously reported variants, and largely ignore potentially causal noncoding variants observed in whole-genome sequencing (WGS) and GWAS. Here we introduce INFERNO2, the first Spark-based annotation and analysis pipeline for large-scale WGS and GWAS studies. INFERNO2 expands upon our INFERNO method by introducing a distributed Spark-based implementation of genomic analyses and WGS-based analyses, coupled with a curated, harmonized functional genomic database (GADB).

INFERNO2 provides a systematic approach to harmonize information from diverse data sources using a cell-type categorization system to prioritize causal noncoding variants and identify the affected regulatory mechanisms, tissues, and target genes. By converting the WGS/GWAS data into Spark Resilient Distributed Datasets (RDDs) and using distributed Spark transformations to efficiently identify loci, users can rapidly query a million of variants approximately a minute in a distributed manner against >60,000 GADB tracks (>20 billion genomic intervals) in parallel. INFERNO2 currently integrates >30 data types from 12 consortia including Roadmap and FANTOM5 enhancers, GTEx eQTLs, DASHR ncRNAs, etc.

We applied INFERNO2 to characterize the landscape of rare noncoding single nucleotide variants using the largest and most diverse late-onset Alzheimer's Disease (LOAD) WGS data from 3,709 case-control subjects across three populations from the Alzheimer's Disease Sequencing Project and Alzheimer's Disease Neuroimaging Initiative. In the LOAD group, we identified approximately 29M, 32M, and 35M variants in Caucasian, Hispanic, and African American populations, respectively, of which 73.4%, 62.6%, and 62.2% were rare variants. Overall, of the 189M analyzed variants across populations, 65% were rare variants, with 1.1% of coding, 31.8% of intronic, 7.7% of UTR, 0.8% of promoter, and 58.6% of intergenic regions, on average. FANTOM5 and Roadmap enhancers overlapped 2.7% and 49.0% of rare variants. These analyses support the utility of INFERNO2 (<https://bitbucket.org/wanglab-upenn/INFERNO2/>) for inferring the molecular mechanisms underlying

noncoding WGS signals in a common complex disease.

## PgmNr 1476: Population-scale identification of variation in tandem repeat regions.

### Authors:

E. Dolzhenko <sup>1</sup>; V. Deshpande <sup>1</sup>; S. Chen <sup>1</sup>; J.J.F.A van Vugt <sup>2</sup>; G. Narzisi <sup>3</sup>; A. Gross <sup>1</sup>; J.H. Veldink <sup>2</sup>; D.R. Bentley <sup>4</sup>; M.A. Eberle <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Genomics Research, Illumina Inc., San Diego, CA, USA; 2) UMC Utrecht Brain Center, Utrecht University, Utrecht, The Netherlands; 3) New York Genome Center, New York, NY, USA; 4) Clinical Genomics Research, Illumina Cambridge Ltd., Cambridge, UK

---

Tandem repeats (TRs) are a major contributor to human genetic variation. Estimates based on the latest high-quality structural variant (SV) truth sets indicate that over half of human SVs correspond to expansions and contractions of TRs. TRs are linked to a variety of Mendelian diseases (e.g. amyotrophic lateral sclerosis, myotonic dystrophy, Friedreich ataxia) and cancer (e.g. microsatellite instability and Lynch syndrome). Subtle changes in the repeat motif and flanking regions have also been shown to have an impact on disease onset and progression. Despite the abundance and clinical importance of TRs, they remain a challenge for standard short read-based computational pipelines. Although specialized methods for TR analysis have been recently developed, the application of these tools genome-wide requires accurate and comprehensive TR catalogs. Existing TR catalogs do not resolve repeat locus structure sufficiently well to be used with this new class of variant callers and, because they are biased to the reference, do not include many TRs that are polymorphic in the population.

To address this problem, we have created a novel method to annotate TRs. This method automatically determines repeat composition and identifies common variants in and around the repeat sequence. The method works by (a) locating approximate positions of TRs by performing a simultaneous scan of read alignments across many whole-genome sequencing (WGS) samples, (b) creating assembly scaffolds for these loci using PacBio CCS long-read samples, and (c) realigning WGS data to the scaffolds to annotate precise boundaries of polymorphic regions and their sequence composition. We have applied this method to high-depth sequencing data for 2,504 WGS samples from the 1000 Genomes Project and four PacBio CCS samples. We are releasing a pilot version of the resulting catalog that includes over 200,000 highly polymorphic TR loci together with their genotypes across all 2,504 samples, including about 5,000 loci containing two or more adjacent repeats. The list of polymorphic repeats identified by our method includes 22 out of 24 known pathogenic short tandem repeats suggesting that polymorphic TRs are not only useful for resolving common structural variation but also an appropriate target for studies aimed at the discovery of novel pathogenic TRs.

# PgmNr 1477: A deep learning framework for interpreting repetitive DNA sequence in heritable disease.

**Authors:**

J.L. Larson; L. Ringel; E. Hallmark; J. Thibert; B.C. Haynes; S. Statt

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Research and Development, Asuragen, Inc., Austin, Texas.

---

**Introduction:**

Pathogenic short tandem repeats (STRs), such as those associated with amyotrophic lateral sclerosis (ALS), fragile X syndrome, and Huntington's disease, elude interpretation by NGS due to their length (often hundreds of nucleotides) and low sequence complexity. Advances in PCR/capillary electrophoresis (CE) chemistry and instrumentation have enabled the analysis of STR DNA fragment size through electropherograms with visually identifiable peaks that are translatable into corresponding genotypes. Existing approaches for PCR/CE peak discernment either require manual inspection or heuristic algorithms tailored to the signal idiosyncrasies of a specific assay or instrument. To circumvent these limitations, we reframed the task of PCR/CE analysis as a computer-vision image classification problem and developed a convolutional neural network (CNN) to robustly detect and interpret STR alleles.

**Methods:**

Biological specimens from the Coriell catalog with known  $G_4C_2$  repeat lengths in *C9orf72*, an ALS-associated gene, were collected and analyzed with the AmpliX<sup>®</sup> PCR/CE *C9orf72* Kit (Asuragen). The resulting PCR/CE traces were portioned into training and testing cohorts for model development and evaluation. Potential genotypic peaks were automatically selected and classified as predicted peaks or background noise with multiple convolutional layers and pooling operations using Keras and TensorFlow. The trained CNN was evaluated and compared to a rule-based algorithm on the Coriell test set and an independent clinical cohort.

**Results:**

The CNN was trained on fewer than 600 PCR/CE traces and evaluated on  $\geq 150$  independent validation samples. The model achieved sensitivity and PPV of  $\geq 99\%$  at the genotype level and  $\geq 99\%$  accuracy at the sample category level (i.e., normal, intermediate, and expanded) in both independent datasets, outperforming a heuristic method. The limit of detection for expanded mosaic peaks in the background of normal *C9orf72* alleles was 20%. The final pipeline was deployed as an analysis plugin with swift ( $< 3$  min per 100 samples) push-button reporting, automated QC checks, and a user-friendly interface for external assessment.

**Conclusion:**

CNNs can greatly improve the accuracy and expedience of PCR/CE-based genotyping for *C9orf72* STR mutations. This is the first deep-learning platform for the automated analysis of PCR/CE data and can be leveraged to accelerate research and diagnostic applications in STR disorders and other PCR/CE applications.

## PgmNr 1478: A new pipeline designed for reliable and accurate short tandem repeat realignment.

### Authors:

J. Lu <sup>1</sup>; C. De Araujo Martins Moreno <sup>2</sup>; M.B. Harms <sup>2</sup>; E.L. Heinzen Cox <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Genomic Medicine, Columbia University; 2) Department of Neurology, Columbia University

---

Short tandem repeats (STRs) are DNA sequences composed of tandemly repeated 1-6 base pair units. STRs comprise ~3% and are widely distributed throughout the human genome. These repetitive sequences have very high mutation rates that lead to both sequence and length variants. STR variants have been associated with a number of severe neurological diseases including Fragile X syndrome, amyotrophic lateral sclerosis, and Huntington's disease. Despite the likely important role of these genetic variants in human disease, accurate calling of STR variants from short-read next-generation sequence data has been extremely challenging due to the high complexity of the DNA sequence that leads to higher rates of sequencing errors and lower mapping efficiency of sequence reads that include repeated sequence. Here, we developed a new pipeline focusing on stable and accurate realignment of sequencing reads containing tandemly repeated sequences. Our pipeline highlights the following three features that distinguishes it from existing STR realignment softwares: (1) simplifying the information required for local realignment of the target STRs provided by user to diminish the impact of any imperfect input and allow easier user customized target lists, including unbiased compilations of STRs in any user interested reference sequence; (2) providing a convenient tool to locally realign the short next generation sequencing reads to the target STR and flanking regions; (3) combining both mapping based flank-guiding method and k-mer based repeat-guiding method to ensure both the accuracy and the sensitivity of STR sequence realignments. Our pipeline successfully realigned >99.9% of simulated STR sequencing reads incorporating a sequencing error rate of 0.5% from individuals with extension or truncation variants of varying sizes at the amyotrophic lateral sclerosis associated *C9ORF72* STR locus. Our pipeline also succeeded in identifying a pathogenic STR expansion variant from an amyotrophic lateral sclerosis patient with a clinically-confirmed and quantitatively assessed pathogenic variant by applying a subsequent calling analysis to the realignment result. In summary, our new pipeline provides a user-friendly choice for genomewide target STR realignments, with reliable and accurate realignment results highly adaptable for subsequent variation calling and analysis.

## PgmNr 1479: Artificial intelligence approach to genomics-image analysis of Alzheimer's disease.

### Authors:

M. Xiong <sup>1</sup>; Y. Liu <sup>1</sup>; Z. Li <sup>1</sup>; Q. Ge <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Biostatistics, Univ Texas Hlth Sci, Houston, Texas.; 2) Department of Mathematics, Fudan University

---

Deep learning demonstrates great potential for diagnosis and prediction power in complex diseases. Despite its great progresses in computer vision, diagnosis and early detection of complex diseases, it is well known that deep learning is 'black boxes' due to their low interpretability to humans. The lack of transparency of deep learning compromises its application to the prediction and mechanism investigation in complex diseases. Overcoming these limitations remains a great challenge. In this report, we develop a novel general framework that integrates deep learning and causal inference for genomic-image analysis. It consists of two stages: (1) develop convolutional neural networks (CNN) to classify AD status and use occlusion map to find image regions that are most distinctive for disease status and (2) the state-of-the-art causal inference tools to determine if the selected image regions are causal for AD and identify genetic variants that cause image variation. The CNN model VGG was chosen for image classification and prediction. Prediction difference analysis for visualizing the response of CNN to a specific input was used to select features for DTI image classification. Wasserstein conditional generative adversarial networks (WCGAN) will be used to discover causal relationships between the brain neuroimaging region and AD, and causal relationships between the brain neuroimaging region and gene as well. A novel statistic for testing causation and its asymptotical distribution are developed. The proposed algorithm is applied to the Alzheimer's Disease Neuroimaging Initiative dataset with diffusion tensor imaging (DTI) in 151 subjects (51 AD and 100 non-AD) that were measured at 4 time points of baseline, 6, 12 and 24 months, and 1,589,061 SNPs. The algorithm achieved prediction accuracies of more than 92% at all four time points. The algorithm also identified brain regions underlying AD consisting of the temporal lobes (including the hippocampus) and the ventricular system. To uncover genetic architecture of brain regions, we conducted genetic imaging causal analysis for these two brain regions. We identified that 43 genes had causal relationships with the frontal, left temporal lobe and 46 genes that showed causation with the right temporal lobe region such as CD33, APP, FGF4, FRMD6, Dock9, H3F3B, SCYL1, PIGC and AKAP5 that were confirmed causations with some diseases in literature.

## PgmNr 1480: Bioinformatics approach to propose autism spectrum disorders priority genes in Mali.

### Authors:

M. Haidara; M. Kouyate; M. Sangare; A. Karembé; F. Doumbia; K. N'Diaye; O. Sidibe; S. Samake; A. Keita

[View Session](#) [Add to Schedule](#)

**Affiliation:** Faculty of Medicine and Odontostomatology, University of Sciences, of Techniques and technologies of Bamako, Bamako, Mali, Mali

---

**Introduction:** Autism Spectrum Disorder (ASD) is a developmental disorder affecting social interaction, communication and repetitive movements of restrictive interests. ASD is familial in 10% of cases.

**Aim:** This work was aimed to propose a panel of ASD priority genes for the genetic study of ASD in Mali.

**Methods:** The literature review and the use of bioinformatics tools have made it possible to identify ASD genes. Priority criteria were used to determine the frequent mutations in the hot regions of the identified genes and to identify regions recognized by restriction enzymes.

**Results:** The priority genes *Tcf4*, *Tsc2*, *Chd8*, *Nlgn3*, *Nme1*, *Mef2c*, *Shank3*, *Dhcr7*, *Tsc1*, *Comt*, *Pten*, *Snta1*, *Akt1*, and *Prodh* were obtained with high priority scores. In the following five genes *Mef2c*, *Pten*, *Prodh*, *Nme1*, and *Dhcr7*, mutations were found at endonuclease restriction sites *BseRI*, *Nspl*, *PfrJS2IV*, *BspGI*, *BsaBI*, and *SpoDI*.

**Discussion:** The list of these priority genes could be used as a starting point for the determination of a panel of ASD genes in Mali. *Chd8*, *Pten*, and *Shank3*, *Mecp2* and *NLGN3* are listed as priority genes by the Simmons Foundation for Autism Research Initiative, too. Restriction by endonuclease may reduce the cost of sequencing eligible genes.

**Conclusion:** Determining the frequency of mutations of these genes in subjects living with ASD and in the general population in Mali will further clarify their relevance in Mali.

## PgmNr 1481: A maximum flow formulation for dimensionality reduction and identification of high-confidence noncoding variants associated with autism spectrum disorder.

### Authors:

M. Varma<sup>1</sup>; K. Paskov<sup>2</sup>; N. Stockham<sup>3</sup>; J.Y. Jung<sup>2,5</sup>; B. Chrisman<sup>4</sup>; P. Washington<sup>4</sup>; M.W. Sun<sup>2,5</sup>; D.P. Wall<sup>2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, Stanford University, Stanford, CA; 2) Department of Biomedical Data Science, Stanford University, Stanford, CA; 3) Department of Neuroscience, Stanford University, Stanford, CA; 4) Department of Bioengineering, Stanford University, Stanford, CA; 5) Department of Pediatrics, Stanford University, Stanford, CA

---

The effect of noncoding variants on the phenotypes of complex genetic disorders remains unclear. Machine learning approaches for predicting putative genetic variants from whole genome sequence (WGS) data have been proposed to elucidate the relationship between genotype and phenotype; however, the predictive power of these methods is severely limited by the presence of high-dimensional variant feature spaces, which lead to model instability and lack of generalizability. Here, we design a novel maximum flow formulation based on linkage disequilibrium (LD) to address this issue and extract a set of stable, high-confidence noncoding variants that are likely to be associated with Autism Spectrum Disorder (ASD).

We analyze 232,193 variants in simple repeat sequences (SRS), collected from 30x coverage WGS of 2182 children with ASD and 379 unrelated control patients with a distinct late-onset disease. We perform 5-fold cross-validation with an  $L_1$ -regularized logistic regression classifier and extract the list of variants assigned non-zero coefficient scores from each validation fold. A model with stable features will be resistant to slight perturbations in the underlying data; to obtain a list of stable variants, we use maximum flow to identify a maximum variant matching across folds based on the presence of LD. The five sets of variants are assembled in a flow network, in which each node  $n$  is defined by a variant and the fold in which it occurs, represented as  $n=(v,f)$ . A directed edge is drawn between a pair of nodes  $n_1$  and  $n_2$  if the following criteria are satisfied: (1)  $n_1$  and  $n_2$  are present in neighboring folds such that  $f_2=f_1+1$  and (2)  $v_1$  and  $v_2$  are in LD as indicated by the  $R^2$  value exceeding 0.8. This maximum flow formulation identified 50 regions (representing 55 SRS variants) that remain stable across all five folds.

To determine if this set of 55 SNPs can serve as a viable biomarker for ASD, we train a logistic regression classifier on this reduced feature set, which we found to perform well on the test set (AUC-ROC=0.812) as well as on an independent dataset consisting of unique ASD samples and non-ASD controls (AUC-ROC=0.922). The classifier also significantly outperforms a bootstrapped analysis with randomly-selected variants ( $p<0.05$ ). Thus, we determine a set of stable SRS variants that are likely to be associated with the ASD phenotype and show that this method can effectively identify



predictive variants from a high-dimensional feature space.

## PgmNr 1482: Shared genetic etiology underlying LOAD and MDD.

### Authors:

J. Barrera <sup>1,2</sup>; D. Sprague <sup>1,2</sup>; M. Lutz <sup>1</sup>; O. Chiba-Falek <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurology, Duke University, Durham, NC.; 2) Genomic and Computational Biology, Duke University, Durham, NC

---

Patients with late-onset Alzheimer's Disease (LOAD) frequently manifest comorbid neuropsychiatric symptoms (NPS), with depression and anxiety being most frequent. Furthermore, depression has been found to be associated with increased risk to develop LOAD. Major depressive disorder (MDD) is a neuropsychiatric condition that patients have an increased prevalence of LOAD. These evidences suggest the possibility of shared etiologies and intersecting pathways between LOAD and MDD. As the first logical step towards understanding their overlapping etiologies, we focused on genetic risk and investigated genetic pleiotropy between LOAD and MDD. We performed analyses for the identification of shared genetic signatures based on comparisons between associated-single nucleotide polymorphisms (SNPs) of the two diseases. Specifically, we assessed genetic pleiotropy by conditional false discovery rates (FDR), fold-enrichment plots and conditional quantile-quantile (q-q) plots using LOAD and MDD genome-wide-association studies (GWAS). GWAS datasets for LOAD were obtained from the International Genomics of Alzheimer's Project and for MDD from the Psychiatric Genomics Consortium (PGC). We found an enrichment of 4-8-fold for LOAD across increasingly stringent levels of significance with the MDD GWAS association (LOAD|MDD), including and excluding the *APOE* region, respectively. Conditional FDR analysis supported known LOAD-risk loci including *SPI1*, *MS4A2/MS4A6A*, *SORL1*, *CR1* and several novel risk-loci including *MADD* and *MFSD2A* as pleiotropic for LOAD and MDD. *MFSD2A*, is particularly interesting as it's required for blood-brain-barrier formation and function, probably by mediating lipid transport. Noteworthy, SNPs in *SPI1*, previously identified as eQTLs, showed FDR significant association with LOAD and MDD. The reverse conditional association (MDD|LOAD) showed moderate enrichment of approximately 7 fold. Overall, our results showed a moderate level of polygenic overlap between LOAD and MDD. Identifying the common genetic signatures for LOAD and MDD will advance our understanding of the shared genetic etiologies and impaired pathways between these two conditions. Our data suggest that *SPI1* (PU.1) may play a role in both MDD and LOAD, implicating common pathways involved in myeloid cell function. This knowledge will provide insights regarding actionable targets for novel therapies to treat depression preceding dementia in an effort to delay or ultimately prevent the onset of LOAD.

## PgmNr 1483: ParseCNV2: Optimized for fast easy curation and CNV disease association.

### Authors:

J.T. Glessner<sup>1,2</sup>; Y. Liu<sup>1,2</sup>; J. Li<sup>1,2</sup>; M. Khan<sup>1,2</sup>; X. Chang<sup>1,2</sup>; P.M.A. Slieman<sup>1,2</sup>; H. Hakonarson<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia; 2) Perelman School of Medicine, University of Pennsylvania

---

Copy number variation (CNV) detection has remained an area of heavy emphasis for algorithm development; however, both CNV curation and disease association remains in its infancy. The current candidate CNV era, where researchers study specific CNVs they believe to be pathological while discarding others, misses the full spectrum of CNVs in a hypothesis free GWAS. We present a next-generation approach to CNV association by natively supporting the popular VCF specification for sequencing derived variants as well as SNP array PennCNV format. While many VCF parsers exist, few if any support the variety of VCF presentations and interpretations of CNV genotypes. Therefore, we implemented a flexible VCF to rawcnv/bed format conversion tool. Key challenges include the variety of ways to represent the end genomic position and the alternative alleles. The end position may have any combination of "LENGTH=" or "END=" in the info field or may need to be inferred by the string length difference between ref and alt alleles. The allele coding for the genotypes can use any combination and order of alternative allele copy number states. ParseCNV2 is a combination of Perl, R, Bash, and C programming. The required inputs are a CNV call file (VCF or rawcnv format) and the case sample definition. ParseCNV2 starts with a comprehensive quality control (QC) pipeline of both samples and individual CNV calls. The pre and post association CNV curation is rigorously supported and emphasized to yield reliable results of highest quality. Chr, start (base pair position based on the genome build), end, p-value, odds ratio (OR), cases (count), controls (count), and filter are statistical association fields in the output. Direction, type, cases, controls, caseIDs, controlIDs are annotated to track the kind of association signals produced. Segmental duplications (SegDups), Database of Genomic Variants (DGV), Guanine/Cytosine base content (GC), cytoband, recurrent events, exon impact, gene(s) impacted and telomere/centromere involvement are provided as genomic feature annotations. Collectively, this provides high confidence and high quality association results. Significant CNVRs are then QC reviewed for further curation and bias screening and either kept or dropped based on predetermined significance criteria. ParseCNV2 allows fast and easy curation and association of CNVs in both population and family-based disease settings. The software is available at <https://github.com/CAG-CNV/ParseCNV2>.

## PgmNr 1484: Introducing privacy preserving machine learning in pharmacogenomics research.

### Authors:

A.C. Hauschild<sup>1,2</sup>; V.S. Marshe<sup>2,3</sup>; M. Maciukiewicz<sup>2</sup>; D.J. Müller<sup>2,3,4,5</sup>; D. Heider<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Marburg, Bioinformatics Department, Medical Informatics Division, Marburg, HE, Germany; 2) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 3) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 4) Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 5) Department of Pharmacology & Toxicology, University of Toronto, Toronto, ON, Canada

---

**Motivation:** In the era of large genomic data sets and machine learning (ML), secure data management is one of the key challenges when developing biomedical software, handling personal, clinical and genomic patient information. Particularly in psychiatric and pharmacogenomic research, where state-of-the-art ML approaches are slowly gaining attention, data confidentiality is a continuous concern. Therefore, privacy-preserving technologies will play a crucial role to pave the way for the application of modern ML in medical diagnostics and treatment optimization of psychiatric disorders.

**Methodology:** In this presentation, we will highlight the pitfalls of traditional ML approaches with respect to data privacy and leakage, as well as the availability and distribution of biomedical and genomic data sets. Subsequently, we will introduce possible solutions utilizing the power of distributed systems, such as federated learning, and the integration of different ensemble learning techniques. The aim is to implement a high-quality centralized model while patient data remains at a potentially large number of secure host locations. In summary, it uses the concepts of distributed computing to (1) train local models on the host devices, (2) build a central aggregated model, and (3) re-distribute the updated model to the hosts.

**Application:** Finally, we will demonstrate potential applications of privacy-preserving ML on large cohorts with extensive clinical and genomic data of patients diagnosed with major depressive disorder and schizophrenia. Thereby, we will outline the importance and performance of federated ML within the scope of personalized psychiatric diagnostics and pharmacogenomic research for treatment optimization.

# PgmNr 1485: Prediction of deleterious effect of noncoding variants mediated by RNA-binding proteins in developmental disorders.

## Authors:

A. Kitaygorodsky <sup>1</sup>; E. Jin <sup>2</sup>; Y. Shen <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Biomedical Informatics, Columbia University, New York, New York.; 2) Department of Computer Science, Columbia University, New York, NY, USA; 3) Department of Systems Biology, Columbia University Medical Center, New York, NY, USA

---

Previous studies identified significant contribution of *de novo* coding variants to developmental disorders such as autism spectrum disorder (ASD), but the impact of *de novo* noncoding variants is not well understood. Most research in noncoding variants has focused on transcriptional regulation through transcription factors and chromatin modifications, while post-transcriptional regulation has been relatively under-studied. We aim to address this gap and improve the understanding of the role of noncoding mutations by investigating disruption of RNA binding protein (RBP) binding sites. In this study, we developed a deep convolutional neural network (CNN) model to learn different RBP binding sites from genomic sequence, and to predict when mutation can cause significant disruption of these sites and disturb normal post-transcriptional regulation. We trained the model using eCLIP binding sequence data with a balanced negative set of sequences randomly sampled from the human genome, controlled for GC content. The eCLIP data was obtained from ENCODE and consists of binding sites of 112 unique RBP types in two cell lines (K562, HepG2). Similar to deep learning methods for identification of transcription factor motifs, both canonical and novel RBP binding motifs were captured within the convolution layers and visualized with techniques like Grad-CAM for validation. Strong binding prediction performance was achieved for the 160 RBPs with a median AUC of 0.890 (10-fold CV). HepG2.HNRNPC (AUC=0.97), K562.TARDBP (AUC=0.96), and HepG2.RBFOX2 (AUC=0.94) performed the best. We applied our method to analyze *de novo* noncoding variants from a whole genome sequencing study of ASD (An et al 2018). We predicted variants as highly loss-of-function (LOF) or gain-of-function (GOF) by maximal likelihood score delta derived from the difference in binding prediction between the alternate and reference sequences across all RBPs that are constrained (ExAC pLI $\geq$ 0.5). We observed significantly more noncoding *de novo* mutations predicted LOF in cases than in controls, suggesting a substantial role of post-transcriptional regulatory noncoding *de novo* variants in ASD. In summary, we developed a deep learning method to predict disruption of RBP binding by noncoding mutations using sequence information not captured by other genetic and epigenetic annotations, and utilized it to prioritize candidate pathogenic variants that potentially disrupt binding of key RBPs.

## PgmNr 1486: Bioinformatics approach to support genetic research in Autism in Mali.

### Authors:

M. Kouyate <sup>1</sup>; M. Sangare <sup>2</sup>; S. Samake <sup>1</sup>; S. Keita <sup>1</sup>; H.G. Kim <sup>3</sup>; D.H. Geschwind <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) ACE, USTTB, Bamako, Mali; 2) FMOS, USTTB, Bamako Mali; 3) Neurological Disorder Center, Qatar Biomedical Research Institute (QBRI), Doha, Qatar; 4) Center for Autism Research and Treatment (CART), UCLA, Los Angeles, USA

---

**Introduction:** Autism is a complex condition resulting from the interaction between genetic predisposition and environmental factors. Bioinformatics can shed light on the genetics of autism. In addition, human genetic studies can be expensive even unaffordable in developing countries in which genetics and mental health are not the health priorities.

**Aim:** To pilot the use of bio-informatics tools to guide scientifically valid, locally relevant and economically sound autism genetic research in Mali.

**Methods:** We used databases NCBI, Assembl, HGMD, LSDB, and DBSNP to identify point mutations and associated endonucleases. Criteria (phenotype, transmission pattern, theoretical expression in the brain, impact of the mutation on the 3D structure of the protein) were used to assign priority to the selected autism genes. We used the protein database (PDB), Modeller (protein 3D structure modeling), Web logo (design the motif of functional domain of the protein) and clustal W (alignment of sequences to find similarities).

**Results:** We found hot point mutations in these high priority autism genes *Mef2c* (Gly27Ala/Leu38Gln), *Pten* (Thr131Ile), *Prodh* (Leu289Met), *Nme1* (Ser120Gly), and *Dhcr7* (Pro227Thr/Glu224Lys) (**Tableau 4**) were respectively associated with restriction restriction enzymes *BseRI*, *NspI*, *PfrjS2IV*, *BspGI*, *BsaBI*, and *SpoDI*. Gly27Ala/Leu38Gln mutations impacted on the 3D structure of the Mef2c protein (**Figure 1**). Alignment of Mef2c protein sequences from different organisms showed high percentage of similarity with highly conserved MADS domain (**Figure 2**).

**Discussion:** The determination of the *Mef2c*, *Pten*, *Prodh*, *Nme1*, and *Dhcr 7* gene mutation frequencies in the Malian population will be interesting in many ways. First, the association of restriction enzymes to the hot spot point mutations will allow us to use PCR + restriction enzyme digestion for mutation screening. Sanger sequencing will be used for confirmation only, which represents an importance saving on the *Sanger sequencing cost*. Second, any discrepancies between the restriction enzyme digestion and sequencing results will lead further discussion and investigation. The knowledge of the 3D structure and potential impact of the mutations on Mef2c protein informed on the protein family and altered function (ex. Leu38Gln).

**Conclusion:** Bio-informatics tools allowed to generate new genetic information on African-led autism research in Mali. Our approach can be applied to other neuropsychiatric disorders.

# PgmNr 1487: Predicting psychiatric phenotypes: Developing and evaluating a standardized, machine learning workflow for genome-wide and clinical data.

## Authors:

V. Marshe<sup>1,2</sup>; A. Hauschild<sup>1,3</sup>; M. Maciukiewicz<sup>1</sup>; J. Reimand<sup>4,5</sup>; D. Müller<sup>1,2,6,7</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 3) University of Marburg, Bioinformatics Department, Medical Informatics Division, Marburg, Germany; 4) Computational Biology Program, Ontario Institute for Cancer Research, Toronto, ON, Canada; 5) Department of Medical Biophysics, University of Toronto, Toronto, ON, Canada; 6) Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 7) Department of Pharmacology & Toxicology, University of Toronto, Toronto, ON, Canada

---

**Introduction:** While state-of-the-art machine learning (ML) approaches are only slowly gaining attention in psychiatric research, ML is already widely applied and shows exceptional diagnostic and prognostic performance in oncology and radiology. Here, we present a standardized ML workflow for psychiatric phenotype prediction using large scale genome-wide data, as well as the evaluation of varying parameters on model performance.

**Methods:** Our sample included 748 individuals from the Alzheimer's Disease Neuroimaging Initiative (ADNI) with genetic (Illumina Omni2.5) and clinical data after standard quality control. We investigated various modelling parameters including phenotype definition, genetic pruning threshold, genetic coding (genotypic vs. allelic SNP coding), sample balance (unbalanced vs. upsampled), feature prioritization feature selection and classifier selection within a nested, 5-fold cross-validation framework. After splitting the sample into independent training (70%) and holdout test (30%) sets, we explored Alzheimer's disease status, beta-amyloid levels, and cognitive performance. Features were prioritization using either functional-based (CADD PHRED scores) or univariate association-based filtering. Subsequently, features were selected using either LASSO or random forest with recursive feature elimination. Lastly, we optimized several classifiers including tree-based, boosting, linear, and kernel-based algorithms. Model performance was assessed using AUC, sensitivity, and specificity.

**Results:** In total, we fit 896 unique parameter combinations. While internally there appeared to be no impact of the coding scheme, externally, additive coding showed better performance (AUC =  $0.53 \pm 0.03$ ) than 'boosted' allelic coding (AUC =  $0.51 \pm 0.04$ ,  $p = 0.006$ ). Sample balancing did not have a significant impact on performance in genome-wide genotyping data (FDR > 0.05). Internally there were significant differences in performance among classifiers, however, these were not reflected in external performance ( $p > 0.05$ ). In addition, features from our best-performing model revealed interesting biological insights showing a potential association of angiogenic pathways with beta-amyloid levels.

**Conclusion:** Our prototype workflow and preliminary results represent a proof of concept showing

the feasibility of such an ML workflow for genome-wide, psychiatric studies and potential for the systematic evaluation of methodological decisions on the modelling process.



## PgmNr 1488: Automated processing of phenotypic data submissions.

### Authors:

R. Mayani <sup>1</sup>; K. Vahi <sup>1</sup>; J.L. Ambite <sup>1</sup>; S. Sharma <sup>1</sup>; M. Azaro <sup>2</sup>; S. Wilson <sup>2</sup>; B. Ruocco <sup>2</sup>; G. Davis <sup>2</sup>; M. Romanella <sup>2</sup>; L. Brzustowicz <sup>2</sup>; E. Deelman <sup>1</sup>; Y. Arens <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Southern California, Los Angeles, CA; 2) Rutgers University, NJ

---

The NIMH Center for Collaborative Genomic Studies on Mental Disorders, now known as the NIMH Repository and Genomics Resource (NRGR), maintains biomaterials, demographic, and phenotypic data from over 200,000 well-characterized individuals with a range of psychiatric illnesses, their family members, and unaffected controls. NRGR receives these data from principal investigators of NIMH-funded studies. The center is then responsible for curating the clinical data submitted and creating collections of well-characterized, high-quality patient and control data and biosamples that are widely used for psychiatric research.

Previously, the curation effort was largely manual with ad-hoc harmonization procedures in place that led to inconsistencies and variance across studies and disorders. To streamline and formalize this process we developed a web-based automated quality control system (AutoQC) for phenotypic data submissions (<https://nimhgenetics.org/submit-your-data/overview>). AutoQC curates submissions and ensures stringent data quality requirements. At the heart of the system is a custom, flexible data dictionary format that enables the system to automatically enforce standard syntactic checks and also allows users to self-describe their phenotypic file format, which the system then enforces. Users upload a zip file comprised of comma- or tab-separated files containing data for NRGR and user-defined dictionaries. AutoQC verifies that users' data submissions satisfy the requirements defined in the corresponding data dictionaries, and performs a variety of additional checks (e.g., individuals identified as fathers are males). In addition, AutoQC implements advanced curation steps that verify users' data against the center's database, ensuring that clinical data is correctly linked to an individual's biosamples, that individuals' diagnosis codes conform to DSM standards, referential integrity constraints across multiple files, and other checks. AutoQC generates a web-based report in real-time for investigators to review. Error-free data submissions are forwarded to experts for final manual curation, for checks which haven't yet, or cannot be completely automated.

Each user submission is implemented as a distributed workflow managed by Pegasus WMS. This facilitates the addition of new curation checks into the framework. If new checks are later added, we can easily re-submit the existing, accepted distributions, resulting in improved quality of the overall datasets.

## PgmNr 1489: Interpretable neural networks for schizophrenia risk prediction based on whole exome sequencing data.

### Authors:

A. van Hilten<sup>1,2</sup>; S.A. Kushner<sup>3</sup>; W.J. Niessen<sup>1,2,4</sup>; G.V. Roshchupkin<sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands; 2) Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands; 3) Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands; 4) Department of Imaging Physics, Delft University of Technology, Delft, The Netherlands; 5) Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands

---

**Objective:** Schizophrenia is a highly heritable mental disorder with a complex etiology. Recent studies have associated more than 100 different loci with schizophrenia, demonstrating the complex polygenic nature of this psychiatric disorder. Modern methods, such as neural networks, allow exploration of complex data with higher-level dependency. Applying such an approach might yield novel insights into the etiology and genetic risk of schizophrenia.

**Methods:** The prediction model was built by training a neural network using a Swedish case control study with 4969 cases and 6245 controls, involving ~1,2 million exome variants. Variant data was reduced to 21,390 genes using a special network layer developed for omics data analysis.

**Results:** The model reached an area under the curve of 0.65 and an accuracy of ~60% in the test set, with a theoretical maximum ~74% upper limit of performance calculated on the basis of the prevalence of schizophrenia and the concordance rate in monozygotic twins. The interpretability of the network allowed us to identify the genes most crucial for network prediction of schizophrenia versus control. The five most discriminant genes were: *HLA-C*, *TTN*, *TRY2P*, *HLA-A* and *LINC00226*.

**Conclusion:** We developed an interpretable neural network architecture for exome data, which incorporates prior biological knowledge (i.e., gene annotations), and can be easily extended to include genomic, tissue, cell type, or other functional annotations. Given that every node in the network is interpretable, we anticipate this approach as having the potential for uncovering novel insights into the genetic architecture of complex diseases, such as schizophrenia.

## **PgmNr 1490: Accurate imputation of summary statistics for non-additive models can assist in the discovery of new, and fine-mapping of existing, GWAS signals.**

### **Authors:**

S. Bacanu <sup>1</sup>; N. Daskalakis <sup>2</sup>; C. Nievergelt <sup>3</sup>; V. Vladimirov <sup>1</sup>; A. Amstadter <sup>1</sup>; B. Riley <sup>1</sup>; K. Kendler <sup>1</sup>; K. Ressler <sup>2</sup>; C. Chatzinakos <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) VIPBG, Virginia Commonwealth Univ, Richmond, Virginia.; 2) Mclean Hospital, Harvard University, Boston MA; 3) UCSD, San Diego, CA

---

Methods for direct imputation of summary statistics, e.g. our group's DISTMIX tool, were shown to be practically as accurate as genotype imputation method, while incurring orders of magnitude lower computational burden when it was paired with a heavily processed reference panel. However, due to the large linkage disequilibrium (LD) between additive summary statistics of neighboring variants, such tools are of limited use in fine-mapping signal regions. As signals for non-additive models (e.g. recessive) decays faster than the typically used additive statistics, imputing summary statistics for non-additive models might assist in fine-mapping of signals. Thus, there is a great need i) for much larger and diverse panels (that can also facilitate imputation of non-additive models, e.g. recessive) in processed form and ii), for accurate LD estimation between variants, to automatically estimate the ethnic composition of the cohort. Thus, for exhaustive analysis of even cosmopolitan cohorts, we propose DISTMIX2 method/software. When compared to DISTMIX and its competitors, the proposed method adds a i) much larger and diverse reference panel, ii) imputation of summary statistics for non-additive models (recessive and dominant) and iii) novel estimation for weights of ethnic mixture based solely on Z-scores (when allele frequencies are not available). To build a larger and more diverse reference panel, we heavily processed the publicly and privately available data to obtain a 33K subject panel which includes ~11K Han Chinese. The non-additive summary statistics are imputed assuming Hardy-Weinberg Equilibrium (HWE) between variants, which is quite reasonable given the HWE testing of variants in GWAS. We apply DISTMIX2 to the GWAS summary statistics from the Psychiatric Genetic Consortium (PGC). Our method uncovered signals in numerous new regions, with most of these findings coming from the rarer variants. For instance, while there were only three (rather marginal signals) in PGC2 PTSD, our method uncovered four large signals (three of them associated with rare variants). Even more, the largest signal (p-value<1e-42 in PKN2 gene) suggests a recessive mode of inheritance for this locus.

## PgmNr 1491: Can machine learning aid in discovery of disease genes? The case of autism spectrum disorder.

### Authors:

M. Gunning<sup>1,2,3</sup>; P. Pavlidis<sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Graduate Program in Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada; 3) Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada

---

Machine learning (ML) algorithms utilizing gene networks have been proposed as an approach for disease gene discovery. For example, for autism spectrum disorder (ASD) ML has been positioned as a supplement to traditional genetic association-based approaches to predict risk genes. However, work in our lab has raised questions about the general utility of ML methods for such tasks due to biases in gene networks and poor real-world performance. Here we report a computational evaluation of eleven published ASD gene prioritization methods, including four “pure” ML algorithms (incorporating gene network information but no genetic association data), four that rely solely on genetic association (as a control), and two “hybrid” ML-genetic methods. For evaluation we took advantage of a set of novel high-confidence ASD genes ( $n=40$ ) recently reported by Satterstrom et al. (2018) and Ruzzo et al. (2018) and not used for training the ML algorithms. We also examined a set of 86 previously reported high-confidence genes, many of which had been used to train the ML methods. We hypothesized that if the ML methods are effective, they should recover both the “known” and “novel” genes with high precision. Instead we found that, compared to genetic association methods, the pure ML methods perform relatively poorly both at recovering previously identified ASD genes (maximum precision at 20% recall [P@20R] 5% compared to 100% for genetics) as well as novel genes (1% vs 7% for genetics). In agreement with our previous reports, most of the ML methods also exhibit bias towards highly-studied and/or highly-connected genes (e.g.  $r=0.44$ ), which implies they are learning generic features of gene networks, not specific to ASD. The hybrid methods perform at best slightly better than simply using genetic association (maximum P@20R known=94%; novel=6%). Importantly, a generic gene-level feature, the haploinsufficiency score (pLI) has comparable performance to sophisticated ML methods (P@20R known=7%; novel=1%). The poor performance and bias of the ML methods compared to this naïve approach suggests there is still a substantial challenge in providing useful ML-based prioritization above that provided by direct genetic analysis.

## PgmNr 1492: Study of brain cell types underlying schizophrenia through a machine learning approach.

### Authors:

Q. Wang<sup>1</sup>; R. Chen<sup>1</sup>; Q. Wei<sup>1</sup>; Y. Ji<sup>1</sup>; Y. Jiang<sup>1</sup>; B. Li<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt University, Nashville, TN; 2) Vanderbilt University Medical Center, Nashville, TN

---

The understanding of genetic architecture of schizophrenia (SCZ) has significantly improved in the past decade. The genetic risks of SCZ are mostly attributable to common variants. However, very little “actionable” knowledge has been established to guide subsequent experimental modeling. One important reason is that the molecular and cellular complexity underlying SCZ remains poorly understood. Understanding of genetic risks at cellular level is crucial, as it would make it possible to identify the cells that are fundamental to the genesis of the disease. The recent development of single-cell RNA sequencing (scRNA-seq) technologies provide us a unique opportunity to dissect the genetic risks of SCZ in a cell-type specific manner. However, mapping the genetic findings to the specific cell types that they best fit to is not an easy task. Few accurate SCZ genes are available for mapping due to the factor that the majority of SCZ-associated SNPs identified by GWAS lie in non-coding regions, making it is not straightforward to infer the exact genes potentially affected by the associated risks. To address this challenge, we proposed a machine learning approach to simultaneously identify disease genes and the cell types they are active in using a bi-clustering algorithm. The basic assumption of our approach is that different cell types contribute to the heterogeneity of disease pathology by affecting different groups of disease genes. The application of our approach to brain scRNA-seq and SCZ GWAS data identifies different disease gene modules and the relevant cell types. We found the majority of SCZ disease genes were highly expressed in neuron cells, confirming previous findings. In addition, we also identified disease gene modules that were active in oligodendrocytes and endothelial cells. Disrupted synaptic connectivity has often been observed in SCZ patients and emerging evidence suggests that this is a result of dysfunctional oligodendrocytes. Also, changed brain endothelial cell gene expression has been observed in SCZ patients. The disease genes associated with neuron cells did not overlap with that of oligodendrocytes and endothelial cells, suggesting that different cell types contribute distinctively to SCZ etiology. These results also show that our approach is a powerful tool to dissect genetic heterogeneity underlying disease pathology.

## PgmNr 1493: Markov affinity-based proteogenomic signal diffusion (MAPSD) reveals dynamics of multi-omics data integration in brain disorders.

### Authors:

A. Doostparast Torshizi <sup>1</sup>; K. Wang <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Raymond G. Perelman Center for Cellular and Molecular Therapeutics, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

---

Heterogeneous pathophysiology of brain disorders, such as schizophrenia (SCZ), autism spectrum disorders (ASD) and Alzheimer's disease (AD), necessitates taking a holistic perspective through extracting knowledge from diverse sources of data modalities. Extensive genome wide association studies (GWAS), genome/exome sequencing studies and transcriptomic studies have led to a large pile of biological knowledge on various brain disorders. Tissue or cell type specific epigenomics and proteomics data are also accumulating rapidly on brain disorders, through large-scale consortium efforts such as PsychENCODE and Human Protein Atlas. How to understand complex brain disorders in the contexts of multi-omics and multi-tissue data remains a challenge. To account for such a critical need, we propose Markov affinity Proteogenomic Signal Diffusion (MAPSD), a novel multi-omic systems biology model to identify and prioritize disease-associated proteins given existing biological evidences including: GWAS hits, rare variants association hits, *de novo* mutations, differentially expressed genes, differential open chromatin and differential methylation data in cell type specific manner. MAPSD superimposes the existing evidences on human protein interactome or any arbitrary protein-protein interaction (PPI) networks and diffuses the disease signals across the network using Laplacian characteristic of network adjacency matrices which acts as a low-pass filter, and is able to remove weak signals from distant proteins (nodes) arrived at un-annotated proteins. MAPSD signifies critical features governing protein function including subcellular localization and normal abundance of proteins in various microdomains. Using data from the Human Protein Atlas, we investigated three brain regions including cerebellum, cerebellar cortex and hippocampus as well as multiple different cell-types such as neuronal cells, glial cells, Purkinje cells, endothelial cells, and the cells in granular and molecular layers. We demonstrated that existing biological knowledge can facilitate multi-omics data integration and improve the understanding of proteome data, and highlight contributions from different brain regions or cell types in different brain disorders.

## PgmNr 1494: Dissecting phenotypic variation in PTSD using genetically-regulated transcriptional variation.

### Authors:

N.P. Daskalakis <sup>1</sup>; C. Chatzinakos <sup>1</sup>; A.A.K. Dehaghi <sup>2</sup>; V. Michopoulos <sup>3</sup>; A. Lott <sup>3</sup>; B. Bradley <sup>3</sup>; T. Jovanovic <sup>3</sup>; K.J. Ressler <sup>2</sup>; M. Kellis <sup>4</sup>; Y. Li <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Depression and Anxiety Disorders, Harvard Medical School/ McLean Hospital, Belmont, Massachusetts.; 2) School of Computer Science, McGill University; 3) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia; Yerkes National Primate Research Center, Atlanta, Georgia; 4) Massachusetts Institute of Technology, Cambridge, MA

---

**Background:** PTSD occurs in some trauma-exposed individuals, but its molecular basis remains uncharacterized. Previously, we have associated genetically regulated gene expression (GReX) across tissues with PTSD revealing the importance of RNA splicing in the prefrontal cortex for PTSD biological mechanisms (Huckins et al. Under review 2019). We additionally, established that brain GReX is more closely associated with brain biomarkers (i.e., neuroimaging), while blood GReX is closer to blood biomarkers (i.e., observed blood gene expression and peripheral cytokines) (Chatzinakos et al. In Preparation 2019). Here, we study the relationship between genetically-regulated transcriptional variation, and phenotypic variation in PTSD, in a cohort of 9400 inner-city, low-socioeconomic-status, primarily-African-American patients of the Grady Memorial Hospital.

**Methods** Each individual was ascertained phenotypically using interview-based assessments, self-reports and lab tests. In addition, we imputed gene expression across 48 tissues using GTEx project's eQTL maps. To study the comorbidity patterns of transcriptomic and phenotypic information across individuals, we used MixEHR, a Bayesian unsupervised learning method (Li & Kellis, 2018). **Results** We identified 3 latent disease topics predictive of PTSD diagnosis (absolute LASSO coefficient > 5).

Based on symptom severity and trauma exposure, 2 of the topics were related to vulnerability (one with high childhood trauma and depression) and one related to resilience. We found enrichments for genes and tissues in the disease topics across multiple tissues.

Top brain tissues with the highest PTSD associations are dorsolateral prefrontal cortex and anterior cingulate cortex, which have been associated with cognition, emotion/stress regulation and PTSD. We also constructed a ranked list of PTSD genes across multiple disease topics. The top genes were: adipose and cardiovascular GXYLT1, a glucoside xylosyltransferase, blood SLC35A4, a glucose transporter, and brain genes related to post-transcriptional and post-translational modifications.

**Conclusions** By combining transcription imputation across the body with Bayesian machine learning in a trauma-exposed sample, we discovered tissue-types and genes that yield new biological insights into the genetic and phenotypic architecture of PTSD.



## PgmNr 1495: Deep-learning-based prediction of late age-related macular degeneration progression.

### Authors:

Q. Yan <sup>1</sup>; D.E. Weeks <sup>2,3</sup>; H. Xin <sup>1</sup>; H. Huang <sup>4,5</sup>; A. Swaroop <sup>6</sup>; E.Y. Chew <sup>7</sup>; W. Chen <sup>1,2,3</sup>; Y. Ding <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Pulmonary Medicine, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, PA; 4) Department of Electrical and Computer Engineering, Swanson School of Engineering, University of Pittsburgh, PA; 5) Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, PA; 6) Neurobiology, Neurodegeneration and Repair Laboratory, National Eye Institute, National Institutes of Health, Bethesda, MD; 7) Division of Epidemiology and Clinical Applications, National Eye Institute, National Institutes of Health, Bethesda, MD

---

Age-related macular degeneration (AMD) is a leading cause of incurable blindness worldwide. Previous genetic studies have shown the progression time to late AMD could be affected by genetic variants and environmental factors. The Age-Related Eye Disease Study (AREDS), a clinical trial from the National Eye Institute, includes genome-wide genotyping data, longitudinal color fundus photographs, and disease severity assessment over a period of 12 years, providing an unprecedented opportunity to investigate prediction models for progression to late-stage AMD. In this report, we jointly used genotypes and fundus images to dynamically predict an eye as having progressed to late AMD (e.g., whether the time to progression to late AMD for the eye exceeds 2 years from the current visit) with a modified deep convolutional neural network (DCNN). In total, we used 31,262 color fundus images in field 2 (centered above the macula) from 1,351 subjects with corresponding genotypes and phenotypes available at baseline and follow-up visits. The first part of this model was derived from a CNN to extract features. After obtaining the output vector of the final convolutional layer, we imputed these extracted image features to a fully connected layer for classifying AMD severity. Thereafter, we incorporated this estimated severity along with reported genetic variants to another layer for predicting the probability of progression to late AMD. Our results showed that the color fundus photos coupled with genotypes could predict late AMD progression with an averaged area under the curve (AUC) value of 0.85 (95%CI: 0.83-0.86). The results using fundus images alone showed an AUC of 0.81(95%CI: 0.80-0.83). We have validated our results in an independent dataset of 200 Caucasians extracted from UK Biobank and the results showed an AUC of 0.9 (95%CI: 0.85-0.94) for predicting whether the eye progresses to late AMD exceeding the 3 years.



## PgmNr 1496: Using phenotypic heterogeneity to model the pathogenic contribution of individual *USH2A* alleles.

### Authors:

A.P. DeLuca<sup>1,2</sup>; S.S. Whitmore<sup>1,2</sup>; E.M. Stone<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Institute for Vision Research, The University of Iowa, Iowa City, Iowa.; 2) Ophthalmology and Visual Sciences, The University of Iowa, Iowa City, Iowa.

---

Phenotypic heterogeneity provides a significant challenge to the interpretation of genetic testing results. *USH2A* is both the most common cause of type 2 Usher syndrome and the most common cause of non-syndromic retinitis pigmentosa (RP), yet little is known about which combinations of alleles in this recessive gene will lead to which disease. We present a technique to model the pathogenic contribution of different alleles using only the genotypes and referring diagnoses of cases gathered from public sources. We employ this model to predict diagnosis based on genotype.

Genotypes and referring diagnosis (Usher or RP) were gathered from *USH2A*-positive cases from the Leiden Open Variation Database as well as 20 published cohorts. Individuals were excluded: 1) if they did not have exactly two reported alleles, 2) if any of their reported alleles was unique in the dataset, and/or, 3) if their referring diagnosis was unclear. This yielded 866 patients (71.7% Usher syndrome). From these patients, a system of inequalities was constructed with coefficients for each allele. This system of inequalities was based on the assumptions that the alleles act in an additive fashion and that Usher syndrome is more severe than non-syndromic RP. This system was solved using a linear programming solver. To evaluate the performance of the model, leave-one-out cross-validation was performed.

The model correctly predicted the phenotype of the patient 84.6% of the time. Additionally, the optimization divided alleles into three types: 1) alleles that always cause type 2 Usher Syndrome when opposite any disease-causing mutation; 2) alleles that cause Usher syndrome when in combination with a similar variant; and, 3) alleles that only cause Usher syndrome when opposite a strongly deleterious allele.

Knowledge of the pathogenic contribution of individual alleles in recessive disease will ultimately allow physicians to provide more accurate prognostic information and facilitate the identification of outliers caused by modifying factors.

## PgmNr 1497: Automated variant ranking: A valuable tool for Mendelian disease analysis.

### Authors:

E. Zampaglione; M. Maher; E. Place; N. Wagner; C. Finn; D. Navarro-Gomez; E. Pierce; K. Bujakowska

[View Session](#) [Add to Schedule](#)

**Affiliation:** Ocular Genomics Institute, Massachusetts Eye and Ear, Department of Ophthalmology, Harvard Medical School, Boston, Massachusetts, MA 02114, USA

---

### Purpose:

Inherited retinal diseases (IRDs) are highly heterogeneous, with over 250 genes known to cause photoreceptor degeneration, following all modes of Mendelian inheritance. High throughput panel sequencing has been increasingly used in the diagnosis of IRDs, with ~50% of cases being genetically solved by single nucleotide variants (SNV) or small insertions/deletions in the exons of known IRD genes. With the plethora of NGS data coupled with a high rate of straight-forward solutions, there is a need for bioinformatics tools to address the repetitive and time-consuming process of variant analysis. To this end, we have developed a Mendelian Analysis Toolkit (MATK), an automated variant ranking software with the purpose of predicting the most likely causal variants for a given patient.

### Methods:

MATK uses a primary input of annotated VCF files, aggregating information from sources such as HGMD, ClinVar, gnomAD, and CADD scores, and providing customizable weights to the information based on expert input. It produces a ranked list of potentially disease-causing variants based on the weighted score. The tool can also perform pedigree aware analysis as well as integrate CNV predictions into the ranking. To determine if MATK performed similarly to a human analyst, the tool was used to prioritize variants in >600 panel-sequenced IRD subjects, a subset of which was compared against expertly curated variants from samples analyzed individually.

### Results:

Considering only first ranked variants, MATK found that 15% of cases were predicted as solved by variants in *USH2A*, 6% by *EYS*, and 4% by *ABCA4*, which corresponds roughly to expected prevalence rates of these genes in IRD etiology. When comparing the tool's output to expert curation, MATK's first or second ranked variant matched human analysts' choices in >90% of cases. The tool was particularly good at identifying and ranking high likely disease-causing variants in cases of recessive inheritance due to loss of function variants.

### Conclusions:

Automated variant ranking is a valuable bioinformatics tool that can expedite the process of identifying likely solving variants in IRD subjects, as shown by a high concordance of MATK with expert curation of individual samples. The ability to programmatically rank variants in large cohorts helps to remove variability that comes from human error and researcher bias, and can be used to avoid tedious re-analysis in the event that variant annotations are added or updated.

## PgmNr 1498: An hybridized *in-silico* model for *Plasmodium falciparum* mRNA seq data analysis elucidates enriched function for several vital unclassified genes.

### Authors:

M. Adebisi <sup>1,2,3</sup>; A. Adebisi <sup>2,3</sup>; O. Olugbara <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Information Technology, Durban University of Technology, Durban, South Africa; 2) Department of Computer Science, Landmark University, Omu-Aran, Kwara-State. Nigeria; 3) Department of Computer and Information Sciences, Covenant University, Ota. Ogun State, Nigeria

The era of genomics and microarray technology has ushered in a new dawn of tremendous amount of biologically useful data. This has led to various forms of research; there are various ways of analyzing such data, one of which is clustering. In this study a Hidden Markov Models, an aspect of clustering technique is presented. HMMs are useful for gene expression analysis for two reasons. The first is that the model has a very good theoretical and mathematical structure. Secondly, HMMs exhibits the Markov property, which allows time dependencies to be considered between elements; this makes it very effective for solving problems related with time series data.

In this work, HMM forward-backward algorithm and the Baum-Welch training algorithm was implemented. The *Plasmodium falciparum* RNA-seq expression data (Otto et al, 2010) was the time series data deployed on HMM. Baum-Welch algorithm was used to train the dataset to determine the maximum likelihood estimate of the Hidden Markov Model parameters. The left-to-right HMM with states varying from 2 to 10 and three observations was constructed. Cluster validation was conducted by performing a likelihood ratio test. This test makes it possible to compare the results of the simpler cluster to that of the complex cluster in order to know when to stop adding more paths and parameters. This method proved to be able to determine the optimal number of clusters in the dataset.

The fitted HMM identify three clusters from the dataset. Cluster one had 506 genes, cluster two had 488 genes and cluster three 701 genes. Sixteen functional enrichment genes were found in the cluster set. Cluster one had four GO functional annotations with percentage of genes ranging from 8.3 to 1.9%, cluster two had five GO functional annotations with percentage of genes ranging from 7.8 to 1.7% and cluster 3 had the highest GO annotation with seven functions and gene percentage ranging from 3.2 to 1.5%. The HMM clustering results was compared with the traditional k-means result and discovered that it performed more efficiently in the sense that the clustering results were more evenly distributed when compared to the K-means result. This research work has successfully cluster the *plasmodium falciparum* gene expression dataset using the Hidden Markov Models, it also obtained enriched functions for other unclassified genes from the cluster set.

## PgmNr 1499: Development and application of methods for resolving molecular diagnoses from patient sequence data.

**Authors:**

D. Alyousfi

[View Session](#) [Add to Schedule](#)**Affiliation:** Human Genetics, University of Southampton, southampton, United Kingdom

---

Identifying molecular causes of disease from sequenced genomes can be extremely challenging usually requiring tiered filtering with the possibility of missing the causal variant(s). This study is concerned with integrating predictors of gene essentiality to understand how gene-level approaches could improve filtering of clinical sequence data. We assessed relationships between essentiality measures in different gene groups including non-essential genes not involved in disease, Mendelian disease genes and genes classed as essential. By examining a range of well-studied and independent gene-level scores, our results suggest that individual methods are less effective than combined scores for predicting genes involved in Mendelian diseases. The results of principal component analysis show that our combined Essentiality Specific Pathogenicity Prioritization (ESPP) score is a more powerful predictor than individual scores. The ESPP score explains more than 40% of the variation across eight diverse scores. Interestingly, the *CHD3* gene was given a relatively high score  $\sim 4.77$  in the classifier, but has only very recently been recognised as having an associated Mendelian phenotype. Success of the classifier is constrained by incomplete understanding of which genes should be assigned to the different gene classes but as gene function is increasingly understood the utility of the classifier should increase. Ultimately, these findings reveal that our combined score better reflects the modularity of many genetic properties measured by each score separately and might help improve diagnostic rates.

## PgmNr 1500: New candidate *Alu* sequences in AK1 Korean reference genome.

### Authors:

K. Bahk; J. Kim; J. Sung

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genome Epidemiology & Health Big Data, Department of Epidemiology, Graduate School of Public Health, Seoul National University, Seoul, Korea

**BACKGROUND:** The *Alu* element family is densely populated repetitive sequences, particularly abundant in primates and humans, characterized by *Alu* restriction endonuclease., with typical ~300bp length. With recent discoveries that *Alu* elements not only have roles in the phylogenetic and evolutionary implications but some functional significances, we attempted to search whether new types of *Alu* elements exist in the ethnic genome. The AK1 is one of the highest quality genome assemblies with haplotig N50 > 16Mb, based on one Korean. Our previous studies showed that 1.8% of total AK1 sequences do not exist in the GRCh38p12, including the alternative and random sequences. We did find about ~3,000 candidate regions with putative coding functions in this non-overlapping 1.8% segments of AK1, but it is elusive whether some unidentified repetitive elements such as *Alu* elements or SINEs/LINEs exist.

**METHODS:** We used the RepeatMasker program, that screens DNA sequences for interspersed repeats, with RMBlast search engine to find *Alu* candidates on these non-overlapping scaffolds. We applied BLASTN to the candidate *Alu* to identify the similarity of subfamily and sequence between the candidate *Alu* elements and the *Alu* in the nucleotide database. Further validation by Sanger sequencing is still ongoing for new types of candidate *Alu* sequences.

**RESULTS AND INTERPRETATIONS:** A total of 371 candidate *Alu* elements were found from the initial search using non-overlapping AK1 genome regions and met the criteria (e-value < 0.001) by the BLASTN comparisons. The *Alu* elements comprise ~3,300 Kbps, indicating that the average number of repeats were ~8.9K. Among 371 *Alu* candidates 343 *Alu* elements were classified into *AluY* (with 8 *AluJb* and 20 *AluSx*) suggesting that the most of the new candidate *Alu* elements are the results of recent evolutionary changes. We are verifying ~20 novel types of *Alu* elements by Sanger sequencing, and further searching the location in the whole genome in the AK1. We expect the identification of new *Alu* subfamilies would help understand the recent human evolutions and ethnic differences in the genetic susceptibilities to diseases.

## PgmNr 1501: Findings from the Critical Assessment of Genome Interpretation, a community experiment to evaluate phenotype prediction.

### Authors:

C. Bakolitsa <sup>1</sup>; G. Andreoletti <sup>2</sup>; R.A. Hoskins <sup>1</sup>; P. Radivojac <sup>3</sup>; J. Moulton <sup>4,5</sup>; S.E. Brenner <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of California, Berkeley, CA 94720; 2) University of California, San Francisco, CA 94143; 3) Northeastern University, Boston, MA 02115; 4) Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD 20850; 5) Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742

---

Interpretation of genomic variation plays an essential role in the analysis of cancer and monogenic disease, and increasingly also in complex trait disease, with applications ranging from basic research to clinical decisions. Yet the field lacks a clear consensus on the appropriate level of confidence to place in variant impact and interpretation methods. The Critical Assessment of Genome Interpretation (CAGI, 'k?-j?') is a community experiment to objectively assess computational methods for predicting the phenotypic impacts of genomic variation. CAGI participants are provided genetic variants and make blind predictions of resulting phenotype. Independent assessors evaluate the predictions by comparing with experimental and clinical data.

CAGI has completed five editions with the goals of establishing the state of art in genome interpretation and of encouraging new methodological developments. Challenges have been predominantly based on human data, mirroring problems in clinical practice and much basic research. The focus has been on interpreting nonsynonymous variants, splicing variants, structural variation, whole-exomes and whole-genomes, with phenotypes ranging from molecular and cellular measurements to organismal phenotypes in inherited disease and cancer. Results from previous CAGI experiments have been described in two special issues of *Human Mutation*.

Each previous edition of CAGI has revealed new aspects of the methods, and there has been significant progress in several areas. Independent assessment has found that top missense prediction methods are highly statistically significant, but individual variant accuracy is limited. Missense methods tend to correlate better with each other than with experiment. Bespoke approaches often enhance performance. Interpretation of non-coding variants shows promise but is not at the level of missense.

In challenges using clinical data, such as the Johns Hopkins and SickKids challenges, predictors identified causal variants overlooked in the initial clinical analysis. Data from the CHEK2 challenge were recently presented to the ClinGen Sequence Variant Interpretation Working Group to explore stronger weighting when methods demonstrate sufficient reliability for clinical use. CAGI results suggest that running multiple uncalibrated methods and considering their consensus may result in undue confidence, so we advise against this.

Detailed information about CAGI may be found at <https://genomeinterpretation.org>.

## PgmNr 1502: Target interval builder: A tool to obtain and annotate clinically-significant intervals.

### Authors:

J.E. Barrows<sup>1,2</sup>; M. Croken<sup>1,2</sup>; R.B. Calder<sup>1,2</sup>; A. Salomatov<sup>1,2</sup>; G. Mendiratta-Vij<sup>1,2</sup>; F.E. Suer<sup>1,2</sup>; L.J. Edelmann<sup>1,2</sup>; E. Schadt<sup>1,2</sup>; A.S. McLellan<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sema4, a Mount Sinai venture, Stamford, CT, 06902; 2) Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029

---

**Background:** Hybrid capture-based target enrichment uses custom-designed oligonucleotide probes to capture specific genomic regions for NGS. Target Interval Builder (TIB) implements all computational steps required to specify target regions for a list of HGNC gene IDs. TIB creates versioned BED files with annotations for known/potential disease-causing variants from HGMD/ClinVar and NCBI RefSeq transcripts (exons, introns, CDS, UTR). These versioned BED files are used to design NGS capture library probes and to mask clinically excluded genes/regions during data analysis.

**Methods/Results:** TIB is implemented using AWS Batch/EC2/EBS/S3, Broad Institute WDL/Cromwell, bedtools, BCftools, Docker, biomaRt, Nginx, Django, and custom-developed Python and Bash scripts. Six Docker images, stored in AWS ECR, encapsulate TIB functionality.

Users provide a text file of HGNC gene IDs and optionally BED files of regions to include or exclude via an AWS-hosted web form. TIB can generate a new reference BED file or use a specified BED file. The software builds the reference BED file based on data from UCSC Genome Browser, HGNC (genenames.org), Ensembl, ClinVar, and HGMD. TIB uses a previously downloaded snapshot stored on AWS S3 if it cannot access a source database. It versions all files using a clear naming convention.

When designing a set of new capture targets, TIB searches the reference BED file for genomic features using the HGNC gene IDs provided at runtime. It filters out non-protein-coding transcripts, if protein-coding transcripts are available, and splits the resulting BED file to generate BED files with and without UTRs and introns. Moreover, metrics are gathered, over the chosen intervals, for parameters that impact NGS performance, e.g., segmental duplications, repeats, and read alignment uniqueness. TIB deposits output files in AWS S3.

TIB annotates intervals with the HGNC approved gene symbol, HGNC ID, and source. Optional annotations include RefSeq transcript, exon number, HGMD accession, Ensembl transcript ID, database timestamp, HGVS variant, ClinVar ID, and type (intron, CDS, etc.).

**Conclusion:** TIB annotates genomic intervals to facilitate design and filtering of targeted NGS capture libraries. Standardized annotation retains interval provenance throughout library design. Poorly performing intervals are intersected with genes, variants, and regions. NGS data can be subset at runtime by filtering annotated intervals against a custom list of HGNC IDs.



## PgmNr 1503: Variant reclassification in a routine diagnostic setting on a highly heterogeneous patient population.

### Authors:

P. Bauer; G. Oprea; A. Beaty-Faber; V. Weckesser; C. Beetz; A. Rolfs

[View Session](#) [Add to Schedule](#)

**Affiliation:** Centogene AG, Rostock, Germany

---

The 2015 guidelines by the American College of Medical Genetics probably represent the most widely accepted framework for classification of genomic variants. The contained categories are: pathogenic (P), likely pathogenic (LP), variant of uncertain significance (VUS), likely benign (LB) and benign (B). With additional relevant information becoming available over time, variants may need to be reclassified.

Along our routine diagnostic workflow, variants that need re-consideration are automatically flagged. Triggers include (i) findings in additional family members, (ii) identification in an unrelated patient, (iii) a change in internal population frequencies, and (iv) availability of biochemical/metabolic data. Subsequent semi-automated review of all ACMG-relevant information may or may not result in eventual reclassification.

We will present a re-classification data set that is based on a 5-months period, covers 248 variants, and has a highly heterogeneous background as regards the ethnic, phenotypic and mutational spectra. 47 reclassifications did not immediately impact on the diagnostic interpretation (i.e. L↔LP and B↔LB); 43 of these (91%) were upgrades from LP to P. Of the remaining 201 reclassifications, n=99 (49%) and n=44 (22%) were VUS to LP/P and VUS to LB/B, respectively. Substantial additional traffic was observed from LP/P to VUS (n=34; 17%), while reclassifications from LP/P to LB/B (n=17; 8%), LP/P to VUS (n=4; 2%) and LB/B to LP/P (n=3; 1%) were much rarer. The reclassifications affected a total of 840 patients (median=2; range 1 to 99). With the exception of LN↔N (2 variants vs. 100 patients), the proportion of specific reclassifications were highly similar between variants and patients.

Our data suggest that the major net effect of re-classification is a reduction of the VUS category. Also within LP/P and LB/B, most classifications become more definite. Despite the heterogeneous background of our data set, these findings are consistent with observations of others focusing on certain diseases or populations. They also confirm that reclassification is an important part of a sustained diagnostic process.

## PgmNr 1504: Identification of elusive copy number variation by targeted coverage depth analysis.

### Authors:

J.R. Belyeu <sup>1,2</sup>; B.S. Pedersen <sup>1,2</sup>; A.R. Quinlan <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, University of Utah, Salt Lake City, Utah; 2) USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, Utah; 3) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT

---

Copy number variation is a major class of human structural variation which affects spontaneous and complex disease as well as human diversity. Advances in whole-genome sequencing have driven the development of computational tools to identify copy number variants (CNVs), and implicated CNVs in many rare and common disease phenotypes.

However, while common, CNVs created by non-allelic homologous recombination (NAHR) between high-identity repeats remain difficult to identify with current methods. This is not surprising since discordant and split-read alignments rarely exist for these CNVs due to the flanking repeats that drive copy number variation, and evidence from sequencing depth is often confounded by the low mappability of the flanking repeats. Worse still, depth-based CNV detection tools typically scan arbitrarily-sized windows across the entire genome, thereby decreasing the likelihood that any indicative change in coverage be captured in a given window, especially when the CNV is small.

Instead, our approach (NAHRly) utilizes the prior knowledge of the locations and compositions of human genomic repeats to focus analysis on NAHR-prone regions. Using annotations of segmental duplication, we have identified ~1000 regions of the human genome with the potential for NAHR-driven copy number variation. To demonstrate the power of this approach, we analyzed 603 whole-genome samples from 33 three-generation families in the CEPH Utah cohort. We developed a peak-identification algorithm that separates depth clusters within the cohort for CNV detection and genotyping. We reveal deletions and duplications in many of the regions predicted to be NAHR-prone and used inheritance and transmission tests to evaluate our findings and tune our method. We report variants identified with this focused method that are missed by both CNVnator and SVTYPER. Our results suggest that NAHRly complements existing CNV detection techniques for the identification of CNVs that were previously very difficult to detect

## PgmNr 1505: A comparison of variant effect prediction algorithms and their performance on specific disease phenotypes.

### Authors:

S. Bhatnagar <sup>1,2</sup>; P.J. Dexheimer <sup>1</sup>; L.J. Martin <sup>3</sup>; S.M. Ware <sup>4</sup>; B.J. Aronow <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 2) Department of Electrical Engineering and Computer Science, University of Cincinnati, OH, USA; 3) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 4) Departments of Pediatrics and Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA

---

Popular variant prediction tools use functional, structural and/or conservation-based features to distinguish damaging variants from background variation. A tool's training variants for damaging behavior may derive from multiple diseases, including those with disparate pathogenetic mechanisms such as loss of function (LoF) or dominant negative effect. We hypothesized that prediction tools have intrinsic differences in performance based on genetic mechanism and disease type. To study this, we analyzed the performance of variant prediction tools by disease phenotypes available in ClinVar (1000+ phenotypes/traits; 100+ diseases; multiple biological systems). We selected ClinVar variants with evidence to assert consistent clinical interpretation (2 star or higher). For simplicity, we dichotomized the predictions by combining pathogenic with likely pathogenic calls and benign with likely benign calls. Variants were annotated through Variant Effect Predictor to gather predictions for the following tools: *Polyphen*, *SIFT*, *CADD*, *REVEL*, *M-Cap*, *MetaSVM*, *MetaLR*, *MutationTaster*, *MutationAssessor*, *FATHMM*, *FATHMM-MKL*, *PON-P2*, *CAROL*, *DANN*, *LRT*, *MTR*, *MPC*, *Condel*, *VEST3*, *PROVEAN*, *GenoCanyon*, *Eigen*, *Eigen-PC*, *GERP++*, *PhyloP*, and *SiPhy*. For each phenotype and all of its associated variants, we compared the ClinVar clinical significance of the variant to each tool's prediction. We computed sensitivity, specificity and F1-score (harmonic mean of precision and recall) per tool-phenotype combination. We also compiled common non-synonymous and LoF variants from 1000 Genomes (GnomAD MAF>0.01), which are assumed to be tolerated, and assessed each tool's performance similarly. The performance of these tools differs considerably by disease phenotype (mean F1-score: ~0.1-0.8). We observed that most of these tools had better sensitivity compared to specificity across disease phenotypes indicating that these tools are better at correctly predicting disease-causing variants compared to benign variants. Ensemble-based classifiers like MetaSVM and REVEL were among the top performing tools for complex diseases including several cardiovascular phenotypes, and limb girdle muscular dystrophy. CADD was a better performer among cancer phenotypes. Our results demonstrate that the ability of bioinformatic tools to predict clinical pathogenicity varies markedly by underlying phenotype. We speculate that explicitly considering underlying disease mechanisms may yield improved variant classification.

## PgmNr 1506: nanotatoR: An annotation tool for genomic structural variants.

### Authors:

S. Bhattacharya <sup>1</sup>; H. Barseghyan <sup>1,2</sup>; U. Boddeti <sup>3</sup>; E.C. Délot <sup>1,2</sup>; E. Vilain <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Children's National Hospital, Washington, District of Columbia.; 2) Department of Genomics and Precision Medicine, School of Medicine and Health Sciences, George Washington University, Washington DC 20052, USA; 3) Columbian College of Arts and Sciences, George Washington University, Washington DC 20052, USA

---

Whole genome sequencing (WGS) has been efficient in identification of single-nucleotide variants (SNVs) and small insertions and deletions (INDELs) but, due to utilization of short reads (~150 basepairs), not effective in identification of larger structural variants (SVs). Introduction of genomic Optical Mapping (OM), which utilizes long fluorescently labeled DNA molecules for *de novo* genome assembly and structural variant calling, substantially increased both sensitivity and specificity for identification of SVs. However, compared with short-read sequencing technologies, OM doesn't have many tools that annotate identified variants (population frequency, overlapping genes, etc.) for filtration and classification.

We developed an R-based annotation package that provides comprehensive information for SV filtration by integrating the patient's OM and expression data sets and reference databases information.

*nanotatoR* uses external databases DGV, gnomAD, 1000 genomes and DECIPHER, to estimate population frequency of the variants. Additionally, *nanotatoR* builds a database of SVs based on the samples provided by the user to estimate the variant's frequency within the user's cohort. User-provided BED files can be used to annotate which genes overlap or are near (upstream and downstream) the identified SV, as well as whether the SV affects specific exon/s or intron/s. Overlap percentages and distances for nearest genes are automatically calculated and can be used for filtration. A primary gene list is extracted from public databases (OMIM, DDD, dbVar, ClinVar, GTR) based on the patient's phenotype and compared with genes overlapping SVs, providing the analysts with an easy way to filter variants. *nanotatoR* also extracts from the GTEx database information about expression level in the sample tissue for the genes and non-coding RNAs overlapping an SV. If RNAseq expression data is available for the sample, it is also aligned to the SV map and displayed in a separate column. The output is given in an Excel file format, subdivided into multiple sheets based on SV type and inheritance pattern (INDELs, *de novo*, inversions, translocations, etc.).

The extensive annotation and the expression pattern of genes near/overlapping variants enables users to identify potential pathogenic SVs with precision with fast turnaround times (currently ~1hr for a human genome OM annotation)

## PgmNr 1507: Imputation after whole genome sequencing to get genotypes for SNPs from the GWAS Catalog.

**Authors:**

M. Bicak; R.J. Klein

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai Hospital, New York, New York.

---

While genome-wide association studies (GWAS) have identified large number of trait-associated SNPs, their functional basis remains unclear. To begin to understand the functional basis for SNPs involved in prostate cancer and inflammatory bowel disease, we asked if they were eQTLs in data from the GTEx Consortium. Surprisingly, upon first examination we found that some SNPs were not present in the GTEx genotype data. Genotypes in GTEx were measured with whole genome sequencing (WGS), which theoretically probes the full spectrum of genetic variation including both common and rare variants. However, some regions of the genome are less amenable to sequencing and may not pass quality control measures. Genotype imputation from genome-wide SNP data has been a widely used approach to “fill in” missing common genetic markers that are not directly genotyped. We hypothesized that such imputation could be useful with whole genome sequencing data to enable us to obtain a measurement for every SNP of interest. We compared imputed genotypes from UK10K+1000Genomes (UK10K+1000G) and Haplotype Reference Consortium (HRC) Panels against WGS genotypes from the GTEx data. We used the Sanger Imputation Server to perform imputation. We found that 92%, 93% and 97% of variants from the GWAS Catalog could be readily found in the WGS, HRC and UK10K+1000G panels respectively. Based on these findings, we propose that imputation with the UK10K+1000G reference panel after WGS may be the best approach to find previously reported SNPs from the GWAS catalog in GTEx and other whole genome sequenced datasets.

## PgmNr 1508: GeneBytes: Interactive and intuitive tool for genome analysis.

### Authors:

M. Cabo Diez; A. Gallego; A. Romera-Lopez; D. Cantalapiedra; R. Rodriguez; S. Lois

[View Session](#) [Add to Schedule](#)

**Affiliation:** Sistemas Genómicos, Valencia, Spain

---

Last years we have witnessed an economic and technological revolution in the clinical sphere. The evolution from the analysis based on genes panels to exomes, from exomes to the whole genome is due to the fact of the decrease of the sequencing cost and to the trend to include studies of personalized medicine as clinical routine, leading to a clinical Big Data, with a huge potential for scientific knowledge. Considering this large amount of information and the continuous time reduction demand on diagnostic, it's necessary to rely on tools able to properly manage and organize a massive volume of information.

With this aim arises GeneBytes, a tool for genomic data mining that has been designed and built by a multidisciplinary team of engineers and doctors. In a diagnostic context, GeneBytes allows the analysis of Structural Variants (balanced translocations and inversions), Copy Number Variants (CNVs) and point mutations, allowing also the possibility to design, save and automatically apply variant prioritization algorithms. This will drastically reduce the time around time per analysis. Moreover, this algorithms and the results obtained can also be exported, guaranteeing the traceability of the whole analysis process.

Filtering functions of the tool include more than 70 features related to variants, including information from to the last versions of reference databases such as HGMD Pro, Clinvar, 1000 genomes, GnomAD, ESP, information about in silico algorithms, clinical guidelands or information about inheritance patterns, giving versatility to the analysis. Among the viewing Features highlights the integration of IGV tool for the display of the sequence alignment, that will help to discriminate between true positive and false positive variants.

Finally, specific functions have been added for Structural Variants and CNV analysis, allowing, for example, to state if the alteration affects or not OMIM disease associated genes. Thus, GeneBytes allows the standardization of NGS genetic data analysis, helps with data interpretation and assures the traceability of the whole process.

## PgmNr 1509: Empirical performance of tree-based inference of phylogenetic networks.

**Authors:**

Z. Cao; J. Zhu; L. Nakhleh

[View Session](#) [Add to Schedule](#)

**Affiliation:** Computer Science, Rice University, Houston, Texas.

---

Phylogenetic networks extend the phylogenetic tree structure and allow for modeling vertical and horizontal evolution in a single framework. Statistical inference of phylogenetic networks is prohibitive and currently limited to small networks. An approach that could significantly improve phylogenetic network space exploration is based on first inferring an evolutionary tree of the species under consideration, and then augmenting the tree into a network by adding a set of 'horizontal' edges to better fit the data.

In this paper, we study the performance of such an approach on networks generated under a birth-hybridization model and explore its feasibility as an alternative to approaches that search the phylogenetic network space directly (without relying on a fixed underlying tree). We find that the concatenation method does poorly at obtaining a 'backbone' tree that could be augmented into the correct network, whereas the popular species tree inference method ASTRAL does significantly better at such a task. We then evaluated the tree-to-network augmentation phase under the minimizing deep coalescence and pseudo-likelihood criteria. We find that even though this is a much faster approach than the direct search of the network space, the accuracy is much poorer, even when the backbone tree is a good starting tree.

Our results show that tree-based inference of phylogenetic networks could yield very poor results. As exploration of the network space directly in search of maximum likelihood estimates or a representative sample of the posterior is very expensive, significant improvements to the computational complexity of phylogenetic network inference are imperative if analyses of large data sets are to be performed. We show that a recently developed divide-and-conquer approach significantly outperforms tree-based inference in terms of accuracy, albeit still at a higher computational cost.

## PgmNr 1510: Using Atlas-CNV to identify large and small deletions in clinical exomes.

### Authors:

T. Chiang <sup>1</sup>; J. Shen <sup>1</sup>; W. He <sup>1</sup>; M. Peng <sup>2</sup>; X. Wang <sup>1</sup>; Y. Yang <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) AiLife Diagnostics, PearLand, Texas.; 2) Veritas-Genetics, China

---

Identifying Copy-Number Variants (CNVs) in exome sequencing is an important analytical step in the clinical pipeline to achieve a higher molecular diagnostic rate. The conventional exome analysis is well suited for detecting single nucleotide variants (SNVs) and small indels but not CNVs. Current CNV calling from exome data has limited power to detect small changes (e.g. single-exon deletions), and therefore needs further improvement. A recent method to identify CNVs (Atlas-CNV) at the single-exon level was reported for capture targeted gene-panels (eMERGESeq) using normalized coverage among batch samples and single-exon metrics to control false calls. Here, we extend those capabilities to capture-based exomes and evaluate its effectiveness in a retrospective blinded study of 34 clinical exomes (9 probands, 7 trios, 1 quad), which include six families with known disease-causing CNVs. Combining the 34 samples as a batch with default calling thresholds, an average of 2,146 deleted exons were called per sample with average quality metrics per exon of  $E_{\text{StdDev}}=0.26$  (standard deviation of  $\log_2$  scores at an exon) and **C-score**=-5 (Z-like score of  $\log_2$  divided by  $E_{\text{StdDev}}$ ). The data analysis focused on identifying potentially disease-causing CNVs that are related to patient phenotypes and correctly called all the six clinically significant deletions: three large deletions encompassing multiple genes; 2q24.3, 4.8 Mb; 22q11.21, 2.5 Mb; 15q11-13, 3.7 Mb, and three small intra-genic deletions: RPS6KA3 (exon 1, chrX), HPS1 (exon 15-16, chr10), NF1 (exon 27, chr17). These deletions were confirmed by orthogonal methods including SNP analyses (n=2), low-pass whole-genome sequencing (n=2), and junction read analysis (n=2). The average  $E_{\text{StdDev}}$  and **C-score** were 0.26 and -4.7 respectively for the large deletions, and 0.29 and -3.3 for the small deletions. Although additional refinements and larger datasets are needed to identify exons with high frequency spurious calls, we conclude that Atlas-CNV has clinical utility to screen both large (>2Mb) and small deletions (single-target exon) in exomes.



## PgmNr 1511: Kibio, a scalable tool to unify, explore and visualize life-sciences massive data.

### Authors:

A. Droit <sup>1</sup>; M. Leclercq <sup>1</sup>; M.P. Scott-Boyer <sup>1</sup>; A. Dessmond <sup>1</sup>; F. Belleau <sup>1</sup>; R. Ongaro-Carcy <sup>1</sup>; O. Perin <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Medicine, CHU de Quebec Research Center, Quebec, Quebec, Canada; 2) Digital Sciences Department L'Oréal Advanced Research, International Microbiology Department, L'Oréal Research and Innovation

---

Linking information between biological databases on the web is a challenging task for any researcher, which needs to manage very large databases and cross the information between them. In this context, we present Kibio (<http://kibio.science>), a new resource that integrates dozen of biological public repositories in the form of a web portal. It offers search, exploration and data linking through dynamic visualizations. Kibio also provides a userspace system, allowing private databases upload and linking to the existing network of databases in Kibio. This extensible environment, based on an advanced visual filtering system, can be used to explore multidimensional Omics data, allowing interactions search, validations of experimental protocols, simplified literature search, hypothesis discovery, characterization of biomarker signatures, identification of non-directly related biological functions and much more. Furthermore, a custom Extract-Transform-Load (ETL) software called Lucan, has been developed to deal with biological data heterogeneity, hence ensuring the consistency between original biological databases and the version imported in Kibio. As of June 2019, Kibio integrates about 40 databases, reaching 2.1TB of public data and 2.5 billions documents whose some contains hundreds of fields.

## PgmNr 1512: PseudoNet: A graph representation learning method for inferring pseudogene functions by borrowing information from coding genes.

### Authors:

K. Fan <sup>1</sup>; Y. Zhang <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Informatics, Ohio State University, Columbus, OH.; 2) The Ohio State University Comprehensive Cancer Center (OSUCCC - James), Columbus, OH

---

Pseudogenes are ubiquitous and abundant in genomes. It has been recognized that some pseudogenes play essential roles in expression regulation of their parent genes, and might contribute to cancer development. Given the fact that experimental identification of pseudogene functions is time-consuming and expensive, it is important to develop computational methods to predict the functions of pseudogenes, and provide testable hypotheses for targeted experimental validation. Since the existing functional annotation of pseudogenes is limited, traditional supervised machine learning methods that require abundant training data are not suitable.

Here we propose PseudoNet, a graph representation learning model that can make use of multiple information and borrow information from coding genes to predict pseudogene function. Our graph-based method can consider both network structure (sequence similarity network) and gene attributes (expression profile) together to make predictions by organizing genes into an attributed network architecture. To be more specific, we first build sequence similarity networks of genes and pseudogenes by using BLAST. We then calculate the co-expression matrix using TCGA RNA expression profile (take BRCA for example), and run node2vec to obtain initial embeddings which are used as node attributes in our PseudoNet model. In summary, sequence information is used to construct network structure while expression profile information is used as node attributes in the networks, so that PseudoNet can consider both features. Then we utilize variational graph auto-encoder (VGAE), a deep learning-based network representation learning algorithm to learn ultimate informative representations for all genes and pseudogenes. The final step is to train a multi-label neural network classifier to predict pseudogene functions by training on coding genes using the learnt representations.

In our benchmark dataset, there are 1325 pseudogenes and 14565 coding genes, and we use Gene Ontology (GO) as the functional annotation. Our model could achieve higher than 0.75 in terms of AUC-ROC on all three ontologies, better than traditional machine learning methods such as SVM. What's more, the learnt informative representations can be used for various downstream analyses, such as performing clustering analysis for discovering inherent patterns, detecting association with drugs or diseases, and visualizing data using t-SNE and PCA.

## PgmNr 1513: Identifying novel regulatory elements using RELICS, a statistical framework for the analysis of CRISPR regulatory screens.

### Authors:

P.C. Fiaux <sup>1,2</sup>; H. Chen <sup>2</sup>; I. Luthra <sup>3</sup>; C. O'Connor <sup>4</sup>; G. McVicker <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics and systems Biology, University of California San Diego, San Diego, CA; 2) IBL, Salk Institute for Biological Studies, San Diego, CA; 3) Engineering, Simon Fraser University, Vancouver, BC, Canada; 4) Flow Cytometry Core, Salk Institute for Biological Studies, San Diego, CA

---

High-throughput CRISPR/Cas9 screens are a powerful new tool for the systematic discovery of regulatory elements in the human genome. In these regulatory screens, thousands of guide RNAs (gRNAs) are delivered to cells to target potential regulatory sequences for mutation, activation or inhibition. The cells are then sorted into high- and low-expression pools based on the expression of a target gene. While, these screens have the potential to perform unbiased discovery of regulatory elements, they generate noisy data and the performance of analysis methods has not been rigorously assessed. Here we describe RELICS, a statistical framework for **R**egulatory **E**lement **L**ocation **I**dentification in **C**RISPR **S**creens. RELICS models the observed guide counts in different expression pools with a generalized linear mixed model. This approach is very flexible, can jointly model multiple expression pools (beyond just high and low), incorporate variability across guides, and accommodate over-dispersion. To assess the performance of RELICS we have developed a simulation framework for generating CRISPR regulatory screen data and simulated 1000s of data sets under a wide variety of experimental and biological conditions. RELICS outperforms existing analysis methods on the simulated data and we have applied it to identify regulatory elements in several published datasets. In addition, we have applied RELICS to data from a paired-guide regulatory screen that we performed for *GATA3* in Jurkat T cells. We identify a total of 23 putative regulatory elements within the 2MB targeted region surrounding *GATA3*. Notably 16 of the identified elements lie within the same topological associating domain as *GATA3*, but only 3 overlap enhancers predicted by ChromHMM.

## PgmNr 1514: Secure storage and query of genomics data with blockchains.

### Authors:

G. Gursoy <sup>1</sup>; S. Wagner <sup>2</sup>; M. Gerstein <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut.; 2) Computer Science, Yale University, New Haven, Connecticut.

---

**INTRODUCTION:** The genomic characterization of millions of individuals promises to be useful for medical research. Having a larger number of studied individuals is assured to boost statistical power; therefore, genomes from more individuals likely will be sequenced going forward. One of the main barriers to expanding sequencing is the ability to provide a secure platform for storage of this sensitive data. While cloud computing offer solutions to maintain and compute on large amounts of data, security, data integrity, and robustness vulnerabilities such as the consequences of single-point-of-failure have not yet been addressed. A potential approach to circumvent these challenges is blockchain technology due to its security, immutability, and decentralization properties.

**OBJECTIVES:** Required storage space and computational power of blockchains is greater than a centralized database application, due to each of the blocks needing to contain the information of the rest of the preceding blocks. Decentralized system creates a higher latency during retrieval of data. The goal is to develop time and space-efficient blockchain-based ledgering solutions for storage and retrieval of genomic data.

**METHODS:** MultiChain is a private blockchain platform that offers data streams embedded in the chain for storage. Using this platform, we first developed a secure storage and query mechanism for genomic access logs (e.g dbGap logs) to prevent potential data misuse such as genomic privacy attacks. Based on the same framework, we then developed a series of tools called SAMChain and SCTools that allow users to efficiently store genome sequencing data (BAM files) in a blockchain, and provide analysis functions such as querying, read streaming, depth analysis, pile-ups for variant calling, and re-creating BAM files. To allow for rapid access, SAMChain is based on turning data streams into hash tables, in which keys are important features of the reads such as their location on the reference genome. These hash tables also allow us to minimize the off-chain operations in the local memory.

**RESULTS:** We compared read streaming, depth analysis, and pile-ups results of SCTools with samtools and confirmed that our tools return accurate results.

**CONCLUSION:** We envision a real-world scenario that individuals create private blockchains to store their personal genomes to share with their healthcare providers. Our storage protocol allow query of data from personal genomes using blockchains.

## PgmNr 1515: Decoding regulatory structures and features from epigenomics profiles: A Roadmap-ENCODE Variational Auto-Encoder (RE-VAE) model.

### Authors:

R. Hu <sup>1</sup>; G. Pei <sup>1</sup>; P. Jia <sup>1</sup>; Z. Zhao <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, USA; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

---

The development of chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing (ChIP-seq) technologies has promoted generation of large-scale epigenomics data, providing us unprecedented opportunities to explore the landscape of epigenomic profiles at scales across both histone marks and tissue types. In addition to many tools directly for data analysis, advanced computational approaches, such as deep learning, have recently become promising to deeply mine the data structures and identify important regulators from complex functional genomics data. We implemented a neural network framework, a Variational Auto-Encoder (VAE) model, to explore the epigenomic data from the Roadmap Epigenomics Project and the Encyclopedia of DNA Elements (ENCODE) project. Our model is applied to 935 reference samples, covering 28 tissues and 12 histone marks. We used the enhancer and promoter regions as the annotation features and ChIP-seq signal values in these regions as the feature values. Through a parameter sweep process, we identified the suitable hyperparameter values and built a VAE model to represent the epigenomics data and to further explore the biological regulation. The resultant Roadmap-ENCODE VAE (RE-VAE) model contained data compression and feature representation. Using the compressed data in the latent space, we found that the majority of histone marks were well clustered but not for tissues or cell types. Tissue or cell specificity was observed only in some histone marks (e.g., H3K4me1 and H3K27ac) and could be characterized when the number of tissue samples is large (e.g., blood and brain). In blood, the contributive regions and genes identified by RE-VAE model were confirmed by tissue-specificity enrichment analysis with an independent tissue expression panel. Finally, we demonstrated that RE-VAE model could detect cancer cell lines with similar epigenomics profiles. In conclusion, we introduced and implemented a VAE model to represent large-scale epigenomics data. The model could be used to explore classifications of histone modifications and tissue/cell specificity and to classify new data with unknown sources.

## PgmNr 1516: Comparing haplotype assembly approaches.

**Authors:**

A. Johnson <sup>1</sup>; S. Sun <sup>2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Statistics, Texas A&M University, College Station, TX.; 2) Mathematics, Texas State University, San Marcos, TX

---

Haplotype information is important for further understanding genetic diseases. Therefore, it is crucial to obtain haplotypes for disease studies. With the development of next generation sequencing (NGS) technologies, it is now possible to obtain haplotypes using sequencing reads. The process of determining haplotypes based on sequencing reads is called haplotype assembly. It is challenging to conduct haplotype assembly because NGS datasets are very large and have complex genetic and technological features. Even though a large number of approaches or software packages have been developed, it is unclear how well these programs perform. Most of them are not well evaluated as they are only compared with a small number (e.g., 1 or 2) of other methods and validated based on different datasets. In this project, we will conduct a comprehensive analysis to compare currently available haplotype assembly approaches (software packages). We will compare these approaches based on their statistical or computational methods, algorithmic features, and evaluation features as well. We will show our comparison results based on both real data and simulated data. With our comparison results, we shall provide users with both detailed input on the performance of current methods and new perspectives on haplotype assembly, which will be helpful for developing more accurate and efficient algorithms.

## PgmNr 1517: Characterizing deleterious variation using sub-population allele frequency information.

### Authors:

K.J. Karczewski<sup>1,2,3</sup>; J. Karjalainen<sup>1,2,3</sup>; M. McConlogue<sup>1,2</sup>; M. Kurki<sup>1,2,3</sup>; A.R. Martin<sup>1,2,3</sup>; B.M. Neale<sup>1,2</sup>; M.J. Daly<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Institute, Cambridge, Massachusetts.; 2) Massachusetts General Hospital, Boston, Massachusetts.; 3) Institute for Molecular Medicine Finland, Helsinki, Finland.

---

As natural selection purges deleterious variation from the population, the allele frequency of a variant is a useful estimator of its functional effect. However, when a population goes through a bottleneck, a subset of deleterious variants will be boosted in allele frequency. Here, we describe the impact of multiple distinct population bottlenecks on the allele frequency distribution of deleterious variants, using exome sequence data from 125,748 individuals and genome sequence data from 15,708 individuals. In particular, we focus on the deleterious consequences on genetic variants, arising from the strong recent bottlenecks in Finnish and Ashkenazi Jewish individuals, as well as the serial Out-of-Africa population bottlenecks.

Consistent with an expectation that bottlenecks allow some extremely rare variants (with higher rates of deleteriousness) to jump dramatically in frequency, we demonstrate that these enriched variants are more likely to be in functionally-relevant coding categories, especially predicted loss-of-function variation. We thus propose a new metric: that the proportion of variants enriched through a bottleneck represents the deleteriousness of that category of variants. We show this metric correlates with previous metrics of variant deleteriousness as well as the severity of the bottleneck. We apply this metric to assess the deleteriousness of different categories of non-coding variation, identifying promoter regions in particular as those that are under the strongest negative selection.

Finally, we assess the influence of these bottlenecks on results from genome-wide association studies, and find that variants enriched by a bottleneck are more likely to be identified as significant hits than equally frequent but non-enriched variants. Using LD-score regression, we characterize the heritability explained in summary statistics from bottlenecked populations, as a function of allele frequency in cosmopolitan populations. These results indicate the power of bottlenecked populations for identifying deleterious classes of variation as well as discovery of genetic variants associated with disease.

## PgmNr 1518: Investigating the associations between epigenomic and metabolomic data using an integrative multivariate approach.

### Authors:

V. Karhunen<sup>1</sup>; F. Jesus<sup>2</sup>; B. Bodinier<sup>1</sup>; M. Chadeau<sup>1</sup>; S. Sebert<sup>3,4,5</sup>; M. Ala-Korpela<sup>3,4,6,7,8,9</sup>; J. Kettunen<sup>3,4,6,10</sup>; M.-R. Järvelin<sup>1,3,4,11</sup>; M. Evangelou<sup>1,12</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom; 2) Department of Life Sciences, Faculty of Natural Sciences, Imperial College London, London, UK; 3) Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland; 4) Biocenter Oulu, University of Oulu, Oulu, Finland; 5) Department for Genomics of Common Diseases, School of Medicine, Imperial College London, London, UK; 6) Computational Medicine, Faculty of Medicine, University of Oulu, Oulu, Finland; 7) Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 8) Systems Epidemiology, Baker Heart and Diabetes Institute, Melbourne, VIC, Australia; 9) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland; 10) National Institute of Health and Welfare, Helsinki, Finland; 11) Department of Life Sciences, College of Health and Life Sciences, Brunel University London, London, UK; 12) Department of Mathematics, Imperial College London, London, UK

---

Metabolites provide information on important biological processes. To fully understand these processes, it is important to link them to gene activity and expression. Epigenetic regulations via DNA methylation are known to reflect gene expression. Thus, we have aimed to analyse the associations between epigenome-wide DNA methylation measured at CpG sites and nuclear magnetic resonance (NMR) targeted metabolomics.

The most common way of analysing epigenome-wide DNA methylation data has been by univariate analysis, where each CpG site is tested for association with a univariate phenotype independently from the rest. In this study, we have investigated the associations through the angle of two-way multivariate association analysis, where the outcome is a matrix of multiple phenotypes and the predictor is the matrix of DNA methylation at CpG sites.

We utilised an integrative approach for two-way multivariate association analysis to investigate the relationships between DNA methylation and NMR metabolomics (both measured in blood samples) in Northern Finland Birth Cohort 1986 dataset (n=470). We used concentrations of 144 metabolites as the multivariate phenotype, and methylation levels at a subset of 27,688 CpG sites as the predictors. The method used is a canonical correlation analysis (CCA) approach modified from the one proposed by Seoane et al. (2014). The classical CCA finds linear combinations of outcomes and predictors with the highest correlations. The proposed method combines the classical CCA with a genetic algorithm optimization method, which is based on principles of evolution, with steps analogous to inheritance, mutation and crossover (Seoane et al. (2014)). This approach also performs variable selection for both predictors and outcomes.



Using the above method we found associations between VLDL and glucose-related metabolites with DNA methylation sites related to cellular metabolism and signalling genes. We also compare these results to the ones from univariate association analyses. Our results showcase the advantages of a two-way multivariate association analysis as it can aid in discovery of new findings, even with modest sample sizes, for multidimensional datasets.

## PgmNr 1519: GMASS: A novel measure for genome assembly structural similarity.

**Authors:**

D. Kwon; J. Lee; J. Kim

[View Session](#) [Add to Schedule](#)

**Affiliation:** Konkuk University, Seoul, Korea, Republic of

---

Thanks to the recent advancements in next-generation sequencing (NGS) technologies, large amount of genomic data, which are short DNA sequences known as reads, has been accumulating. Diverse assemblers have been developed to generate high quality *de novo* assemblies using the NGS reads, but their output is very different because of algorithmic differences. However, there are not properly structured measures to show the similarity or difference in assemblies.

We developed a new measure, called the GMASS score, for comparing two genome assemblies in terms of their structure. The GMASS score was developed based on the distribution pattern of the number and coverage of similar regions between a pair of assemblies. The new measure was able to show structural similarity between assemblies when evaluated by simulated assembly datasets. The application of the GMASS score to compare assemblies in recently published benchmark datasets showed the divergent performance of current assemblers as well as its ability to compare assemblies. The GMASS will contribute to the understanding of assembly output and developing *de novo* assemblers, and serve as a valuable measure for studying genomes of various species including human.

## PgmNr 1520: AtacWorks: A deep convolutional neural network to enhance ATAC-Seq data.

### Authors:

A. Lal<sup>1</sup>; N. Yakovenko<sup>1</sup>; Z.D. Chiang<sup>2,3</sup>; J.D. Buenrostro<sup>2,3</sup>; J. Israeli<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NVIDIA, Santa Clara, California.; 2) Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA;  
3) Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA

---

ATAC-Seq (Assay for Transposase-Accessible Chromatin using Sequencing) is a DNA sequencing technique used to measure genome-wide chromatin accessibility. Since its development in 2013, ATAC-Seq has been used to identify regulatory DNA elements, to localize epigenetic changes underlying development and disease, and to construct cellular regulatory networks. More recently, single-cell ATAC-Seq has enabled mapping of accessible chromatin in individual cells, leading to the identification of cell-type specific regulatory elements and networks.

The quality of the ATAC-Seq signal depends on the amount of input material for the experiment. ATAC-Seq derived from few cells can be so sparse that open chromatin regions fail to be identified. This leads to errors in characterizing rare cell types in a single-cell ATAC-Seq experiment.

We introduce AtacWorks, a deep convolutional neural network that takes as input a sparse, low coverage ATAC-Seq signal, and reconstructs a higher-quality signal. To predict chromatin accessibility at each position on the genome, our model uses information from a wide surrounding region, combining local features of the ATAC-Seq signal with large-scale features spanning several kilobases. We applied AtacWorks to both bulk ATAC-Seq (where it generates a high-quality signal from low sequencing coverage) and single-cell ATAC-Seq (where it generates a high-quality signal from a small number of cells).

We demonstrate that enhancing noisy ATAC-Seq data using our model increases the accuracy of open chromatin identification for downstream analysis. In an experiment using single-cell ATAC-Seq from 600 cells, our model doubles the number of true open chromatin sites that can be identified from the low-quality signal, from 40.0% of the true sites to 80.7%. We also show that an AtacWorks model trained on data from one cell type can accurately reconstruct data from other cell types. This suggests that our model identifies common patterns in ATAC-Seq data across diverse biological datasets.

## PgmNr 1521: BioTuring Browser: A modern data visualization and analytics platform for single-cell sequencing data.

**Authors:**

T. Le; T. Phan; T. Truong; T. Tran; D. Phung; T. Tran; K. Nguyen; A. Nguyen; N. Pham; H. Pham; S. Pham

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bioinformatics, BioTuring Inc, San Diego, California.

---

Single-cell sequencing technologies provide unprecedented tissue resolution and quickly allow breakthroughs in a wide spectrum of biological research in both academic laboratories and pharmaceutical companies. On the other hand, the amount of information generated in a typical scRNA-Seq study grows exponentially and creates huge computational challenges in data processing, analytics, and visualization in both private and public single-cell data sets. To address these challenges, we introduce BioTuring Browser, a modern data visualization and analytics application to explore scRNA-Seq dataset in our curated database as well as combining with newly generated data. Our database consists of more than 3.4 million cells from 12 different tissues and 56 single-cell studies, all supported with state-of-the-art analytics and visualization packages for single cell transcriptome, and CITE-Seq data. The visualization and computation in BBrowser are optimized so that it can effectively analyze up to 1.3 million cells at a time on a standard laptop, and allow users to quickly export, reanalyze, or remove batch effects and merge data from different studies. BioTuring Browser addresses major computational challenges in single-cell analytics, fosters transparency in science, and helps scientists explore single-cell data more effectively.

# PgmNr 1522: Discovering SNVs and indels from RNA-Seq: Comparison of results of whole transcriptome sequencing to those of whole genome sequencing.

## Authors:

J. Lee <sup>1,2</sup>; Q. Choi <sup>3</sup>; S.J. Kee <sup>2</sup>; S.H. Kim <sup>1,2</sup>; J.H. Shin <sup>2</sup>; M.G. Shin <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Laboratory Medicine, Chonnam National University Hwasun Hospital, Hwasun, Republic of Korea; 2) Department of Laboratory Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea; 3) Department of Laboratory Medicine, Chungnam National University College of Medicine, Daejeon, Republic of Korea

---

**Introduction:** Molecular test for genetic variations has been DNA sequencing, and RNA sequencing (RNA-Seq) has been mainly used for gene expression analysis. Recently, attempts have been made to detect genetic variants from RNA-Seq results, as splice-aware aligners' functionality has greatly improved and RNA-Seq utilization has increased. In this study, we compared the results of whole transcriptome sequencing (WTS) and those of whole genome sequencing (WGS) obtained from the same individual to find out the possibility of RNA-Seq as a tool for discovering genetic variants (SNVs and Indels).

**Methods:** RNA and DNA were extracted from peripheral blood of one healthy adult male. WTS was performed as 200 cycles using HiSeq2500 system (Illumina, USA). We aligned reads to the GRCh38 human genome assembly using STAR v2.7.0f. After sorting and deduplication using Picard v2.20.0, variants were called using freebayes v1.2.0. The variant calling was restricted to the coding DNA sequence (CDS) region of all protein-coding genes. WGS was performed 300 cycles using HiSeq X-10 system (Illumina). We aligned reads from the sequencer using bwa v0.7.17. After sorting and deduplication, variants were called using freebayes. The variant calling was also restricted to the CDS region. After merging vcf files of WTS and WGS using bcftools v1.9, we compared the results between two methods.

**Results:** A total of 9,376 variants (SNVs and Indels) were found from CDS regions of WTS data. On the other hand, a total of 8,591 variants were found in CDS regions of WGS. Among them, 8,248 variants were found in both methods, the concordance rate was 84.9%. Discrepancies between the results of WTS and those of WGS were not randomly distributed throughout the entire genes but were clustered in specific genes, mainly in the *HLA*, *IGH*, *IGK*, *IGL*, and *TRB* subgroups that make up the immune system. When these five subgroups were excluded, 8,140 variants were detected in WTS data and 8,311 variants in WGS data. Among them, 7,943 variants were identical, showing 94.3% as concordance rate.

**Conclusions:** In this study, we confirmed that germline variants on DNA could be reliably predicted from RNA-Seq results using splice-aware aligner and CDS region-specific variant calling. However, special attention should be paid to the prediction of variants in the genes related to the immune system.

## PgmNr 1523: Development of a whole genome sequencing-based assay for detecting copy number variations at a clinical grade level.

### Authors:

W.P. Lee <sup>1,2,3</sup>; Q. Zhu <sup>3</sup>; X. Yang <sup>2,3,4</sup>; S. Liu <sup>3</sup>; E. Cerveira <sup>3</sup>; M. Ryan <sup>3</sup>; A. Mil-Homens <sup>3</sup>; L. Bellfy <sup>3</sup>; C. Zhang <sup>1,3</sup>; C. Lee <sup>1,3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Precision Medicine Center, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China; 2) MOE Key Lab for Intelligent Networks & Networks Security, School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an, China; 3) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 4) Department of Computer Science and Technology, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, China; 5) Department of Life Sciences, Ewha Womans University, Seoul, South Korea

Despite the growing use of next-generation sequencing (NGS)-based assays for the detection of SNPs and INDELs in clinical settings, it remains challenging to reliably detect copy number variants (CNVs) for clinical diagnostics. At the clinic, it lacks robust end-to-end tools or pipelines that will process raw sequencing datasets and detect CNVs. Although several whole genome sequencing (WGS) based CNV calling algorithms have been developed, none of them is widely accepted for pathogenic or high-risk CNV discrimination due to either high false positive or false negative rate. Building pipelines using a combination of existing calling algorithms can gain strength and preclude limitation of each tool. However, replicating those pipelines is not realistic without a full release of them, including versions and parameters. Moreover, using an ensemble of calling algorithms may improve accuracy, but it will add complications and thus take longer to make a diagnostic decision.

As a solution, we developed JAX-CNV, a multi-tiered computational algorithm for the detection of CNVs on WGS data. The performance of JAX-CNV was evaluated on 31 samples (coverage 30-45) from patients with constitutional disorders, which were examined in parallel by the clinically-validated chromosomal microarray analysis (CMA) assay in our CLIA-certified laboratory. As a result, JAX-CNV identified all 104 CNVs reported by the CMA assay including pathogenic CNVs in the 101Kb-94.2Mb size range. The result demonstrated 100% concordance between JAX-CNV and the clinical CMA assay. Moreover, JAX-CNV detected an additional 767 CNVs that were not captured by the CMA. To assess the false discovery rate of JAX-CNV, we randomly selected 16 novel discovered CNVs from two samples and successfully validated 14 by Droplet digital PCR (ddPCR). One deletion was not conclusive and the other one duplication was considered as false positive where it locates in segmental duplication and simple repeat regions. The robustness test of JAX-CNV is performed by downsampling to lower coverage sequencing data (30x, 20x, 15x, 10x, and 9x) and found 100% sensitivity for the detection of CNVs >300Kb (the current cutoff set by CAP for clinical CMAs) at as low as 9x sequencing coverage. This study demonstrates that WGS paired with JAX-CNV has the potential to replace CMAs for clinical diagnostics.

## PgmNr 1524: Quality control for full genome sequencing with Sequila.

### Authors:

A. Lesniewska <sup>1</sup>; S. Kubicki <sup>1</sup>; M. Okoniewski <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Technology, Department of Computer Science, Piotrowo 2, Poznań, Poland; 2) ETH Zurich, Scientific IT Services, Weinbergstrasse 11, Zurich, Switzerland

The number of sequenced full genomes is growing steadily in recent years. It is done mainly in Illumina high-throughput technology and is becoming the basis of precision medicine. Every sequenced genome, sequenced in low coverage is resulting in a large dataset with sizes ranging in terabytes. Thus the analysis techniques once useful eg. for exomes, become too costly computationally.

A partial answer to the growth of the dataset size is the application of novel software, eg. based upon Apache Spark[1] that takes advantage of map-reduce scaling and memory optimization. Recent versions of Sequila[2] are providing fast coverage function, which may be used eg. for copy number analysis or for checking the coverage-based quality of a full genome, and to do this in not prohibitive time.

Calculating the genome-wide coverage function is the first step. Then a set of genomic regions with sub-standard coverage needs to be identified. There are specific thresholds below which variant calling becomes statistically not reliable[3,4]. The set of those sub-standard regions needs to be then compared with gene coordinates and low-mappability regions. It is important also to point out possible reasons for low coverage. The reasons may include mappability, large deletions, copy number issues or technical or algorithmic artefacts of a specific type of sequencing.

Our concept of quality control for full genomes includes the following:

- \* fast calculation of genome-wide read coverage function
- \* finding sub-standard genomic regions with low coverage (LCRs)
- \* finding the overlap of LCRs with low mappability regions
- \* finding the overlap of LCRs with genes and exons
- \* attempting to distinguish between various types of reasons for LCRs

Sequila includes now optimized features for finding coverage and overlapping large sets of genome coordinates, which can be used for the productive solutions of the functionalities above. We present the recent algorithmic approaches with implementation as well as tests on full genomes with different level of coverage. Important issues that still need to be solved technically are eg. scalability or memory tuning. Overall, the results in the poster show how efficiently the measures can be calculated and how to interpret them. This will be the basis for a dedicated software package for the full genome QC.

- [1] Wiewiórka et al. Bioinformatics, 2014
- [2] Wiewiórka et al. Bioinformatics, 2018
- [3] Meienberg et al. Hum Gen, 2016
- [4] Meienberg et al. NAR, 2015



## PgmNr 1525: GPU-accelerated next generation sequencing bioinformatic pipeline.

**Authors:**

M. Linan; Y. Itan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NEW YORK.

---

Current software used for gold-standard next generation sequencing (NGS) bioinformatic pipelines (BWA, PICARD, GATK, Samtools and variant annotation software) are all programmed to utilize conventional central processing units (CPU). Slow computational runtime remains a major hurdle in NGS bioinformatic pipelines despite the ability to multi-thread some of the pipeline's software. Currently running full pipelines for thousands of whole exome or whole genome sequencing (WES and WGS, respectively) samples, requires several months even on high-performance computing CPU clusters. Graphical processing units (GPU) have not yet been efficiently employed in NGS bioinformatic pipeline acceleration but have been demonstrated to accelerate computational runtime in computational chemistry, molecular dynamics and medical imaging. To address this problem, we have developed a Kubernetes GPU-accelerated bioinformatic pipeline where we adapted the code of the different NGS bioinformatic pipeline software to utilize GPU. We tested the applicability and performance of GPU frameworks to accelerate BWA, PICARD, GATK, Samtools and variant annotation software (ANNOVAR, SnpEff and VEP). We show through multiple experiments the extent of acceleration in each software with varying combinations of multithreaded CPUs and GPUs. Overall, we demonstrate that GPU is effective for enhancing the computational speed of NGS bioinformatic pipeline analyses. We also show that with further development the combination of CPU and GPU for these analyses will likely become the gold standard in the field.

## PgmNr 1526: Baited abrogation of contamination.

**Authors:**

F. Mendez; R. Jiang; S. White; W. Lee

[View Session](#) [Add to Schedule](#)

**Affiliation:** Helix Opco LLC, San Carlos, California.

---

Whole exome and whole genome sequencing (WES and WGS, respectively) have proven valuable tools to hasten diagnosis of genetic disorders and enable discovery of heritable traits and disease. These approaches rely on next generation sequencing (NGS) technologies and have been scaling up due in part to the ease of sample collection with saliva, both in the clinic and at home. NGS methods are very sensitive, but also susceptible to the effects of contamination. Saliva-based collection methods require collection protocols (i.e. waiting 30 minutes) that attempt to minimize the impact of exogenous factors such as food on data quality. Nonetheless analytical tools are necessary to combat possible sample contamination that could lead to incorrect genotype calls, which impact ancestry estimates, estimates of relatedness, and clinical diagnosis. A number of statistical methods (like Freemix) enable the efficient detection of human cross-contamination; however, these methods were not designed for cross-species contamination. The effects of cross-species contamination on human genotype calls have not been carefully explored, possibly due to the lack of large saliva based collections that are subsequently sequenced on NGS technologies. Potential sources of contamination in saliva include oral microbiome and ingested food. Bacterial sequences are very divergent from those of humans, but some animal sequences may be mapped and incorrectly used in variant calling.

We have studied the effect of mapping sequences from nine farm and pet animals to the human reference genome and found that a considerable number of human genes in a contaminated human sample could be flagged as having pathogenic mutations. This includes more than a third of the genes included in the ACMG 59 list of genes with actionable mutations. We tested various levels of animal contamination and observed that amounts of foreign DNA that are not detected as contaminating by Freemix can still result in genotyping errors. We implemented a filtering method that greatly alleviates this problem, removing most incorrect calls without significantly affecting the coverage by reads of human origin. We then applied this method to human samples that were collected from saliva and exome sequenced. We show that our method significantly reduces the risk of incorrect genotype calls with potential healthcare consequences, thereby improving consumer confidence in genomic testing results.

## PgmNr 1527: AutoMap: High performance homozygosity mapping using sequencing data.

### Authors:

M. Quinodoz <sup>1,2</sup>; V. Peter <sup>1,2</sup>; N. Bedoni <sup>1</sup>; K. Cisarova <sup>1,2</sup>; B. Royer-Bertrand <sup>3</sup>; A. Salmaninejad <sup>4</sup>; N. Sepahi <sup>6</sup>; A. Pasdar <sup>4,5</sup>; M. Mojarrad <sup>4</sup>; A. Ghanbari Asad <sup>6</sup>; L. Coutinho-Santos <sup>7</sup>; A. Superti-Furga <sup>3</sup>; C. Rivolta <sup>1,2,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Département de Biologie Computationnelle, Université de Lausanne, Lausanne, Vaud, Switzerland; 2) Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom; 3) Division of Genetic Medicine, Lausanne University Hospital (CHUV), University of Lausanne, Lausanne, Switzerland; 4) Department of Medical Genetics, Faculty of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran; 5) Division of Applied Medicine, Medical School, University of Aberdeen, Foresterhill, Aberdeen, UK; 6) Noncommunicable Diseases Research Center, Fasa University Of Sciences, Fasa, Iran; 7) Instituto de Oftalmologia Dr Gama Pinto, Lisbon, Portugal; 8) Institute of Molecular and Clinical Ophthalmology Basel (IOB), Basel, Switzerland; 9) University of Basel, Basel, Switzerland

**Purpose:** Through the identification of genomic runs of homozygosity (ROHs), homozygosity (autozygosity) is a powerful approach to find autosomal recessive mutations, especially in consanguineous families, via Most tools for ROH discovery rely on data from genotyping arrays, querying polymorphic variants at fixed positions. However, current investigation in medical genetics are mostly based on exome sequencing (ES), producing genotypes at variable locations of the coding genome with different allele frequencies. For this reason, existing tools were adapted or created for the use of ES data, but with limited performances. We have developed AutoMap (Autozygosity Mapper) to allow investigators to use directly ES data for the identification of ROH, without the need of performing additional genotyping tests.

**Methods:** AutoMap directly uses standard VCF files produced by various variant callers as primary source of data. A careful variant filtering step relying on coverage and alternative reads count is implemented, prior to the identification of ROHs, by a sliding-window approach and subsequent filtering. The tool was first trained on ES data from 26 samples from consanguineous families, and further validated in 26 additional samples with same ethnic background and degree of consanguinity. The performances of AutoMap were compared to other existing tools by using PLINK on genotyping array data as a reference.

**Results:** AutoMap displayed similar performance for both training and validation cohorts, with sensitivity of 90.3% and 91.6% and specificity of 81.1% and 83.2% respectively. Our method showed a significantly higher F-score with respect to 8 existing tools. When applied to unpublished ES data, AutoMap allowed the discovery of homozygous non-coding variants that in turn led to the identification of disease genes for two novel conditions including the Liberfarb syndrome.

**Conclusion:** AutoMap is a reliable tool that can predict, in less than a minute, ROHs with extremely high specificity and sensitivity. It can run efficiently by using a single VCF file derived from individual ES experiments. Our tool is available both via a web-based interface, for a quick analysis with default parameters, as well as a command-line package, allowing assessment of a large numbers of samples and the customization of many parameters. Finally, AutoMap can be used without adjustments also

on VCFs from whole-genome sequencing (WGS) experiments.

## PgmNr 1528: Constraint and phenotypic similarity characterization of drug targets and their paralogs to inform on drug side effect burden.

### Authors:

D. Rhodes; P. Cacheiro; D. Smedley; M.R. Barnes; M. Caulfield

[View Session](#) [Add to Schedule](#)

**Affiliation:** William Harvey Research Institute, Queen Mary University of London, London, London, United Kingdom

---

Side effects are a major concern in drug development due to their contribution to clinical trial failures or reducing patient compliance. A possible source of side effects include non-specific binding to paralogs - homologous genes resulting from prior gene duplication events. Here we describe efforts to systematically identify whether an association between the number of paralogs of a target protein and the number of side effects exists. There are multiple challenges associated to this analysis, among them: how to properly define relevant paralog groups and how to identify relevant side effects amongst the redundant or infrequent side effects reported.

First we curated 2,764 high confidence sets of paralogs (components) covering 11,882 unique proteins based on Ensembl 95 annotations and subsequent filtering using pairwise gene ontology semantic similarity, amino acid sequence similarity, and finally network based community detection algorithms. 519 proteins targeted by FDA approved drugs with known mechanism of action were labelled across 172 components. Drug targets were amongst the most constrained genes of the component as measured by their loss-of-function observed/expected upper bound fraction (LOEUF) score; with 62% of drug targets being more constrained than the mean constraint of their component. Components containing drug targets are on average double the size (mean 8.55 vs 4.5) and showed greater within-component phenotypic similarity (mean score 71.3 vs 64.8) than non-drug target containing components. Phenotypic similarity scores were computed using mouse orthologue phenotype data from the Mouse Genome Informatics resource.

We expanded our list of approved drug targets to include drugs with strong evidence of binding for downstream analysis. We then summarised SEs (from the Side Effect Resource (SIDER 4.1)) for drug-target containing components to per component measures of SEs using the intersection of SEs reported for any protein targeted by more than one drug. Initial findings suggest a positive correlations between side effect number and severity with component size. Further investigation will aim to further curate the component SE burden using SE frequency data.

A better understanding of the relationship between a drug target and its paralog group, will provide additional information on the appropriateness of new candidate targets and consequently may help prioritise targets with favourable SE profiles.

## **PgmNr 1529: Novel methods for structural variant detection using Linked-Reads: Analysis of 68 human whole genomes across 11 HapMap populations.**

### **Authors:**

C. Ricketts <sup>1</sup>; D. Meleshko <sup>1</sup>; C. Alkan <sup>2</sup>; I. Hajirasouliha <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute of Computational Biomedicine, Weill Cornell Medicine of Cornell University, New York, NY 10021; 2) Department of Computer Engineering, Bilkent University, Ankara, Turkey

---

Linked-Read technologies such as the 10x Chromium system promise a comprehensive structural variant (SV) screening of a large cohorts at a reasonable cost. In this work, we present two computational methods for SV calling using Linked-Read sequencing: (i) Novel-X, a method for detection and assembly of novel sequence insertions in a sequenced genome which is based on a local assembly of interconnecting barcodes, and (ii) VALOR2, a mapping based algorithm for characterizing complex large-scale SVs. Using these methods, we performed a search for large duplications, inverted duplications, insertions, inversions and novel sequence insertions of 68 whole human genome samples sequenced using 10X technology. This cohort consists of samples from different populations from the HapMap project that allowed us to study interracial dependencies of the obtained calls. For the novel sequence insertion we were able to find that number of insertions in Asian samples is higher than in the samples of African or European origins. This finding is consistent with Sherman et al. 2019. Moreover, we were able to find 16 insertions presented in  $\geq 50$  samples, some of them exceeding 1000 bp in length that proves that current human reference genome is far from being complete. In total we identified 676 inversions, 60 duplications, 13 inverted duplications, and 12 Mbp of novel insertion sequences across all samples. Interestingly, African genomes also represented the highest total number of inversions, duplications and inverted duplications among populations of similar sample sizes. Our work also shows a higher degree of conserved variants in Yoruba/Nigerian and USA/African genomes as no other population has more than 10 variants shared across  $>50\%$  of the genomes ( $N \geq 5$ ). This is consistent with Gurdasani et al. 2015, showing low differentiation within African populations.

## PgmNr 1530: Machine learning approach to literature mining for the genetics of complex diseases.

### Authors:

J. Schuster <sup>1,2</sup>; M. Superdock <sup>2</sup>; A. Agudelo <sup>1</sup>; P. Stey <sup>3,4</sup>; J. Padbury <sup>1,2,5</sup>; I.N. Sarkar <sup>2,4</sup>; A. Uzun <sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Women and Infants Hospital of Rhode Island, Providence, Rhode Island.; 2) Pediatrics, Warren Alpert Medical School of Brown University, Providence, Rhode Island; 3) Computing and Information Services, Brown University, Providence, Rhode Island; 4) Center for Biomedical Informatics, Brown University, Providence, Rhode Island; 5) Center for Computational Molecular Biology, Brown University, Providence, Rhode Island

---

To better understand the genetic mechanisms of complex disease, we hypothesized that it was necessary to identify relevant, phenotype-specific gene sets that combined the curation of public literature, review of experimental databases, and interpolation of pathway-associated genes for a given disease or phenotype. Using this strategy, we built publicly available databases for two complex reproductive disorders: The Database for Preterm birth (dbPTB) and The Database for Preeclampsia (dbPEC). The completeness and accuracy of these genetic databases is essential for our understanding of these complex disease phenotypes. Unfortunately, given the exponential increase in biomedical literature, it is becoming increasingly difficult to manually maintain these databases. We used these two previously developed, highly-curated databases of literature describing the genetic architecture for complex diseases as reference data sets to develop a machine learning-based approach to optimize article selection for manual curation. Our approach employed automated text-mining tools to extract rich features from titles and abstracts, supervised machine learning to predict article relevance, and human expertise to extract relevant genetic data. We used logistic regression, random forests, and neural networks to classify articles that were relevant to the diseases of interest that should be considered for further manual analysis. We examined features that we hypothesized would best classify articles with genetically relevant content. Various combinations of these features were used to classify articles and the performance of these feature sets were compared to the standard Bag of Words feature set. Several of the feature sets, which include the MESH terms, gene significance triplets (sentences in which there was a co-occurrence between the "GENE" semantic group and UMLS concepts related to statistical significance) and semtype features, outperformed Bag of Words in terms of workload savings (% reduction of required manual curation), F-score, and Fg-score at 95% gene sensitivity. Workload savings ranged from 0.814-0.846 for the dbPTB data set and 0.301-0.371 for the dbPEC data set. A rich database of metadata and annotations for each article is generated which allows for rapid query of individual features. Our results demonstrate that machine learning algorithms can identify articles of interest for creation of a database of gene sets for complex diseases.

## **PgmNr 1531: Truwl.com: A web platform to publish bioinformatics methods in a reusable, findable, understandable fashion.**

**Authors:**

K.B. Seby; T. Thompson

[View Session](#) [Add to Schedule](#)

**Affiliation:** xD Bio, Inc., Whitefish, Montana.

---

The lack of solutions to find, understand, and apply established bioinformatics methods has led to a lack of reproducibility, underuse of open source tools, continued dependence on outdated tools, and time-consuming reimplementations of established methods. Bioinformatics tools and workflows published on truwl.com (truwl) are exposed to crawlers through server-side rendering and provide context through tags, commenting, and voting making methods easily findable through search engine queries. The proper use of tools and workflows can be understood through collections of use cases that show examples of how tools and workflows have been used, including their input and output files and parameter settings. The ability to easily explore the relationships between tools, workflows, and their inputs and outputs is key to understanding how methods are applied and provides the necessary context to vet and adapt methods. Truwl makes exploring these relationships simple through a web interface: file inputs link to pages that describe them and link to any known locations where the files can be obtained; workflows are displayed as interactive graphs that can be navigated to the tools that they use and are similarly linked to sets of use cases that enable users to see how the workflows are used in practice. Behind the web interface, tools and workflows displayed on truwl are described by files written in the Common Workflow Language (CWL) and Workflow Description Language (WDL) that are standards to make methods executable by a variety of workflow management systems. Furthermore globally unique identifiers are assigned to each version of tool, workflow, and associated use cases allowing unambiguous citation not only to the method but to how the method was used. The combination of a user friendly web interface, focus on findability, platform agnostic executable descriptions, and community curation tools promises to allow users to share, find, and implement the most appropriate and up to date bioinformatics methods.



## PgmNr 1532: Automated reporting of germline variants using machine learning approaches.

### Authors:

Y. Sivasamy; N. Agrawal; M.S. Niveditha; S. Katragadda

[View Session](#) [Add to Schedule](#)

**Affiliation:** Strand life sciences, Bangalore, India

---

Usage of NGS technologies has become an integral part of clinical settings for genetic screening, diagnostics, and clinical assessment. For better diagnosis and treatment of a disease like cancer, knowledge, and understanding of genomic variants is the key factor. In order to classify a variant as reportable or not, a user has to depend on multiple sources of information and form complex rules to reach a conclusion. Here, we used machine learning algorithms to combine information from the public data sources and our proprietary database of 10000 patients to classify the variant. This reduces the processing time (to about less than 5 min) and the workload of the clinician to validate the variants and generate a report for the patient.

We used Hereditary Breast and Ovarian Cancer germline panel and subsetting our database for variants of genes in this panel to create a classifier model of report generator using supervised learning algorithms. Initial data cleanup was done to exclude the outliers, missing results, and some exceptional cases. Finally, we arrived at a list of 3746 variants from 12 genes of the selected gene list and used it for classifier modeling. These variants were annotated with attributes like Gene name, Mutation type, Strand's ACMG label, Clinvar ACMG labels, Minor allele frequency from ExAC, EVS and 1000 Genomes, dbNSFP predictor scores, and Conservation scores.

We used the WEKA library's Logistic Regression, Naive Bayes, Random Forest, Random Tree, Support Vector Machine with various kernels to choose the algorithm that best suited our purpose of classifying the variant as reportable (yes) or not (no). From the input variants dataset, 80% of the variants were selected randomly for training purpose and the remaining 20% were used for testing. Based on the initial analysis, SVM with RBF kernel suited well for this one-vs-one classification and gave consistent results with an overall accuracy of 98.91%, precision, and recall of 0.989. For validation, about 81 variants that were not taken in for model creation were used. From these variants, 90% were predicted correctly (53 variants as yes and 20 as no). The 10% (8 variants) incorrectly predicted variants were all false positives and these can be quickly resolved by the variant scientist. From the above results, we were able to demonstrate our capability of modeling report generator and this generic approach can be applied to any gene(s) and variant(s) annotated with the required attributes.

## **PgmNr 1533: Machine learning approach to predicting important interacting proteins between human red blood cells and *Plasmodium falciparum* schizont.**

### **Authors:**

J. Soyemi <sup>1,3</sup>; E.F. Adebisi <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Computer Science, Federal Polytechnic, Ilaro, Ilaro, Ogun, Nigeria; 2) Department of Computer and Information Sciences, Covenant University, Ota, Nigeria; 3) Covenant University Bioinformatics Research (CuBRE), Covenant University, Ogun State, Nigeria

---

**Introduction:** The need to develop vaccines that targets the parasite erythrocytic invasion stage becomes imperative because the estimated reduction in malaria in Africa has not been commensurate with the much effort put in place by the world health organization. Identification of important interacting proteins between host and parasite at those crucial stages precedes vaccine development. Previous studies have predicted several interactions between host and parasite, however, the mechanism related to invasion is poorly understood. Again, invasion has been identified as a potential vaccine target, being one of the few stages when the parasite is directly exposed to the host immune system. **Aim:** Therefore, this study identified important interacting proteins (IIPs) that are responsible for merozoite invasion of the human Red Blood Cells (RBCs). **Methods:** RNA-Seq gene expression profiles of human RBCs and merozoite were obtained from experimental data with the differentially expressed genes obtained after statistically analysis in R. Correlation distance metric, K-nearest neighbour algorithm and Bipartite graph were implemented to make the final predictions. **Results and Future Direction:** The predicted results shows 63% IIPs that corroborate with predictions from previous studies and 37% novels. The IIPs predicted should be experimentally validated and then use as drug targets for vaccine development against the parasite invasion of the red blood cells

## PgmNr 1534: A pipeline for large-scale simulation of population-based linkage.

### Authors:

E.L. Sticca <sup>1</sup>; G.M. Belbin <sup>2</sup>; J. Soto <sup>3</sup>; R. Shemirani <sup>4</sup>; D. Nelson <sup>5</sup>; S. Gravel <sup>5</sup>; J.L. Ambite <sup>4</sup>; G.L. Wojcik <sup>3</sup>; E.E. Kenny <sup>2</sup>; C.R. Gignoux <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) HMGPP, University of Colorado Anschutz Medical Campus, Aurora, CO.; 2) Icahn School of Medicine, Mount Sinai; 3) Stanford University; 4) Information Sciences Institute, University of Southern California; 5) McGill University

---

The advent of large-scale sequencing projects such as TOPMed and GSP allow for interrogation of rare, population specific variants. To test the limits of detection of modern genotyping arrays' power to discover rare variant associations compared to whole exome or whole genome sequencing experiments, we developed a pipeline simulating human evolutionary history that generates variants recapitulating global patterns of population structure and identity by descent (IBD). IBD, by detecting patterns of recent relatedness, is often proposed as an alternative strategy for interrogating recent variants that remain rare in the population when only array data is present. This pipeline generates realistic distributions of IBD segments by including evolutionary parameters simulating cryptic relatedness, population substructure and demography, along with a realistic array design model using a similar tag SNP selection strategy to the MEGA array. The pipeline may be run over a variety of parameters by investigations to determine how to optimize their study design based upon demographic history of the populations of interest, underlying genetic architecture of the trait of interest and recruitment sample sizes. Using a mix of python for user-friendliness and C++ for computational efficiency, the pipeline is containerized using Docker to allow for rapid deployment across disparate computational infrastructures. Using this pipeline, we describe the relationship between IBD segments and variants across the rare spectrum, and provide recommendations for investigators interested in rare variant detection and trait association studies. We plan to use this IBD generation pipeline to evaluate novel methods of scalable IBD tract detection (iLASH) and segment clustering (iCURL) to improve power of association testing over traditional GWAS methods because long IBD segments account for recent, rare variation and the pervasive cryptic relatedness present in modern biobank-sized datasets.

## PgmNr 1535: High-level optimizations over query engines ensemble: Accelerating distributed genomic data science.

**Authors:**

A. Szmurlo; M. Wiewiórka; T. Gambin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Computer Science Institute, Warsaw University of Technology, Warszawa, mazowieckie, Poland

---

**Background:** Nucleic acid sequencing is well adopted in performing molecular diagnosis and is the very core of large-scale research projects producing immense amount of data which are meaningless until further examined by algorithms. Presently computational analysis step is the major bottleneck in clinics as well as sequencing research projects however there is a handful of ideas for improving performance of bioinformatics pipelines. Although one of the possibilities, distributed computing is reasonable, available distributed software suffers from either limited functionality, or rather poor overall performance as it is usually being optimized for one of the two kinds of analyses (single sample analysis for clinical purposes vs. cohort/case control research studies).

**Methods:** In recognition of various access patterns instead of using single query engine we plan to construct and use an ensemble of data stores each optimized for specific conditions. We designed efficient primary model exposing different views of the data and helper structures with synchronization mechanisms. Our high-level optimizer, aware of these elements, is able to route the query to the suitable component which will most certainly be superior under encountered conditions.

**Benchmarking:** We have performed tests on 1000 Genomes Project data on our Hadoop-based cluster. We have confirmed that both efficient implementation of appropriate distributed algorithms and routing query to suitable query engine significantly reduces the computation time for painful bioinformatics operations.

**Conclusions:** We expect that optimized access to data will streamline tertiary analyses, resolve the challenge of manipulating heterogeneous genomic data sets and will lay down foundation to wider adoption of data-heavy machine learning methods bringing a new stimuli into the personalized medicine.

## PgmNr 1536: *TypeTE*: A tool for genotyping mobile element insertions in whole-genome sequencing data.

### Authors:

J. Thomas <sup>1</sup>; C. Goubert <sup>2</sup>; L.M. Payer <sup>3</sup>; J.M. Kidd <sup>4</sup>; J. Feusier <sup>1</sup>; W.S. Watkins <sup>1</sup>; K.H. Burns <sup>3</sup>; C. Feschotte <sup>2</sup>; L.B. Jorde <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Utah, School of Medicine, Salt Lake City, Utah.; 2) Department of Molecular Biology and Genetics, 107 Biotechnology Building, Cornell University, Ithaca, New York 14853, USA; 3) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 4) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109

---

The proliferative activity of non-LTR retrotransposons (*Alu*, *LINE-1* and *SVA*) over the last 100 million years has generated thousands of copies, accounting for ~32% of the human genome. Many of those insertions are polymorphic, where individuals or populations share the presence or absence of a mobile element insertion (MEI). Such polymorphic MEIs, in addition to contributing to structural variation, can alter gene expression. Polymorphic MEIs are also used as markers of ancestry in population genetics to understand the structure and evolution of populations. These studies depend on accurate genotyping of MEIs for correct estimation of their allele frequencies, and for phasing these insertions with other variants that comprise the haplotype. In fact, we found that ~17% of the MEI genotype calls from 1000 genome project data (1KGP) are erroneous when compared to a validated set of genotypes (>200) generated by polymerase chain reaction (PCR). Accurate MEI genotyping using short-read alignments is inherently challenging due their repetitiveness, and dedicated tools are scarce. We thus have developed a computational pipeline, *TypeTE*, that aims to genotype both reference and non-reference MEIs (specifically *Alu* insertions) from whole-genome paired-end resequencing data aligned to the reference assembly. The pipeline performs an in-depth analysis of the *Alu* breakpoints detected by other tools. First, all mapped reads including discordant reads and mates are extracted from each locus. The consensus sequence of *Alu* at each locus is either generated through a local *de novo* assembly or extracted from the Repbase database. The pipeline then identifies the signatures of non-LTR retrotransposition (i.e. poly-A tail and target site duplications) and the strand orientation of the *Alu* insertion. The extracted reads are then mapped to two alleles (with and without the *Alu* insertion) separately, and the genotype likelihoods are calculated using established methods. To confirm the accuracy of prediction, we compared *TypeTE* predicted genotypes for both reference and non-reference loci to >200 PCR-based genotypes, which is considered the 'gold standard'. Altogether, *TypeTE* improved genotype accuracy from 83% to 92% in the 1KGP dataset. *TypeTE* is adaptable to other retrotransposon families and is a valuable addition to the toolbox for population genetics and genomics.

## PgmNr 1537: Skipping: A versatile genome assembler for read clouds technology.

### Authors:

T. Tran <sup>1</sup>; H. Che <sup>1</sup>; T. Phan <sup>1</sup>; T. Truong <sup>1</sup>; T. Chen <sup>2</sup>; S. Pham <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) BioTuring Inc, San Diego, CA 92121; 2) UST Corporation, Carlsbad, CA 92011

---

The high-throughput barcoding technology, which tags reads that originate from a single DNA molecule with a unique barcode, has been the foundation for generating Synthetic Long Reads (SLR) and linked-reads. The accuracy of short-read data combined with the long-range information encoded in the read clouds formed by linked-reads make protocols like 10X GemCode and TELL-Seq the preferred solutions for haplotype genome phasing and structural variant calling. Linked-read technologies also bring new hope to completely solve the long-standing genome and metagenome assembly problems. Existing linked-read assemblers use linked-read data to scaffold contigs generated from traditional assemblers and significantly improve the assembly. However, this strategy does not utilize the assembly graph and limits its ability in resolving complex repeat structures. We have developed *Skipping*, a versatile genome assembler for read clouds (linked-read) data, to support *de novo* assembly for both small and large genomes, as well as complex metagenomes. *Skipping* assembler constructs the de Bruijn graph from raw sequencing data then leverages the barcode information to resolve complex regions in the graph caused by repeats in the genome. To increase the continuity of the draft assembly, we use k-mer mapping approach to link pairs of consecutive contigs that share a large number of barcodes. For *E. coli* DH10B dataset, *Skipping* assembler outputs one complete chromosome with highly contiguous quality (NA50: 4.5 Mb). In a mock metagenome dataset, *Skipping* assembler's result has the largest contig and the largest N50 compared to other SLR assemblers' results. We also use data from the Vertebrate Genomes Project to benchmark *Skipping* with other SLR assemblers to see the advantage of *Skipping* in terms of accuracy and speed for medium and large genomes. Although *Skipping* is originally an end-to-end assembler, it can also be incorporated into any existing assembly pipeline as a scaffolding module using linked-read data, which helps improve the quality of draft assemblies.

## PgmNr 1538: Proteinarium: Multi-sample protein-protein interaction analysis and visualization tool.

### Authors:

A. Uzun<sup>1,2,3</sup>; D. Armanious<sup>4</sup>; J. Schuster<sup>1,2</sup>; S. Istrail<sup>3,4</sup>; J. Padbury<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Alpert Medical School, Brown University, Providence, Rhode Island; 2) Pediatrics, Women and Infants Hospital, Providence, Rhode Island.; 3) Center for Computational Molecular Biology, Brown University, Providence, Rhode Island.; 4) Computer Science, Brown University, Providence, Rhode Island.

---

The generation of high throughput screening methods accelerated and improved building protein-protein interaction (PPI) networks. Large scale protein interaction maps show that genes involved in related phenotypes frequently interact physically at the level of proteins in model organisms as well as in humans. This is also true for complex diseases phenotypes. Proteins associated with similar disease phenotypes have a strong affinity to interact with each other. Therefore, identifying disease specific biological networks and subnetworks dynamically through the analysis of experimental data is a valuable application in order to understand the hidden layers of relationships between complex diseases and their associated phenotypes in individuals or patient groups. Hence, new analysis and visualization tools are needed to better identify the shared and individual network pathways within complex diseases or phenotypes. For that purpose, we developed Proteinarium to analyze and visualize the networks of complex diseases and their associated phenotypes in the context of clustered samples based on their network similarities. The feature of this program is that the user can input separate gene lists for each sample as a distinct input which are further examined through protein-protein interaction analyses. The input can be derived from transcriptome analysis or any high-throughput screening approach. For each sample; Proteinarium (1) converts the list of genes into proteins, (2) maps proteins onto the interactome where STRING database, v11 for humans is used, (3) applies Dijkstra's Algorithm to generate PPI networks. Then, Jaccard distance is used to identify the network similarities between samples. Next, samples are clustered according to their network similarities using UPGMA method and displayed as a dendrogram. Each branch of the dendrogram can be visualized to present layered graph of PPI networks. Proteinarium provides several outputs. For example, users can also analyze the network(s) on other platforms like Cytoscape or dendrograms can be visualized on other programs. Proteinarium, through the derivation of shared PPI networks, allows researchers to address important questions on high throughput data for a variety of disease phenotypes based on their PPI networks. Proteinarium is freely available at <https://github.com/Armanious/Proteinarium>.

## PgmNr 1539: An automated functional annotation pipeline rapidly prioritizes clinically relevant genes for neurodevelopmental disorders which may impact treatment.

### Authors:

O.J. Veatch <sup>1</sup>; M.G. Butler <sup>2</sup>; S.H. Elsea <sup>3</sup>; B.A. Malow <sup>4</sup>; J.S. Sutcliffe <sup>5</sup>; J.H. Moore <sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Sleep and Circadian Neurobiology, University of Pennsylvania-Perelman School of Medicine, Philadelphia, Pennsylvania.; 2) University of Kansas Medical Center, Kansas City, Kansas; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 4) Sleep Disorders Division, Department of Neurology, Vanderbilt University Medical Center, Nashville, TN; 5) Vanderbilt Genetics Institute, Department of Molecular Physiology & Biophysics, Department of Psychiatry, Vanderbilt University Medical Center, Nashville, TN; 6) Department of Biostatistics, Epidemiology, & Informatics, University of Pennsylvania, Philadelphia, Pennsylvania

---

There are hundreds of genetic risk variants for neurodevelopmental disorders (NDDs) and the list is continually growing. To understand how genetic data can be used to inform treatment, it is important to develop efficient ways to sort through association study and other related results and prioritize clinically relevant genes. We developed an automated translational bioinformatics pipeline using databases that incorporate evidence from multiple sources to identify: 1) genes associated with NDDs that are expressed in the brain with evidence for convergent biological function, 2) genes with evidence for functional consequences relevant to NDDs in mice, 3) genes with evidence of pathogenic genetic variants and/or predict potential genes via evidence for direct interactions with the protein products of known pathogenic genes, and 4) genes encoding proteins targeted by approved pharmaceutical compounds with evidence for pharmacogenomics effects influencing individual treatment response. We identified 916 unique NDD candidate genes which were annotated and compared with 1,000 random sets of  $n=916$  human protein coding genes defined in Ensembl v86. Compared to all of the random gene sets, more NDD candidate genes ( $p < 1.0 \times 10^{-3}$ ) were annotated in functional categories of interest. Specifically, more NDD genes were associated with phenotypes reflecting nervous system and/or behavior dysfunctions when knocked-out in mice (ProportionNDDSet=0.11,  $\mu$ ProportionRandomSets=0.08 $\pm$ 0.01). Proteins encoded by NDD genes showed more evidence for direct interactions with proteins encoded by American College of Medical Genetics-recommended clinically actionable genes (ProportionNDDSet=0.51,  $\mu$ ProportionRandomSets=0.33 $\pm$ 0.01) and were targets for FDA-approved drugs (ProportionNDDSet=0.14,  $\mu$ ProportionRandomSets=0.03 $\pm$ 0.01) or bioactive molecules with drug-like properties (ProportionNDDSet=0.15,  $\mu$ ProportionRandomSets=0.08 $\pm$ 0.01). In total, 22 NDD genes were prioritized based on functional evidence in all annotation categories, compared to an average of  $7 \pm 3$  genes from random sets. Our automated approach for functional annotation of NDD candidate genes may help to rapidly prioritize potentially actionable results from genetic studies, which may inform future work focused on supporting clinical decisions regarding the benefits of genetic testing for optimizing personalized approaches to treatment on a case-by-case basis.



## PgmNr 1540: Using long and linked reads to generate a new Genome in a Bottle small variant benchmark.

### Authors:

J. Wagner<sup>1</sup>; A. Carroll<sup>6</sup>; I.T. Fiddes<sup>3</sup>; A.M. Wenger<sup>2</sup>; W.J. Rowell<sup>2</sup>; N. Olson<sup>1</sup>; L. Harris<sup>1</sup>; J. McDaniel<sup>1</sup>; C. Xiao<sup>5</sup>; M. Salit<sup>4</sup>; J. Zook<sup>1</sup>; Genome in a Bottle Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899; 2) Pacific Biosciences, 1305 O'Brien Drive, Menlo Park CA 94025; 3) 10x Genomics, 7068 Koll Center Parkway, Pleasanton CA 94566; 4) Joint Initiative for Metrology in Biology, Stanford, CA 94305; 5) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 45 Center Drive, Bethesda, MD 20894; 6) Google, Inc. Mountain View, CA

---

The Genome in a Bottle (GIAB) consortium performs authoritative characterization of 7 human genomes. Here, we describe work towards generating a new GIAB small variant benchmark that incorporates long and linked read sequencing data. The GIAB collection includes 2 son-mother-father trios from the Personal Genome Project that are broadly-consented, including for commercial redistribution. The GIAB benchmarks are created by integrating variant calls from multiple sequencing technologies and analysis methods. This integration systematically evaluates and arbitrates amongst the technologies and methods to identify consensus calls, and regions containing those calls, that can be relied upon as benchmarks. The benchmark variants and genomic regions cover, for example, 87.8% of assembled bases in chromosomes 1-22 of GRCh37 for one genome (HG002). Because many clinically-relevant variants, such as those in *CYP21A2* lie outside the current GIAB benchmark regions, expanding these is important for medical applications. Short read variant callers perform poorly in segmental duplications and low-complexity repeat-rich regions as well as other regions with high homology. We utilize PacBio CCS long read data and 10x Genomics linked reads to expand the GIAB benchmark regions and reduce errors in current regions. Preliminary analyses suggest that long and linked reads might be able to add between 40,601 and 457,130 benchmark small variants along with expanding the coverage of GRCh37 by approximately 20 to 214 million base pairs. The draft benchmark covers substantially more challenging regions, such that the false negative rate for short read-based methods increases by a factor of about 7.6 to 17.2 relative to the current benchmark. Furthermore, initial inspections show a few thousand potential errors in the current benchmark in difficult-to-map regions, such as LINEs, where the current short read-based methods are inaccurate. We generated draft benchmark variant calls, worked with GIAB consortium members for evaluation, and are currently developing more robust calls in segmental duplications. We are developing a similar benchmark set for GRCh38 because it represents segmental duplications better than GRCh37. To confirm the accuracy of the draft benchmark, we performed long range PCR and Sanger sequencing to test variants including those in *CYP21A2* for which we confirmed 12 variants. This work will enable benchmarking in challenging genomic regions.

## PgmNr 1541: Physician-driven genomic analysis using IOBIO web tools.

### Authors:

A. Ward <sup>1,2</sup>; T. Di Sera <sup>1</sup>; M. Velinder <sup>1</sup>; A. Ekawade <sup>1</sup>; G. Marth <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Utah, Salt Lake City, MA; 2) Frameshift Genomics, Boston, MA

The diagnosis of inherited diseases, cancers and other genetic disorders are already benefiting from DNA sequencing and analysis. Given the difficulty of interpreting the genetic basis of these disorders, it is becoming increasingly important for subspecialty clinicians to carry out analysis in their own patients' genomic data. Existing genomic analysis software tools, however, require substantial bioinformatician involvement. We are developing a tool to guide physicians through all necessary steps of clinical diagnostic genomic analysis, from data quality review, candidate gene generation, variant interrogation and prioritization, to reporting of all findings, in a visual, compelling fashion. This clin.iobio tool is based on our popular IOBIO web apps, and performs analyses in real time, without requiring the help of bioinformaticians.

**Clin.iobio** provides a flexible analysis workflow, combining multiple IOBIO apps into a single platform. Following the lead of other medical procedures, a checklist of tasks promotes analysis consistency and repeatability. A user begins by building a candidate gene list using **genepanel.iobio** (<http://genepanel.iobio.io>). This app combines NCBI's genetic testing registry, the Phenolyzer tool, and the user's knowledge of likely genes to generate this gene list. These genes are then propagated to **gene.iobio** (<http://gene.iobio.io>) for real-time variant annotation and analysis. Variants are automatically prioritized by gene.iobio, and users are encouraged to assign a significance level and attach notes to variants to promote collaboration. Finally, variants drawn from the whole exome or genome, independent of the candidate gene list, and conforming to predetermined filters are presented for review. This ensures all potentially disease-causing variants in all genes are examined. The analysis findings are summarized in a final, interactive report. An important feature of **clin.iobio**, is that it facilitates dynamic analysis and re-analysis. If no candidate variants are discovered, the user can quickly create a more expansive gene list, bringing in additional variants for review. This dynamic and collaborative analysis platform will help move genomic analysis into the clinical practice of subspecialty clinicians.

## PgmNr 1542: Fast estimation of genome-wide and partitioned genetic correlation for complex traits.

### Authors:

Y. Wu<sup>1</sup>; D. Palmer<sup>3,4</sup>; A. Ganna<sup>3,4,5,6</sup>; S. Sankararaman<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computer Science, UCLA, Los Angeles, California.; 2) Department of Human Genetics, UCLA, Los Angeles, California; 3) Analytic and Translational Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 6) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

---

Genetic correlation can provide insights into shared genetic pathways as well as providing a starting point to investigate causal relationships among traits. Further, partitioning of genetic correlation provides additional insights into identifying groups of genetic variants that contribute disproportionately to the genome-wide genetic correlation. While several methods for estimating genome-wide genetic correlation have been proposed, no existing method can partition genetic correlation on large-scale individual data. We propose a method based on bivariate linear Mixed models (LMMs) that can estimate genome-wide genetic correlation as well as partition genetic correlation.

The bivariate LMM jointly models the effect sizes of a given SNP on each of the pair of traits being analyzed. The parameters of the bivariate LMM, i.e., the variance components, are related to the heritability of each trait as well their genetic correlation. We further extend the bivariate LMM with additional information on grouping of SNPs. Specifically, we assume that the effect sizes of SNP from the same group for two traits are independent and follow a multivariate distribution with group-specific parameters. However, inference in bivariate LMMs, typically achieved by maximizing the likelihood, poses serious computational challenges that make it impractical to apply at scale

We propose, RGcor-mc, a scalable randomized Method-of-Moments (MoM) estimator of partitioned genetic correlations in bivariate LMMs. RGcor-mc leverages the structure of genotype data to obtain runtimes that scale sub-linearly with the number of individuals in the input dataset (assuming the number of SNPs is held constant). First, we perform extensive simulations to validate the accuracy of RGcor-mc to estimate genome-wide genetic correlation. We show RGcor-mc is as accurate as running GCTA on each group of SNPs separately. RGcor-mc is scalable and can compute the genetic correlations on the UK biobank dataset consisting of 430,000 individuals and 460,000 SNPs in 3 hours on a stand-alone compute machine, and partition the genetic correlation with no additional cost. We validate that the genome-wide and local genetic correlation are consistent with previous studies. We applied our method on UK biobank, and found novel significant genetic correlations. We document that rheumatoid arthritis (RA) and Crohn's diseases have negative genetic correlation ( $g=-0.22$ ,  $p=2.2e-13$ ).

## PgmNr 1543: ENmix: An efficient and comprehensive analysis tool set for Illumina HumanMethylation BeadChip.

**Authors:**

Z. Xu; J. Taylor

[View Session](#) [Add to Schedule](#)

**Affiliation:** NIH/NIEHS, Research Triangle Park, North Carolina.

---

The Illumina Methylation BeadChip is being broadly utilized in large epigenome-wide association studies. Methylation differences between groups are often very small and proper preprocessing of array data are critical to maximizing signal and minimizing noise. We developed a comprehensive R package, ENmix, to facilitate the analysis by removing unwanted experimental noise and improve accuracy and reproducibility of methylation measures. This new-enhanced package provides a set of quality control and data pre-processing tools, including background correction, dye bias correction, probe-type bias adjustment, low quality sample and probe filters, a low quality data point filter and imputation of missing data. In addition, we implemented a set of complementary functions, including data import from idat files, data visualization, identification of probes with multimodal distributions due to SNPs or other factors, exploration of data variance structure using principal component regression analysis plots, control surrogate variable information, an efficient algorithm oxBS-MLE to estimate 5-methylcytosine and 5-hydroxymethylcytosine levels, cell type proportion estimators and methylation age calculators. This stand-alone software is specifically designed to support large-scale data analysis, provides multi-processor parallel computing options for all functions, includes pipeline defaults for new users and supports 27K 450K and EPIC arrays.

## PgmNr 1544: Applying AMELIE to the NIH Intramural Cohort of the Undiagnosed Diseases Network.

**Authors:**

G.Y. Yu; S.M. Jafri; N. Balanda; A. Abdelbaki; C. Lau; B. Pusey; C. Toro; C.J. Tifft; W.A. Gahl; D.R. Adams

[View Session](#) [Add to Schedule](#)

**Affiliation:** NIH/NHGRI Bethesda, MD

---

The NIH Undiagnosed Diseases Program (UDP) was started in 2008 to find diagnoses for individuals with severe medical conditions that remained undiagnosed despite extensive workup. The UDP joined with other medical centers across the country to form the Undiagnosed Diseases Network (UDN) in 2013. Genomic medicine approaches have formed an important role in finding diagnoses for UDN participants. A central challenge in this field is the identification of diagnostic and research worthy DNA variation in the patient genome or exome. AMELIE (Automatic Mendelian Literature Evaluation) developed in Stanford is a computer tool to rank candidate genes based on phenotypes by finding gene-phenotype associations mined from literature. It takes a list of gene symbols and term identifications of the Human Phenotype Ontology and returns a ranked list of genes. We applied this tool to a subset of diagnosed cases enrolled in UDP to assess its usefulness in identification of variants from genome or exome sequencing data for a diagnosis. In 62 out of 83 diagnosed cases (75%), the genes harboring the diagnosis-associated variants are among the ranked top five. AMELIE is useful in ranking variants for identification of known disease-gene associations. It allows a long list of candidate variants including those intragenic non-coding regions to be automatically examined for reported associations. Further work on a subset of UDP undiagnosed cases and a few example cases will be presented.

## PgmNr 1545: Predicting pathogenicity of missense variants using deep learning.

### Authors:

H. Zhang <sup>1</sup>; W.K. Chuang <sup>2</sup>; Y. Shen <sup>1,3,4</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Systems Biology, Columbia University, New York, NY.; 2) Departments of Pediatrics and Medicine, Columbia University, New York, NY.; 3) Department of Biomedical Informatics, Columbia University, New York, NY.; 4) JP Sulzberger Columbia Genome Center, Columbia University, New York, NY.

---

Accurate predictions of genetic effects of missense variants is critical to the interpretation of genome sequence. Previously published methods have been used to improve power in genetic studies and to identify pathogenic variants in clinical genetic testing. However, the performance of these methods is suboptimal, due to issues related to the complexity of the mechanisms of pathogenic variants and suboptimal usage of large training data sets. In addition, it is often hard to interpret predicted scores. Here we report on a new prediction method MVP2, aimed to improve prediction performance and interpretability. MVP2 uses a deep learning approach to learn protein-context-specific pathogenicity of amino acid changes from large number of curated pathogenic variants. MVP2 learns the latent representation of amino acid types using context-wise embedding technique and uses convolutional neural network to learn the patterns of context including evolutionary conservation, structural properties and subgenic coding constraints in general population. To assess the utility in clinical genetic diagnostic testing, we obtained high-throughput functional data of ~1600-1800 missense variants in PTEN, TP53, and BRCA1 from recent publications as ground truth. We tested the ability to identify pathogenic variants excluding all known pathogenic variants in ClinVar. The area under the precision-recall curve of MVP2 is 0.69, 0.77 and 0.71 in PTEN, TP53 and BRCA1, respectively, substantially better than the best of previously published methods (0.50, 0.60 and 0.68, respectively). We also showed that our model and the scores are interpretable. Amino acids with similar biophysical properties are clustered together based on embedding vectors learned during training. We also found that amino acids mutated to proline in BRCA1's helix regions are predicted to be highly pathogenic, even in positions not well conserved. This is consistent with the fact that proline rarely occur in the middle of alpha-helix. This showed that, without prior knowledge, MVP2 captures context-dependent impact to protein function from amino acid substitutions. In summary, we show that MVP2 improves predictions of pathogenicity of missense variants by capturing protein context-dependent information using deep learning, and it has the potential to significantly increase the power in genetic studies and enhance the utility of clinical genetic testing.

## PgmNr 1546: AgentBind: Profiling context-specific determinants of transcription factor binding affinity with deep learning techniques.

### Authors:

A. Zheng <sup>1</sup>; M. Lamkin <sup>2</sup>; C. Wu <sup>3,4</sup>; H. Su <sup>1</sup>; M. Gymrek <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA.; 2) Department of Bioengineering, University of California San Diego, La Jolla, CA.; 3) Graduate Program in Bioinformatics and Systems Biology, University of California San Diego, La Jolla, CA.; 4) Department of Medicine, University of California San Diego, La Jolla, CA.

---

Binding of transcription factors (TFs) to DNA is one of the major transcriptional regulation mechanisms. Disruption of TF binding has been shown to affect a wide variety of phenotypes, including cancers and diabetes. Most TFs have intrinsic binding preferences for short DNA motifs of 5-20bp. Intriguingly, while millions of instances of a given TF motif may exist in the human genome, typically only a small fraction (less than 1%) of them are actually bound, suggesting sequence context around the motif also plays a significant role. To interpret the role of sequence context in TF-DNA recognition, we designed a novel deep learning framework, AgentBind. Our framework uses a convolutional neural network to classify bound vs. unbound TF motifs based on 1 kb of sequence context and make predictions of binding affinity for new samples (*binding scores*). It then applies Grad-CAM, a state-of-the-art model interpretation technique, to score the contribution of each nucleotide to the classification label (*context scores*). These scores can then be used to predict the impact of individual non-coding variants.

We evaluated AgentBind on 38 TFs profiled by ENCODE. Using only sequence information, AgentBind could accurately classify bound vs. unbound TF motifs (average AUC=0.945) and outperformed IMPACT, an alternative method which makes predictions based on dozens of epigenomics profiles. By interpreting the well-trained models, we were able to quantitatively annotate each nucleotide in the context and identified the k-mers enriched in the high-scored regions. This analysis identified many known co-factor pairs, such as SP1 and NFY.

We then tested the ability of AgentBind in predicting the impact of non-coding variants. We used LD-score regression to quantify the heritability explained by single-nucleotide polymorphisms (SNPs). Strikingly, SNPs in regions with high *binding scores* for TFs Gata3, Stat3, and Foxp3 were strongly enriched for explaining the heritability of rheumatoid arthritis (enrichment 20-80). We then analyzed SNPs from the 1000 Genomes Project and found that SNPs with high *context scores* (top 1%) were enriched for singleton variants compared to SNPs with low scores (p=0.0069), suggesting they are under increased selection. Taken together, these results suggest AgentBind scores accurately pinpoint genetic variants influencing binding of nearby TFs and can be used as a precise prioritization metric to identify disease-associated variants.

## PgmNr 1547: Evaluating nanopore sequencing data processing pipelines for structure variation identification.

### Authors:

A. Zhou <sup>1</sup>; T. Lin <sup>1</sup>; J. Xing <sup>1,2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Genetics Department, Rutgers, the State University of New Jersey, Piscataway, NJ, USA; 2) Human Genetic Institute of New Jersey, Rutgers, the State University of New Jersey, Piscataway, NJ, USA

---

### Background

Structural variations (SVs) account for about 1% of the differences between human genomes and play a significant role in phenotypic variation and disease susceptibility. Widely adopted next generation sequencing technologies have short read-length which limits their ability to identify SVs. The emerging nanopore sequencing technology applies real-time single-molecule sequencing and can generate long sequence reads. This technology can potentially provide better SV identification. However, the available tools for aligning long-read data and detecting SVs have not been thoroughly evaluated.

### Results

Using three human nanopore datasets, including both empirical data and simulated reads, we evaluated four alignment tools and three SV detection tools. We also evaluated the performance of the consensus calls from multiple SV callers and the impact of sequencing depth on SV detection. SV callers' performance varies depending on the SV types. For an initial assessment we recommend using aligner minimap2 in combination with the SV caller Sniffles because of their speed and relatively balanced performance. For users with specific requirements for SV types, precision or recall rates, other combinations or consensus call sets could provide better results.

### Conclusions

We established a workflow for evaluating aligners and SV callers for nanopore sequencing data. Our results indicate that optimizations are needed to improve SV detection accuracy and sensitivity. The nanopore technology keeps improving and the nanopore sequencing community is likely to grow accordingly. In turn, better benchmark call sets will be available to more accurately assess performance of available tools and facilitate further tool development.



## PgmNr 1548: A rare variants association test by weighting variants using sigmoid function.

### Authors:

N. Zhu <sup>1,2</sup>; W.K. Chung <sup>1</sup>; Y.F. Shen <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics molecular genetics, Columbia University, New York City, New York.; 2) Department of system biology, Columbia University, New York City, New York

---

Statistical tests of burden of rare variants in individual genes is one of the main approaches to identify new risk genes in genome or exome sequencing studies of human diseases. In many studies, it is necessary to use large-scale data generated outside of the study as controls. This strategy improves statistical power of detecting association of rare variants, but inevitably introduces technical batch effect between cases and controls. A practical method would be robust to such subtle technical batch effect. Additionally, missense variants are the most common type of genetic variation in protein-coding regions, but many missense variants are not deleterious, and including these in association tests would decrease signal to noise ratio and reduce power. Previous it was shown that pre-select deleterious missense variants based on in silico prediction can improve power (Zuk et al 2014); however, the optimal threshold of defining deleterious variants is likely different in different genes (Wilfert et al 2016). Furthermore, different risk variants from the same gene likely have different effect size.

We reason that relative risk of a missense variant in a true risk gene is correlated with in silico prediction scores, and the correlation can be modeled by a sigmoid function. In this work, we propose radTest, a burden test method using a logistic function to weight the variants by in silico prediction of deleteriousness. In each gene, we try a range of midpoint values to reach maximal significance in case-control comparison, and then permute the data to build background model of the test statistics to calibrate p-values. To benchmark performance, we obtained 193 rare variants in cases and controls from recent publications in BMP2 (Machado et al 2015), a major risk gene of pulmonary arterial hypertension. This data set allows us to simulate conditions without making unrealistic assumptions about dependency of effect size on predicted deleteriousness. We did observe a sigmoid shape in relative risk over predicted deleteriousness in the data. We then model a range of effect size by randomly switch case/control labels for a subset of rare variants, and showed that compared with current standard methods, radTest has superior power to detect association under realistic conditions with modest effect size. Finally, we show that radTest is robust to randomly injected small batch effect, making it applicable to genetic studies in which such batch effect is unavoidable.

## PgmNr 1549: Galaxy as an integrative environment for microbiome analysis and visualization.

### Authors:

D. Blankenberg; The Galaxy Team

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genomic Medicine Institute, Cleveland Clinic, Lerner Research Institute, Cleveland, Ohio.

---

A microbiome is a prime example of a complex biological system, a tangled web of multi-layered interactions, of which we are only beginning to be able to explore, elucidate, and command. The accelerating expansion in the kind and availability of high-throughput assays and technologies coupled to the increasing accessibility, refinement, and accuracy of molecular techniques to probe microbiomes continues to push the capability boundaries of computational analysis and data visualization. These physical experimental efforts can result in a wide-range of disparate datatypes and datasets that span across multiple domains and breach the threshold into big data classifications.

As microbiome datasets increase in size, scope, and complexity, how can we ensure that researchers are empowered to make effective use of not only the data, but also analysis tools, pipelines, visualizations, and compute resources? Furthermore, how can we make certain that researchers are able to do so in a Findable, Accessible, Interoperable, and Reusable (FAIR) manner?

Here we discuss the current status of Galaxy as a FAIR-capable integrative environment for microbiome analysis and visualization, including recent developments and future directions. Galaxy provides an ideal foundation for our microbiome analysis and visualization environment.

Galaxy (homepage: <https://galaxyproject.org>, main public server: <https://usegalaxy.org>) is a web-based scientific analysis platform used by tens of thousands of scientists across the world to analyze large biomedical datasets such as those found in genomics, proteomics, metabolomics, and imaging. It is a widely used and deployed open source platform for data integration and analysis in the life sciences. It can be used via a web GUI or a programmatic API. It has been deployed on everything from laptops to servers, to public cloud providers and supercomputers. There are well over 125 publicly accessible Galaxy instances, plus many national and public cloud deployments, and published containers and VMs. Many additional private installations have also been deployed.

## PgmNr 1550: FIREcaller: An R package for detecting frequent interacting regions from Hi-C data.

### Authors:

C. Crowley <sup>1</sup>; Y. Yang <sup>1</sup>; Y. Qiu <sup>2,3</sup>; B. Hu <sup>1</sup>; H. Won <sup>1</sup>; B. Ren <sup>2,4,5</sup>; M. Hu <sup>6</sup>; Y. Li <sup>1,7,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 2) Ludwig Institute for Cancer Research San Diego Branch, San Diego, CA.; 3) Bioinformatics and Systems Biology Graduate Program, University of California San Diego, La Jolla, CA.; 4) Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, CA.; 5) Institute of Genomic Medicine and Moores Cancer Center, University of California San Diego, La Jolla, CA.; 6) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH.; 7) Department of Biostatistics, University of North Carolina at Chapel Hill, NC.; 8) Department of Computer Science, University of North Carolina at Chapel Hill, NC.

High-throughput chromatin conformation capture (Hi-C) has been widely used to measure genome-wide chromatin spatial organization since first introduced in 2009. Analyzing Hi-C data has led to the discovery of structural readouts at a cascade of resolutions, including: A/B compartments, topologically associating domains (TADs), chromatin loops at Kb resolution, and significant chromatin interactions. Among these Hi-C readouts, TADs and chromatin loops are largely conserved across cell types, while A/B compartments and chromatin interactions exhibit rather moderate levels of cell type specificity.

To identify Hi-C readouts that are better indicative of cell type or tissue-specific chromatin spatial organizations, we previously reported thousands of frequently interacting regions (FIREs) by studying a compendium of Hi-C data across 14 human primary tissues and 7 cell types. We defined FIREs as 40Kb genomic regions, which have substantially more frequent interactions with their neighboring regions. FIREs are enriched for tissue-specific enhancers and nearby tissue-specifically expressed genes, suggesting their potential relevance to tissue-specific transcription regulatory programs. In our original study, we relied on an in-house pipeline to identify FIREs, limiting the general application of FIRE analysis and the full exploration of tissue-specific chromatin interaction features from Hi-C data. Here, we present FIREcaller, a stand-alone, user-friendly R package to detect FIREs from Hi-C data. FIREcaller takes Hi-C contact frequency matrix as input, performs within- and cross-sample normalization, and outputs continuous FIRE scores, dichotomous FIREs, and super-FIREs. To demonstrate the application of our FIREcaller, we used the human hippocampus Hi-C data from our previous study, as well as cortical tissue samples across two developmental epochs, fetal and adult. Among our results, we show that in the human hippocampus tissue, a super-FIRE overlaps with a schizophrenia-associated GWAS SNP rs9960767, leading to a hypothesis that the chromatin spatial organization may play a role in gene regulation in this region, which may ultimately affect the risk of schizophrenia. In addition, in the cortical tissue, we observe that FIREs dynamic across brain developmental stages are closely associated with developmental gene regulation. Overall, we believe that our FIREcaller will become a useful tool in studying tissue-specific chromatin spatial organization features.

## PgmNr 1551: *Mezcal*: Rapid and accurate local ancestry inference at scale using conditional random fields.

### Authors:

I.K. Jordan<sup>1,2,3</sup>; S.D. Nagar<sup>1,2,3</sup>; H.F. Espitia-Navarro<sup>1,2,3</sup>; L. Rishishwar<sup>2,3</sup>; A.B. Conley<sup>2,3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA; 2) IHRC-Georgia Tech Applied Bioinformatics Laboratory, Atlanta, GA; 3) PanAmerican Bioinformatics Institute, Cali, Colombia

---

The genetic architecture of human health is shaped by our ancestry, with individuals' ancestral origins influencing linkage disequilibrium patterns and the presence of disease-associated mutations. Genome sequence or whole genome genotype data can be analyzed to infer patterns of local genetic ancestry, i.e. the ancestral source of specific chromosomal regions. With the increasing number of large-scale human genome sequencing projects and biobanks that sample from cosmopolitan populations, there is a growing need for rapid and accurate approaches to local ancestry inference (LAI). However, scalable tools for the reliable inference of local ancestry are currently not publicly available to the research community. We have developed the LAI implementation *Mezcal*, which uses a conditional random field (CRF), parameterized on random forest classifiers, to predict the local ancestry composition of human genome sequences drawn from diverse, admixed populations. Our CRF is distinguished by the use of variable transition probabilities, which are modeled on the local genomic context. Our LAI implementation allows users to save their custom random forest classifiers generated via the analysis of ancestral reference and admixed genomes for the rapid characterization of individual samples as they become available in the future, thereby overcoming the N+1 problem often faced by bioinformatics scientists. We tested the performance of *Mezcal* on simulated and real-world human genome sequence datasets, characterized by combinations of two and three distinct continental ancestry components. Our approach to local ancestry inference yields a 50x time performance improvement over the current state-of-the-art, with no appreciable loss of accuracy. We were able to achieve an average of 98% accuracy for the calculation of African, European, and Native American ancestry components from six admixed populations in the Americas.

## PgmNr 1552: Use of burden test to identify the association of rare functional variants to rare Mendelian phenotypes.

### Authors:

R. Martin <sup>1</sup>; D. Valle <sup>1</sup>; H. Ling <sup>2</sup>; E. Wohler <sup>1</sup>; N. Sobreira <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA.; 2) Center for Inherited Disease Research, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

---

The identification of causative rare variants associated with rare Mendelian phenotypes is challenging but it can be facilitated by the increasing availability of genomic data. Here we present an analysis based on a burden test using a multisample VCF file containing whole exome sequencing (WES) data from 2662 samples sequenced as part of the Baylor-Hopkins Center for Mendelian Genomics (BHCMG) project. FASTQ files were aligned to the reference genome (GRCh37) with the Burrows-Wheeler Alignment (BWA 0.5.10) resulting in SAM/BAM output. GATK best practices were applied in order to obtain a multisample VCF file. Non-Single Nucleotide Variants (SNVs) were ruled out and the final multisample VCF was annotated with ANNOVAR. SNVs were filtered based on RefSeq, ExAC, gnomAD and Kaviar annotations to include functional (missense, nonsense, and splicing variants) rare (MAF  $\leq$  0.01) SNVs. 1635 unrelated samples were assigned as controls. For each phenotype investigated, the case group included all the affected probands. A contingency 2x2 table containing the sum of individuals presenting at least one functional rare SNV was built for each gene containing control and case groups data. Fisher's exact test was determined and  $p < 0.05$  was considered as statistically significant. Selected genes were also annotated with gnomAD missense Z-score, gnomAD pLI score and RVIS. Next, for each cohort investigated, we selected the genes that were more mutated among the patients than among the controls ( $p < 0.05$ ) and had a missense Z-score  $> 1$  or a pLI score  $> 0.9$  and an RVIS percentile  $< 25\%$ . This pipeline was applied to four different cohorts including a group of 16 probands with Tarlov cyst, 36 probands with hyperhidrosis, 24 probands with early-onset inflammatory bowel disease, and 7 probands with systemic artery to pulmonary artery malformation (SAPAM). The final candidate genes for each cohort are being further investigated as possible novel disease genes. As an example, the analysis of the SAPAM cohort study identified 3 candidate genes: *RNF145*, *EHMT1*, and *MCF2L*. Meanwhile, we had already identified the *MCF2L* gene as causative of this phenotype by our standard cohort analysis where we found this gene mutated in 3 out of the 7 families being investigated. Therefore, we expect that this alternative approach based on the burden test will enable novel disease genes discoveries by the identification of the most conserved genes that are more mutated among our cohort groups.

## PgmNr 1553: TransAnno: Novel LiftOver software for a new genome assembly.

### Authors:

Y. Okamura <sup>1,2</sup>; K. Kinoshita <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Advanced Research Center for Innovations in Next-Generation Medicine, Tohoku University, Sendai, Miyagi, Japan; 2) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 3) Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan

---

Recently, ethnic specific genomes, such as AK1 (Seo, J., *et.al.*, (2016) Nature) or HX1 (Shi, L., *et.al.*, (2016) Nat Com) were published. These genome assemblies are promising to improve the discovery new disease-causing variations. However, thousands of studies have been performed on international reference genomes, such as GRCh37 or GRCh38, and a lot of information is only available at GRCh37 or GRCh38. When we use other genomes, we have to perform lift-over a lot of annotation data international reference genomes to ethnic genome.

To perform lift-over genomic variation annotations, CrossMap (Zhao, H., *et.al.*, (2014) Bioinformatics) and Picard (<https://broadinstitute.github.io/picard/>) are widely used. These tools work well to transfer among international reference genomes, but not work well to transfer among ethnic genomes. These tools have some problems to transfer annotations, when reference sequences were changed, and changes were introduced in de novo assembly based ethnic genomes. Especially, insertions and deletions were difficult to be handled with these tools.

Gene annotation is another major problem to use new genome. There is no commonly used tool to perform lift-over gene annotations. Since most software does not care about relationship among exons, transcripts and genes, we cannot use them to perform lift-over gene annotations. GENCODE project has software to transfer their annotation to GRCh37, but their algorithm could not be used to transfer other genomes.

To overcome these difficulties, we developed a new LiftOver tool, TransAnno. TransAnno can 3 tasks, the first task is creating a chain file from minimap2 (Li, H. (2018). Bioinformatics) result. Because of the great performance of minimap2, only a few hours are required to create a new chain file. The second task is performing liftover GENCODE or other gene annotations with considering exon and transcripts. The third task is performing liftover VCF file, even if reference sequence was changed. TransAnno can perform liftover important annotations and makes ethnic specific genomes more useful.

# PgmNr 1554: A new software ecosystem for high resolution ribosome profiling.

**Authors:**

H. Ozadam; C. Cenik

[View Session](#) [Add to Schedule](#)**Affiliation:** Department of Molecular Biosciences, University of Texas at Austin, Austin, TX.

---

Human genetic variation can alter protein function and gene expression to impact phenotype. Our understanding of how human genetic differences control RNA expression is rapidly expanding. Yet, protein abundance is poorly predicted by RNA expression alone. This discrepancy can be partially explained by translational control. The state-of-the-art method for transcriptome-wide measurements of translation is ribosome profiling. In this method, ribosome protected RNA fragments are sequenced to quantify transcript-level ribosome occupancy. In these experiments, sequence read lengths are variable and carry critical information about the translation state of the ribosomes. Therefore, analyzing ribosome profiling data involves computation and storage of multiple metrics for different ribosome footprint lengths. The current solution to this problem of naively using text files is inefficient in organization, computation and storage.

Here, we developed a complete software ecosystem including a new efficient binary file format for ribosome profiling. The first component, RiboFlow, is a pipeline that processes raw ribosome profiling reads to generate sequence alignments, and related statistics. RiboFlow is highly portable across a large number of computational environments with built-in capabilities for parallelization. Importantly, RiboFlow assembles the results into a new binary file format, that we named “ribo”. Ribo files store user-defined metadata and all quantities of interest grouped by different ribosome footprint lengths. The second component, Ribogadgets, provide an interface in Python and R, for writing and reading ribo files. Using Ribogadgets, users can efficiently access ribosome profiling metrics, generate plots and output results to text files.

Taken together, this new software ecosystem provides a complete solution for researchers to study translation seamlessly. Moreover, we provide ~20-fold improvement in storage compared to the conventional approach. As a case study, we will present an application of our software solution to a human ribosome profiling dataset exploring the impact of amino acid starvation. We expect our infrastructure to facilitate large-scale characterization of the relationship between human genetic variation and translation control.

## PgmNr 1555: MOPower: A multi-omics study power calculator to determine optimal sample size and choice of integration model.

### Authors:

H. Syed <sup>1,2</sup>; G. Otto <sup>1,2</sup>; D. Kelberman <sup>1,2</sup>; S. Castellano <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UCL Great Ormond Street Institute of Child Health, University College London, London, United Kingdom; 2) NIHR Great Ormond Street Hospital Biomedical Research Centre, London, United Kingdom

---

Multi-omics studies are being increasingly used to uncover the underlying mechanisms of clinical phenotypes by integrating information from the genome to the microbiome. Correctly integrating these multiple data types is currently the biggest challenge, with each existing method suited to a different kind of dataset. This question becomes more complicated in rare disease studies where the sample size is small, and where individual omics approaches have failed to achieve significant results. To address this challenge, we have developed the interactive R-shiny web application and Python reporting tool MOPower. This program simulates data using statistical distributions or real study/reference data via bootstrap resampling to ensure realistic disease-specific simulations. Omics features included in the simulation are SNP's, gene expression and DNA methylation for binary, longitudinal or time-to-event outcomes. Users have control over the study design parameters such as fold change, the minor allele frequency of variants and mean gene expression per group. Each data replicate is analysed using multi-omics integration models such as multivariate projection models, factor analysis and cluster analyses. Depending on the integration model, power is interpreted as the correlation between omics features or the power to detect an association with the phenotype of interest through analysis of  $n$  number of replicates. The output from the power analysis is displayed on the screen, but a detailed report can also be produced in HTML or PDF format. This report is customizable by the user and can include power and sample size calculations on several different factors, displaying the corresponding power plot, a false discovery rate plot and a statement of results interpretation. Furthermore, a comparison is made between each integration model for the user's specific power calculation scenario. Initial results suggest that the number of omics features included per analysis with varying sample sizes influences power. Additionally, factors such as increased dispersion for gene expression correlates with decreased power. Ultimately, MOPower will support in the design of multi-omics studies, providing crucial information on the ideal sample size and choice of an analytical model.



## PgmNr 1556: A webtool for NGS data management, interpretation, and reporting.

### Authors:

Wang; T. Zhuang; J. Shen; X. Lu; T. Gao; Y. Yang

[View Session](#) [Add to Schedule](#)

**Affiliation:** AiLife Diagnostics Inc, Pearland, Texas

---

NGS has been widely used in research and clinical settings. However, data interpretation remains the bottleneck to fully realizing genomic data. For WES or WGS, it is inefficient and impractical to search for causal variants without the aid of prioritization tools. However, the sensitivity and reliability of such tools is currently unclear. In addition, current tools often lack the flexibility that enables users to test different hypotheses of genetic etiologies underlying the complex nature of genotype-phenotype correlations.

Here we report a cloud-based webtool to empower NGS data process. The tool includes modules of LIMS, QC, variant calling from VCF or FASTQ, annotation, knowledge database inquiry, data interpretation, and reporting in one secured system. To facilitate interpretation and reporting, a prioritization score for each variant is automatically calculated based on variant classification, population and research data, as well as phenotype and inheritance correlation with known disorders. The streamlined process also allows for flexibility so reviewers can modify the parameters and add additional information (e.g. conclusions from functional studies) and update the prioritization scores accordingly.

As a pilot study, we analyzed ~100 clinical WES cases, of which previous studies identified likely disease-causing variants in 30 cases. Each case had ~200 rare variants in disease genes, which were fully reviewed by the previous manual analyses. By using the prioritization score in this tool, the contributing variants were identified by reviewing only the top 20 ranked variants for all cases. However, one VUS that was called as possibly contributing by manual review in a proband only exome case was scored as having no phenotype overlap by the tool, due to that the phenotype associated with the gene was not included in the HPO database.

Our results showed that the new tool can effectively streamline the NGS process. In particular, the prioritization scores, which take into consideration of both genetic and phenotype evidence, can be used as a guide to facilitate data interpretation. It should be noted a small percentage of disease genes, esp. newly published ones, may not have HPO annotation. Therefore for cases with negative or uncertain results, or partial molecular diagnosis, further searches based on genetic evidence and manual phenotype evaluations should be performed to ensure a thorough analysis.

# PgmNr 1557: MMIRNA\_ER: A software package for studying miRNA and mRNA expression relationships.

**Authors:**

Z. Xu

[View Session](#) [Add to Schedule](#)**Affiliation:** Seven Lakes High School, Katy, TX.

---

**Background:**

MicroRNAs or miRNAs, an abundant family type of non-coding RNAs, participate in post-transcriptional regulation through binding to the 3' UTRs of target genes or mRNAs and have been reported prevalent in metazoan genomes. A single miRNA can target many mRNAs, while many miRNAs are able to cooperatively target a single mRNA. This allows for fine-tuned gene expression regulation. Most existing miRNA and mRNA data from biological experiments only have individual samples generated and are not ready to be deployed for downstream analysis. The types of data are from either related or unrelated individuals, or from the same individual but different stages of disease progression. To study the interactions between miRNAs and mRNAs, several preprocessing steps focusing on normalizing and grouping individual raw data files into a systematic format are often necessary.

**Methods and Results:**

In this study, we have developed an efficient software package which can automatically take a list of input files for miRNA and mRNA with expression information from multiple samples for a given genetic disorder to determine their relationship. The program can scan through two input expression files (e.g. miRNA and mRNA) filled with a list of samples to report target pairs and get average expression calculated for each target pair. We have processed the input files of miRNA and mRNA from multiple cancer types available in The Cancer Genome Atlas (TCGA) project and reported the output and statistics (e.g. running time).

**Conclusions:**

Our program is user-friendly and flexible and can be used to process data files for studying the relationships between miRNA and mRNA target pairs by providing an easy computation of fold change and target pairs between two or more biological conditions.

## PgmNr 1558: Systematic detection of consanguinity through runs of homozygosity facilitates molecular diagnosis of rare disease patients.

### Authors:

L. Matalonga <sup>1,2</sup>; A. Papakonstantinou <sup>1,2</sup>; S. Laurie <sup>1,2</sup>; D. Piscia <sup>1,2</sup>; I. Gut <sup>1,2,3</sup>; O. Riess <sup>4,5</sup>; G.J. van Ommen <sup>6</sup>; H. Lochmüller <sup>1,7,8</sup>; S. Beltran <sup>1,2,3</sup>; NeurOmics consortium, BBMRI-LPC consortium and RD-Connect consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona, Spain; 2) Barcelona Institute of Science and Technology (BIST); 3) Universitat Pompeu Fabra (UPF), Barcelona, Spain; 4) Institute of Medical Genetics and Applied Genomics, Medical Faculty and University Hospital, Eberhard Karls University, Calwerstraße 7, 72076, Tübingen, Germany; 5) NGS Competence Center Tübingen (NCCT), Eberhard Karls University, Calwerstraße 7, 72076, Tübingen, Germany; 6) Leiden University Medical Center, Albinusdreef 2, 2333 ZA, Leiden, the Netherlands; 7) Department of Neuropediatrics and Muscle Disorders, Medical Center - University of Freiburg, Faculty of Medicine, Freiburg, Germany; 8) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON K1H 8L1, Canada; Division of Neurology, Department of Medicine, The Ottawa Hospital, Ottawa, Ontario, K1Y 4E9, Canada

---

Autozygosity is known to be associated with an increased risk of genetic rare disease (RD) and is thus a relevant factor to take into consideration when undertaking clinical genetic studies. NGS technologies allow the precise detection of genomic regions where a reduction in heterozygosity is evident and offer the opportunity to estimate autozygosity at the exome and genome level. Here more than 2400 whole exome sequencing (WES) datasets from the RD-Connect Genome-Phenome analysis platform (GPAP) (<https://platform.rd-connect.eu/>) were analysed and autozygosity assessed based on detection of long (>1Mbp) runs of homozygosity (RoH) using the PLINK software. The relationship between the number and length of detected RoH and different clinical and experimental parameters such as patient continental origin, family history, WES quality metrics, reported consanguinity status in clinical records and sequencing strategy was evaluated. This allowed us to establish four consanguinity ranges ("consanguineous", "probably consanguineous", "probably non-consanguineous" and "non-consanguineous") based on the overall RoH length to indicate if an individual is likely to be the progeny of a consanguineous pairing or not. To validate these findings, a subset of 280 cases from NeurOmics and BBMRI-LPC projects were classified according to the consanguinity ranges, followed by genomic data analysis applying RoH specific region filtering (>1Mbp) in the RD-Connect GPAP. This approach resulted in the re-classification of the consanguinity status of 10% of the patients, either through the detection of possible unstated consanguinity or by empirically correcting the consanguinity reported in the corresponding clinical record. Moreover, filtering of genomic data to RoH regions in experimentally classified consanguineous cases enabled us to reduce by 40-60% the number of candidate variants to be assessed. Finally, pathogenic variants were found within a RoH in 40% of the consanguineous cases enabling to reach a final molecular diagnosis for those patients. The consanguinity ranges defined provide clinical geneticists with an estimate of the degree of

inbreeding within a pedigree, enabling confirmation or (re)classification of consanguineous status, hence facilitating and accelerating the molecular diagnosis of RD patients who undergone WES. We acknowledge NeurOmics and BBMRI-LPC consortium members for their contribution

## PgmNr 1559: Exploring exonic copy number variants within a large clinical cohort with a variety of Mendelian disease traits.

### Authors:

H. Du <sup>1</sup>; Z. Coban-Akdemir <sup>1</sup>; T. Gambin <sup>2,3</sup>; X. Song <sup>1</sup>; R. Du <sup>1</sup>; H. Hijazi <sup>1</sup>; A. Jolly <sup>1</sup>; C.M. Grochowski <sup>1</sup>; C.M.B. Carvalho <sup>1</sup>; J.E. Posey <sup>1</sup>; J.R. Lupski <sup>1,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 2) Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland; 3) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas; 5) Department of Pediatrics, Baylor College of Medicine, Houston, Texas; 6) Texas Children's Hospital, Houston, Texas

---

The Baylor-Hopkins Center for Mendelian Genomics (BHCMG) program has performed exome sequencing (ES) and rare variant family based genomics in > 3000 families with a wide variety of Mendelian disease traits. Novel discoveries involving pathogenic variant alleles in over 850 'disease genes' (Dec 2011-May 2018) have focused on rare single nucleotide variants (SNVs) and indels. Additional copy number variant (CNV) analysis may further contribute to molecular diagnoses and offer a starting point to understand disease pathobiology. Many small exonic copy number variants (eCNVs), i.e., intragenic duplication or deletion variants of a few exons or even single exon, are too small (< 30 kb) to be detected by clinical array. Sequencing methods provide a scalable and unbiased way to investigate eCNVs genome-wide. The current methods developed are only able to consistently detect medium-sized CNVs in whole-exome sequencing (WES) data but not small ones (e.g., most 'read depth' - based copy number algorithms are optimized to detect heterozygous CNVs encompassing at least three consecutive exons).

In this study, we developed eCNVFinder, an algorithm that improves the signal-to-noise ratio of a single exon by clustering read depth data from samples before CNV calling, enabling the detection of rare small eCNVs from WES. When applied the method to a set of orthogonal validated CNVs from BCHMG, the method successfully recalled 24 out of 27 heterozygous deletions ranging from 2.3 kb to 5 Mb in size, 7 out of 8 triplications ranging from 3.9 kb to 2.5 Mb, and 12 out of 22 heterozygous duplications ranging from 3.6 kb to 3.5 Mb. An overall recall rate of 75.4% was achieved and 2 events involving only 2 exons were able to be detected. Additionally, by applying this method to BCHMG ES database, we detected eCNVs of *PARK2* in 4 subjects within an early onset Parkinson family that were not detected by other CNV calling algorithms such asXHMM. These results show eCNVFinder can identify additional eCNVs and potentially improve pathogenic CNV variant detection, which may in turn augment molecular diagnosis for patients.

## PgmNr 1560: Haplotype-aware CNV detection using Parliament2 on the eMERGE III Network.

### Authors:

T. Wu; F. Sedlazeck; S. Kalla; D. Muzzy; R. Gibbs; E. Venner; eMERGE Consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

---

Structural variation (SV) is a common type of genomic variation in the human genome and is frequently associated with genetic diseases. In the eMerge Phase III project, a targeted next-generation sequencing panel was designed to target 109 genes and more than 1500 SNVs from over 25,000 participant samples. Atlas-CNV, a clinically-validated, coverage-based copy number variant (CNV) calling tool was applied for the detection of CNVs. However, this approach can be further improved by using alleles from population data, potentially increasing the resolution and sensitivity of the method.

To address this issue, we have applied Parliament 2, an approach that has been previously applied to whole-genome and whole-exome data from other NIH-funded large-scale sequencing programs. The cloud-based Parliament 2 analysis software leverages multiple SV analysis methods, allowing the detection of multiple SV types though focusing mainly on CNVs in this analysis. First, we remapped data from all 25,000 eMERGE participants (starting from FASTQ(HGSC) or BAM(Broad)) to the GRCh38 reference genome. This was done to ensure comparability to large population datasets derived from WGS sequencing projects. The resulting bam files then go through the Parliament2 analysis on DNAnexus. This process uses analysis tools such as Lumpy, Manta, Breakseq, Breakdancer, Delly, and CNVnator to call all SVs.

Next, using SURVIVOR, we are annotating CNV calls against population-level SV information from previous WGS and WES studies, assigning each CNV a population frequency that can aid in interpretation and pathogenicity assignment. In addition, we are comparing the harmonized eMERGE events to ethnicity-specific alleles and using this data to better resolve our current set of CNV calls as well as potentially identifying complex events that are difficult to identify using a single-ethnicity reference. Lastly, for deletions, we will assess the concordance of the CNV calls with SNP data, potentially reducing false-positives. Our analysis so far has resulted in the detection novel copy number variants in over 100 clinically-relevant genes such as DSC2, CYP2D6, LDLR, MYH11, F5, PMS2, TGFB1, BRCA1, BRCA2, and TP53.

## PgmNr 1561: An accurate and powerful method for copy number variation detection.

### Authors:

F. Xiao <sup>1</sup>; X. Luo <sup>1</sup>; N. Hao <sup>2</sup>; Y. Niu <sup>2</sup>; G. Cai <sup>3</sup>; C. Amos <sup>4</sup>; H. Zhang <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology and Biostatistics, Univ South Carolina, Columbia, South Carolina.; 2) Department of Mathematics, Arizona University, Tucson, AZ; 3) Department of Environmental Health Sciences, Univ South Carolina, Columbia, South Carolina.; 4) Institute of Clinical and Translational Medicine, Baylor College of Medicine, Houston, TX; 5) Department of Biostatistics, Yale University, New Haven, CT

---

Integration of multiple genetic sources for copy number variation detection (CNV) is a powerful approach to improve the identification of variants associated with complex traits. Although it has been shown that the widely used change point based methods can increase statistical power to identify variants, it remains challenging to effectively detect CNVs with weak signals due to the noisy nature of genotyping intensity data. We previously developed modSaRa, a normal mean-based model on a screening and ranking algorithm for copy number variation identification which presented desirable sensitivity with high computational efficiency. To boost statistical power for the identification of variants, here we present a novel improvement that integrates the relative allelic intensity with external information from empirical statistics with modeling, which we called modSaRa2. Simulation studies illustrated that modSaRa2 markedly improved both sensitivity and specificity over existing methods for analyzing array-based data. The improvement in weak CNV signal detection is the most substantial, while it also simultaneously improves stability when CNV size varies. The application of the new method to a whole genome melanoma dataset identified novel candidate melanoma risk associated deletions on chromosome bands 1p22.2 and duplications on 6p22, 6q25, and 19p13 regions, which may facilitate the understanding of the possible roles of germline copy number variants in the etiology of melanoma.

## PgmNr 1562: DEFOR: Depth- and frequency-based somatic copy number alteration detector.

### Authors:

H. Zhang <sup>1</sup>; X. Zhan <sup>1</sup>; J. Brugarolas <sup>2</sup>; Y. Zie <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Population and Data Sciences, UT SOUTHWESTERN MEDICAL CENTER, Dallas, Texas.; 2) Department of Internal Medicine and Kidney Cancer Program, Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas.

---

Somatic copy number alteration (SCNA) is the change in copy number that arises in somatic cells. SCNAs have been observed frequently in tumors, and recurrent SCNAs have also been identified in several types of tumors. SCNA detection plays an important role in studying the mechanisms of tumor development and in guiding therapeutics. In recent years, SCNA detection using high-throughput sequencing has become popular because of rapid developments in sequencing technology.

Although several bioinformatics methods have been developed for SCNA detection using exome-sequencing data, these methods rely on a strong assumption that the median or mean value of the copy numbers of the genome is two (normal status). This assumption is generally used to give an estimation of the depth or depth ratio that represents the normal status. The assumption usually holds for germline copy number variations, because most parts of the genome don't have copy number changes. However, the assumption may not hold in tumor cells where large scale copy number alterations were observed frequently. In such situations, the mean or median value of copy numbers across the genome may be not two, and so the copy number estimate based on that assumption would be no longer reliable. Therefore, existing methods may not perform well in calling SCNAs for the unstable tumor genomes. Samples from tumor tissue are usually contaminated with normal cells from nearby normal tissue, and the purity of different tumor samples varies substantially. Because the purity of the tumor sample affects the observed copy number, it adds another layer of variations in SCNA detection of tumor samples.

We developed a new method, DEFOR, to detect SCNAs in tumor samples from exome-sequencing data. DEFOR supports the estimation of copy numbers in six different statuses and adopts a model considering allele frequency, depth and purity. The performance of DEFOR is outstanding compared with the other five existing methods in the evaluation, even for unstable tumor genomes with a larger proportion of SCNAs.



## PgmNr 1563: Undiagnosed Diseases Program Database (UDPdb) of genomic data: A model for cohort data consolidation, insights and reanalysis.

### Authors:

N.J. Balanda <sup>1</sup>; D.L. Ross <sup>1</sup>; A.S. Brown <sup>1</sup>; B.N. Pusey <sup>1</sup>; C.C. Lau <sup>1</sup>; D.R. Adams <sup>1,2</sup>; W.A. Gahl <sup>1,2</sup>; Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Undiagnosed Diseases Program, NIH, Bethesda, MD.; 2) Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD

---

The Undiagnosed Diseases Program (UDP) commenced in 2008 with its primary goal to uncover causes of previously undiagnosed diseases. The development of the program coincided with the development of genomic sequencing technology. As high throughput sequencing advanced, the UDP adapted these methods into its standard operating procedure, contributing to a diagnosis rate of ~25% for UDP patients. In the last 10 years, the UDP has sequenced more than 400 genomes and 1600 exomes of affected individuals and their genetically informative family members. Under standard procedure, this genomic data is analyzed on a per-case basis, leveraging the power of adding family members to the analysis when possible. The bioinformatic team in the UDP is constantly looking to test and develop new ways to revisit old data, with hopes of improving the diagnostic rate through reanalysis of old cases with new techniques. With a growing cohort, there is an increasing potential for interrogating the cohort as a whole. However, in order to ask high-yielding questions from the cohort, genomic information from all of the individuals must be consolidated. Here, we consolidate genomic variants from all available UDP sequence data into a comprehensive cohort sequence database. To do so, each genome/exome is reworked such that alignment, variant calling, and joint genotyping methods are standardized and uniform across all data. The resulting variant calls are then annotated and passed into GEMINI, a portable SQLite database, enabling novel interrogation of the existing UDP sequence cohort and supplemented analyses for new patients. Having data consolidated in this format enables efficient parallelization of reanalysis and rapid interrogation of gene or patient-specific questions. Through reanalysis, we have identified new diagnoses in our cohort in the genes *NTRK2*, *ACER3*, *NDUFA6*, *DHDDS*, and *RNF216*. Other potential diagnostic candidates are being investigated to confirm splicing effects. In addition to reanalysis, by leveraging the data of the entire cohort, we are able to prioritize bioinformatic outliers as gene discovery candidates by looking specifically at predicted deleterious mutations in genes that are intolerant to variation. The UDPdb is a model for large cohort data consolidation and empowers new analysis methodology in the Undiagnosed Diseases Program.

## PgmNr 1564: Compendium of Arab genetic variants and metabolic trait associations.

### Authors:

A. Channanath; P. Hebbar; S.E. John; F. Al mulla; T.A. Thanaraj

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetics and Bioinformatics, Dasman Diabetes Institute, Kuwait, Kuwait

---

The Arab region is characterized by a high prevalence of metabolic disorders. There is a lack of publicly-available data on genetic variants from Arab populations. We have developed a database of genetic variants comprising variants from Exome sequencing and genotyping arrays as identified on native individuals and catalogued from Kuwait. Further we catalogued statistical associations signals of metabolic traits with imputed genotypes using array data. We provide a user-friendly web query interface to access the genetic information.

### Data:

**Exome sequencing:** We sequenced 291 whole exomes of unrelated, healthy native Arab individuals from Kuwait to a median coverage of 45X and characterized 170,508 single-nucleotide variants (SNVs).

**Genome-wide genotyping:** of ~1300 & 1900 Arab ethnic individuals generated from Illumina human OmniExpress and cardiometabo arrays respectively.

**Imputation using 1KGP haplotypes:** We cataloged variants imputed from two GWAS data sets individually using 1000G Phase 3 reference panel.

**Genome-wide associations:** We subsequently statistically tested association of common ( $\geq 5\%$  frequency), well imputed variants across 13 metabolic quantitative traits with RAW and inverse normal variance traits. Importantly, this data includes 1222 unique variants (from 232 genes) from 1442 associations with  $p\text{-value} < 0.05$  for 13 metabolic traits in Arab population upon comparing with 313 metabolic trait variant associations published in GWAS Catalog.

The genetic variants were systematically annotated by way of using various other databases and tools in ANNOVAR package and we catalogued computational predictions of pathogenicity of variants such as SIFT, Polphen2, MutationTaster and MutationAssessor, FATHMM etc. In addition, variants were systematically annotated across a number of relevant databases.

We implemented the current trending technologies such as MongoDB (to store data), JQuery (web interface) along with Python framework called Django. Th queries can be performed using genetic

variant IDs (rsIDs), gene names, positions or ranges of genomic positions. This database is freely available at [www.dgr.dasmaninstitute.org](http://www.dgr.dasmaninstitute.org) for the academic and not for profit research institutions.

## **PgmNr 1565: Gabriella Miller Kids First Data Resource Center: Collaborative platforms for accelerating research in genetic etiology of pediatric diseases.**

### **Authors:**

A.P. Heath <sup>1</sup>; C. Yung <sup>2</sup>; Y. Zhu <sup>1</sup>; M. Mattioni <sup>3</sup>; Z.L. Flamig <sup>5</sup>; Y. Tang <sup>5</sup>; B. Farrow <sup>1</sup>; J. Lilly <sup>1</sup>; Y. Guo <sup>1</sup>; P. Raman <sup>9</sup>; P. Storm <sup>1,10</sup>; S.L. Volchenboum <sup>4,8</sup>; J. Nazarian <sup>6</sup>; N. Vasilevsky <sup>12</sup>; J. DiGiovanna <sup>3</sup>; M. Haendel <sup>12,13</sup>; R. Grossman <sup>5</sup>; B. Davis-Dusenbery <sup>3</sup>; D. Taylor <sup>7,9</sup>; V. Ferretti <sup>11</sup>; A. Resnick <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Data Driven Discovery in Biomedicine, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Ontario Institute of Cancer Research, Toronto, Ontario, Canada; 3) Seven Bridges Genomics, Cambridge, MA, US; 4) Center for Research Informatics, The University of Chicago, Chicago, Illinois, USA; 5) Center for Translational Data Science, The University of Chicago, Chicago, Illinois, USA; 6) Children's National Medical Center, Washington, DC, USA; 7) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 8) Department of Pediatrics, Biological Sciences Division, The University of Chicago, Chicago, Illinois, USA; 9) Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 10) Department of Neurosurgery, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 11) CHU Sainte-Justine Research Center, Montreal, Quebec, Canada; 12) Oregon Health & Science University, Portland, Oregon, USA; 13) Oregon State University, Corvallis, Oregon

---

Since launching to the public in October 2018, the Gabriella Miller Kids First (Kids First) Data Resource Center (DRC), has made an increasing number of pediatric, especially trio and family focused, genomic studies available to the research community. By the end of 2019, genomic and clinical data from over 15,000 participant samples will be available across a variety of structural birth defect and pediatric cancer cohorts. The DRC has architected a secure, cloud-based platform that supports the ability of researchers to not only find, access and reuse data, but also integrate, collaborate, and analyze data quickly at scale. A “best-of-breed” approach has been taken to develop a portal, with reusable components from Overture, as the entry point for Kids First data. From there, users can use integrations with platforms such as Cavatica for bioinformatics workflows and pedCBioPortal for cancer genomic visualizations. Additionally, a set of framework services, powered by Gen3, provide a foundation for interoperability with other large-scale data sources, platforms and a growing ecosystem of analysis and visualization applications that provide capabilities for rapid in-place analysis without download.

Recognizing the value in both the source data collected by the original studies and the power in having harmonized genomic and clinical data for cross-study analysis, the DRC makes both available in the platform. The new “explore data” feature on the portal allows users to search and browse in real-time across all Kids First studies in order to identify virtual cohorts of interest for further study. Within the portal, these cohorts can be saved and shared with collaborators for iterative refinement and analysis. With appropriate approvals, the associated data can be accessed and analyzed seamlessly in Cavatica, or other platforms with interoperable framework services. To lower the barrier

to entry, the DRC has been piloting rapid-access cloud credits and providing support to researchers performing custom analysis on the platform. On the near term roadmap is the ability to create cohorts utilizing genomic features and improved visualizations for germline and family-based genomic data.

# PgmNr 1566: lncRNAKB: A comprehensive knowledgebase of long non-coding RNAs.

## Authors:

F. Seifuddin; A. Suresh; K. Singh; Y.C. Chen; V. Chaitankar; I. Tunc; M. Pirooznia

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bioinformatics and Computational Biology, National Heart, Lung, and Blood Institute, National Institutes of Health, BETHESDA, MD.

---

**Motivation:** There are several databases that exist for annotation of human long non-coding RNAs (lncRNAs) that contain between 20,000 to 100,000 entries. These databases contain unique and overlapping lncRNAs that have been identified by next generation sequencing (NGS) methods. The information on lncRNAs provided in these databases is not rigorous thus, making it difficult to understand their molecular and cellular functions. Consequently, there is a need to systematically and carefully combine these annotations, create a non-redundant resource and provide valuable functional information on lncRNAs.

**Results:** We have created the long non-coding RNA knowledgebase (lncRNAKB) by methodically integrating six widely used lncRNAs annotation databases (CHESS, FANTOM, LNCipedia, NONCODE, MiTranscriptome and BIGTranscriptome). We present an annotation of a large number of unique lncRNAs ( $n=77,199$ ). The lncRNAKB incorporates coding potential, classification/localization with respect to messenger RNAs (mRNAs), gene expression, tissue-specificity scores, expression quantitative trait loci (eQTL)-regulated lncRNA genes, phylogenetic conservation and functional characterization to identify co-expressed mRNAs that would provide potential understanding on lncRNAs function. A machine learning approach was used to calculate the coding potential scores and classify the lncRNAs in the lncRNAKB annotation database as putative lncRNAs or mRNAs. Gene expression data of 9,074 RNA-seq samples, collected from the Genotype Tissue Expression (GTEx) project was used to provide tissue-specific expression profiles and tissue-specificity scores in 31 solid organ human normal tissues. Using whole genome sequence (WGS) genotype data of 652 subjects and tissue-specific gene expression data from the GTEx project we calculated *cis*-eQTLs in all tissues. We calculated and compared evolutionarily exon conservation between lncRNAs and protein-coding genes (PCGs) using an alignment of 30 vertebrate species. We used Weighted Gene Co-expression Network Analysis (WGCNA) to identify co-expression modules encompassing lncRNA-mRNA pairs that were subjected to enrichment analysis using Gene Ontology (GO) pathways to identify meaningful biological processes that lncRNAs could be potentially involved in and created dynamic Cytoscape networks for exploration and visualization. All components are provided in a user-friendly web interface. lncRNAKB is available at <http://www.lncnakb.org>

## PgmNr 1567: VISDB: A comprehensive database for human disease-related virus integration sites.

### Authors:

D. Tang<sup>1</sup>; B. Li<sup>1</sup>; T. Xu<sup>1</sup>; R. Hu<sup>1</sup>; Z. Zhao<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, 77030, USA; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

---

Viral infection is a significant cause of a variety of human disease, including cancer. Viruses may intrude human cellular system and even insert their genetic material into the host genomes, thus affecting the normal function of the molecules in the host cells. With the advance of next generation sequencing, many studies have discovered virus infection in the host cells and in some cases, integration to the host genomes. The studies also found that the integration sites (VIS) prefer next to oncogenes and common fragile sites. Viral integration can be a key driving event in virus related cancer such as hepatocellular carcinoma and cervical carcinoma. Thus, it is important to investigate the integration events in the human genome and understand the molecular mechanism of viral integration. In this study, we systematically collected and curated all the VIS in literature and databases, and developed the Virus Integration Site Database (VISDB, <https://bioinfo.uth.edu/VISDB/index.php>). The data features include characteristics of the malignant disease, chromosome region, genomic position, target and nearby genes, fragile sites, viral-host junction sequence and potential virus mediated host microRNAs (miRNAs). The current version of VISDB includes 77,336 integration sites in 5 DNA oncovirus (HBV, HPV, EBV, MCV, AAV2) and 4 RNA retrovirus (HIV, MLV, HTLV-1, XMRV) curated from 109 publications, and correlated with 14986 target genes (including 447 oncogenes and 678 tumor suppressor genes), 123 fragile sites, 27 diseases and 2276 human samples. The data was further curated by linking to public databases such as NCBI GenBank, OGene, TSGene, HumCFS, etc. We present a universal virus integration site model and provide a visualization method to display the integration event and target genes, upstream and downstream genes in the vicinity of integration site. VISDB provides several functions including data search, browse, curation, download and feedback and upload for the user. The database offers a unique resources that can help users for better understanding the fragile (risky) genes or loci by virus integration, especially oncogenes and suppressor genes in cancer cases.

## PgmNr 1568: TogoVar: Integrated database among Japanese genomic variation databases.

### Authors:

L. Toyo-oka <sup>1</sup>; N. Mitsuhashi <sup>1</sup>; M. Kawashima <sup>1</sup>; T. Fujiwara <sup>2</sup>; T. Katayama <sup>2</sup>; S. Kawashima <sup>2</sup>; Y. Tateisi <sup>1</sup>; T. Takagi <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Bioscience Database Center, Japan Science and Technology Agency, Chiyoda-ku, Tokyo, Japan; 2) Database Center for Life Science, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Japan; 3) Toyama University of International Studies

---

Population-specific information on genomics is important to reveal causal variations when conducting human genetic studies and interpreting the result. The Exome Aggregation Consortium (ExAC) and the Genome Aggregation Database (gnomAD) have been providing population-specific genomic information. However, there is a limitation on Japanese specific allele-frequency information in ExAC and gnomAD because the sample number of Japanese in those databases are only seventy-six. Thus, the Integrative Japanese Genome Variation Database (ijGVD) and Japanese Multi Omics Reference Panel (jMorp) are well-used databases as a reference for population-specific genetic database of Japanese due to the biggest sample number of Japanese Whole Genome Sequencing (WGS) data up to 3,554. Another reference database for allele frequency of Japanese is Human Genetic Variation Database (HGVD) which contains allele frequency information from 1,208 Whole Exome Sequencing (WES) samples. Because there is no database integrating information on allele frequency data from ExAC, ijGVD and HGVD, ClinVar information and related publication data, we launched TogoVar database (<https://togovar.biosciencedbc.jp/>) as one-stop service to obtain the information. Further, we created another two new databases of allele frequency named JGA-SNP and JGA-NGS. The former one contains allele frequencies from SNP-chip analyses and the latter one contains that from WES analyses. The original data are deposited in the NBDC Human Database (<https://humandbs.biosciencedbc.jp/en/>) that our database center has conveyed to share genomic-phenotypic data in cooperation with the DNA Data Bank of Japan since 2013. Thus, allele frequency information of both JGA-SNP and JGA-NGS will be updated when NBDC Human Database receive controlled-access data from data submitters.

TogoVar users can find variants interested in by using gene name, rs number, HGVS representation and disease name from over 80 million variants. Further, users can filter a range of data based on criteria they define: allele frequency, database name, genetic consequence, variant calling quality, and the score generated by SIFT and Polyphen2.

In the poster, the recent achievement on data integration and establishment for searching function will be presented.



## PgmNr 1569: Functional Annotation of Variants Online Resource (FAVOR).

### Authors:

H. Zhou<sup>1</sup>; X. Li<sup>1</sup>; Z. Li<sup>1</sup>; S. Kumar<sup>2</sup>; A. Asok<sup>1</sup>; B. Li<sup>3</sup>; S. Sunyaev<sup>6,7</sup>; S. Liu<sup>1,4</sup>; Z. Weng<sup>5</sup>; M. Gerstein<sup>2</sup>; X. Lin<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biostatistics Department, Harvard T. H. Chan School of Public Health, Boston, MA.; 2) Yale School of Medicine, Yale University, New Haven CT; 3) Vanderbilt University Medical Center, Vanderbilt University, Nashville, TN; 4) Dana-Farber Cancer Institute, Boston, MA.; 5) University of Massachusetts Medical School, Worcester, MA; 6) Brigham and Women's Hospital, Boston, MA; 7) Harvard Medical School, Boston, MA

---

Functional annotation data has proven effectiveness in wide variety of studies, including rare-variant association test, fine mapping, etc. We created a most comprehensive functional annotation database by merging functional annotation data from following major sources: popular functional annotation databases, Epigenetics database, annotation data from collaboration labs and also from many individual studies and publications. Currently our functional annotation database has more than 260 annotation channels, and the annotation database is growing rapidly and steadily.

To improve the conherency, efficiency and usability of applying functional annotation data in downstream analysis of genetic variants, we pioneered the effort of building in the functional annotation results into GDS format for the first time, resulted in the creation of AGDS. It represented a well-rounded solution for storing functional annotation results for genetic variants, with optimal search and retrieval performance while keep data size very small. In AGDS file genotype data is recorded with matched huge number of functional annotation scores for each variant, this increases the effective application of functional annotation data in variants analysis. As an R friendly data structure, AGDS achieves perfect format and software compatibility, with seamlessly conversation to other file formats and compatibility with all the tools that working with GDS.

Because functional annotation data is very useful for different fields, building an easy-to-access web portal will be critical to empower many studies. We created FAVOR (Functional Annotation of Variants Online Resource) using R Shiny, which is the first web portal implementation using GDS. FAVOR supports single variant query and batch annotation. For single variant query, users can input genomic position or rs number. For batch annotation, users can upload variant list to FAVOR with annotation results display in table format on web interface and also have options to download annotation results in three different formats (GDS, VCF, CSV). The back-end GDS database has removed the genotype data from all-in-one GDS, this modification is especially suitable for web portal development and also for public accessing and sharing. Moreover, the offline and online retrieval and query operations stay the same.

We hope our effort of creating the functional annotation database and FAVOR will be one of the core knowledgebases for genetics research.

## PgmNr 1570: Scout: A decision support framework for clinical diagnostics.

### Authors:

L. Stranneheim <sup>1,2,4</sup>; M. Magnusson <sup>1,2,4</sup>; C. Rasi <sup>4</sup>; D. Nilsson <sup>1,3,4</sup>; V. Wirta <sup>5,6</sup>; A. Wedell <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Medicine and Surgery, Karolinska University Hospital, Stockholm, Sweden; 2) Centre for Inherited Metabolic Diseases, Karolinska University Hospital, Stockholm, Sweden; 3) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 4) Science for Life Laboratory, Karolinska Institutet Science Park, Solna, Sweden; 5) SciLifeLab, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden.; 6) SciLifeLab, Department of Microbiology, Tumor and Cell biology, Karolinska Institutet, Stockholm, Sweden

---

A number of paradigm shifts in our ability to sequence DNA have paved the way to utilising whole genome sequencing in a clinical setting. Massively parallel sequencing (MPS) is now rapidly becoming the first tier test in clinical genomics with dramatic improvements in diagnostic yield and time to diagnosis. However, the enormous information content and sheer size of the generated data from a single MPS run represents a huge threshold for clinical implementation.

Variant calling format (VCF) or the binary equivalent (BCF) files are the ubiquitous storage formats for variant calls identified in a MPS data. In principle, they are not more than tab-delimited text files. They lack a powerful and standardized data model to enable sophisticated genetic analysis, which can be translated in a clinical setting. Converting VCF into a GEMINI database goes a long way to enable this type of complex processing. However, using this requires knowledge of SQL and the internal schema, which cannot be expected from a clinical or research user.

*Scout* is an open-source decision support framework for clinical diagnostics with graphical user interface (GUI) for MPS associated data. Scout lets you organize, collaborate, share, inspect, filter, annotate, explore and analyze variant calls in a GUI via a standard web browser. It provides a unified and intuitive interface for all loaded resources and can be used by anyone without computer programming literacy. It has already been used for processing of more than 5000 clinical rare disease samples in the Stockholm healthcare region, covering a wide range of genetic disorders and providing specific molecular diagnoses for several hundred patients.

## PgmNr 1571: An analytical pipeline for decreasing the burden of returning secondary findings.

### Authors:

E. Reble<sup>1</sup>; M. Gutierrez-Salazar<sup>3</sup>; K.R. Zakoor<sup>3</sup>; M. Clausen<sup>1</sup>; R. Kodida<sup>1</sup>; S. Shickh<sup>1,2</sup>; C. Mighton<sup>1,2</sup>; R.H. Kim<sup>3,4,6</sup>; K.A. Schrader<sup>5</sup>; I. Cohn<sup>6</sup>; J. Lerner-Ellis<sup>2,3</sup>; Y. Bombard<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) St. Michael's Hospital, Toronto, ON, Canada; 2) University of Toronto, Toronto, ON, Canada; 3) Mount Sinai Hospital, Sinai Health System, Toronto, ON, Canada; 4) University Health Network, Toronto, ON, Canada; 5) BC Cancer Agency, Vancouver, BC, Canada; 6) The Hospital for Sick Children, Toronto, ON, Canada

---

**Background:** Advances in genomic sequencing technology have led to the use of exome and genome sequencing instead of panel testing, which has increased the potential to identify secondary findings (SFs). There is a need for new bioinformatic pipelines tailored to filter and analyze the large amounts of data associated with SFs beyond just medically actionable results in order to address guidelines for the disclosure of clinically relevant SFs.

**Aim:** We aimed to develop an analytical pipeline for the efficient analysis and return of SFs.

**Methods:** We expanded on and adapted Berg's 2011 binning framework for the analysis and return of SFs. Our pipeline consists of comprehensive gene lists for five categories ('bins') of SFs and incorporates filtration parameters for prioritization of variants in each bin. Gene lists were curated following evidence-based guidelines, established criteria and based on consultations with genetics experts. Our variant prioritization framework was developed by integrating disease-associations, population frequency, gene-variant databases and ACMG variant classification criteria. We applied the pipeline to 42 exomes to assess the feasibility and efficiency of the analysis and return of SFs.

**Results:** The five categories ('bins') of SF genes consisted of: (1) 90 genes associated with medically actionable diseases and 29 pharmacogenomics variants, (2) 17 common disease risk variants, (3) 3,854 genes associated with Mendelian diseases, (4) 7 genes associated with early-onset neurological disorders, and (5) 688 genes associated with carrier status. Analysis of the 42 exomes using our binning and variant prioritization framework resulted in a decrease of >98% of variants compared to the raw analysis, yielding a manageable number of variants for manual review and allowing for decreased cost and turnaround time for the return of SFs.

**Conclusions:** Our analytical pipeline increases the feasibility of analyzing and returning SFs in clinical practice and represents a flexible approach to accommodate alternate workflows, version updates and new guidelines for reporting SFs. Our bioinformatics pipeline for the analysis of SFs represents an effective and efficient step towards streamlining the return of SFs to realize their public health benefits.

## PgmNr 1572: From NGS data to diagnosis: Fast HPO-based variant filtering of NGS data using Moon by Diploid.

### Authors:

F. Sleutels <sup>1</sup>; H. Bruggenwirth <sup>1</sup>; L. Hoefsloot <sup>1</sup>; C. Kint <sup>2</sup>; P. Schols <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Genetics, Erasmus medical center, Rotterdam, The Netherlands; 2) Diploid, Leuven, Belgium

Panel-based filtering of variants from Whole Exome Sequencing (WES) has been our standard for years. Such gene-driven analysis consists of a collection of genes ("the panel"), with known (or evidence-based) involvement in the clinical presentation. Panel-based filtering poses challenges when a mix of clinic is presented that does not easily match a specific panel or when the number of variants to consider grows (e.g. for large gene panels, WGS, diverse ethnic backgrounds and consanguinity).

The transition to a phenotype-based analysis is driven by our aspiration to increase throughput and accelerate and standardize our NGS powered genetic diagnoses. Essential tools that nurse this transition are the Human Phenotype Ontology project (HPO) and up-to-date genotype-phenotype databases (e.g. OMIM). Here we present our experiences to transit to a fast phenotype-driven NGS analysis using Moon by Diploid.

### Methods:

For retrospective analysis, patients were selected with a challenging clinical presentation, genotype or genetic diagnosis. 13 WES samples (5 singletons, 7 trios, 1 duo) and 2 WGS samples. For prospective analysis, 11 patients were enrolled (9 singletons and 3 trios), for which a panel-based analysis was requested. NGS data (Illumina) from WGS or WES (Agilent Sureselect captured) was variant called (GATK haplotype caller) yielding VCF files which were processed with our default analysis flow (Agilent Bench Lab NGS) and with Moon (Diploid).

### Results and Discussion:

Our gene-driven analysis flow is compared to the phenotype-driven approach and advantages and challenges are discussed. Advantages of the phenotype-driven analysis are: No need to administrate and update gene panels. Secondly, the standardized and objective data analysis shortlists the clinically most relevant variants. Moon links to Mastermind (Genomenon) to show the scientific literature for the variant at a glance. And finally, when new variants get shortlisted, users gets an attention warning.

Challenges that we encountered: For some oncogenetic traits, variants are not shortlisted because they are considered a risk factor with incomplete penetrance (e.g ATM). Defining and reporting risk factors is actively debated in the field. Finally, discovery of new disease-contributing genes is not catered for, but for most clinical genetics labs this is out of the scope of routine advise.

Future perfections of our NGS analyses: Automate reanalysis and automate intake of HPO-clinic using

Moons API.

## PgmNr 1573: Enhanced transcriptome-wide analytic framework identifies novel associations with metabolic and virologic traits in HIV-positive adults.

### Authors:

B. Li <sup>1</sup>; Y. Veturi <sup>2</sup>; Y. Bradford <sup>2</sup>; S. Verma <sup>2</sup>; A. Verma <sup>2</sup>; D.W. Haas <sup>3,4</sup>; M.D. Ritchie <sup>2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomics and Computational Biology, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA; 3) Departments of Medicine, Pharmacology, Pathology, Microbiology & Immunology, Vanderbilt University School of Medicine, Nashville, TN, USA; 4) Department of Internal Medicine, Meharry Medical College, Nashville, TN, USA; 5) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA

---

In studies to identify genes and biological pathways that underlie complex traits, it may be difficult to relate noncoding genetic variants with downstream genes that are affected. Various methods have been developed to overcome this difficulty. One such method is the transcriptome-wide association study (TWAS), which aggregates transcriptional regulatory effects of cis coding and noncoding variants to identify genes whose genetically-regulated transcriptional changes are associated with traits of interest. Available TWAS methods vary regarding underlying biological assumptions about tissue specificity of transcriptional regulatory mechanisms. This affected whether TWAS methods tend to identify tissue-specific or multi-tissue associations in our previous study. Thus, we designed a simulation analysis to examine how the interplay between selected TWAS methods and tissue specificity of genes affects power and type I error rate in gene prioritization. We found that eQTLs with higher prediction accuracy improved TWAS power. Importantly, single-tissue TWAS (i.e., PrediXcan) had greater power in prioritizing single tissue-expressed genes. But it was likely to prioritize false positive tissues for genes that are expressed in multiple tissues. For multiple tissue-expressed genes, cross-tissue TWAS (i.e., UTMOST) had overall equal or better power and controlled type I error rate. Based on simulation results, we applied an enhanced TWAS analytic framework to baseline (i.e. pre-treatment) laboratory values from AIDS Clinical Trial Group (ACTG) studies 384, A5095, A5142, A5202 and A5257. We identified several novel associations, including *KCND3* (which encodes subunits of voltage-gated rapidly inactivating potassium channels) with plasma log<sub>10</sub> HIV-1 RNA ( $p =$ ). *KCND3* has also been previously associated with immunologic response to vaccine. We identified novel pleiotropic genes, including *CWC22* (which encodes CWC22 spliceosome-associated protein homolog) with cholesterol, glucose, high-density lipoprotein cholesterol, and total bilirubin. Overall, when causal tissue is known, single-tissue TWAS will suffice. Otherwise, our simulation suggested using single-tissue TWAS for single tissue-expressed genes, and cross-tissue TWAS for multiple tissue-expressed genes, separately. This enhanced TWAS analytic framework which considers tissue specificity of gene expression, improves power and controlled type I rate for TWAS gene prioritization.

## PgmNr 1574: Pattern and trend of human genomic variations in the form of SNPs INDELs.

**Authors:**

P. Liang; V.K. Chundi

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dept Biological Sci, Brock Univ, St. Catharines, Ontario, Canada

---

Single nucleotide polymorphisms (SNPs) and small insertions/deletions (INDELs) are the most common genetic variations in the human genome. They have been shown to associate with phenotype variation including genetic disease. Based on data in a recent version of the NCBI dbSNP database (Build 150), there are 305,651,992 SNPs and 19,177,943 INDELs, and together as all small sequence variants, they represent approximately 11% of the human reference genome sequences. In this study, we aimed first to examine the characteristics of SNPs and INDELs based on their location and variation type. We then identified the ancestral alleles for these variants and examined the patterns of variation from the ancestral state. Our results show that the frequency of small variants averages at 104 SNPs/kb and 6.5 INDELs/kb for a total of ~11% of the genome. Chromosome 16 and 21 represent the least and most conserved autosomes, respectively, while the sex chromosomes are shown to have a much lower density of SNPs and INDELs being more than 30% lower in the X chromosome and more than 85% lower in the Y chromosome. By gene context, these variations are biased towards genic regions, and further, they are biased towards protein-coding genes and intron regions within the genic regions. Within the coding regions, SNPs and INDELs are biased towards missense and frameshift variations, respectively. Further, genes with the highest level of variation showed enrichment in functions related to environmental sensing and immune responses, while those with least variation associate with critical processes such as mRNA splicing and processing. Through a comparative genomics approach, we determined the ancestral state for most of these variants and our results indicate that ~0.79% of the genome has been subject to SNP and INDEL variation since the last common human ancestor. Our study represents the first comprehensive data analysis of human variation in SNPs and INDELs and the determination of their ancestral state, providing useful resources for human genetics study and new insights into human evolution.

# PgmNr 1575: Leveraging large datasets accumulated through population carrier screening to inform variant classification.

**Authors:**

D.H. Tran; S. Cox; K.E. Kasenit; E. Collins; M. Meredith; E. Zmuda

[View Session](#) [Add to Schedule](#)

**Affiliation:** Myriad Women's Health, South San Francisco, California.

---

**Introduction**

Sequencing-based genetic testing requires real-time interpretation of variants encountered in patients. We have performed expanded carrier screening of recessive Mendelian disease genes for more than 1,000,000 samples, detecting more than 1,400,000 distinct variant alleles. Our variant database includes a substantial number of curated case-studies garnered from the literature along with allele frequencies in our sample cohort. Here we demonstrate how a centralized database of evidence enables accurate variant interpretation at scale and improves variant classification methodologies.

**Methods**

We collated data collected from more than 40,000 manual reviews of literature and database evidence accumulated on a variant-by-variant basis. To accurately estimate the proportion of disease alleles represented by a given variant, we calculated the total number of disease alleles from our curated literature database. We filtered for studies with greater than 20 cases with a valid Pubmed ID, resulting in 2672 studies encompassing 267 genes. In order to identify and remove potential double-counting of disease alleles, we assessed whether study cohorts were reported in multiple references.

**Results**

Focusing on 154 primarily single-gene diseases, we established an estimate of the total disease allele counts given our filtering criteria. Our estimates range from 32 disease alleles for HYS1-associated Hydrolethalus syndrome to 83,436 alleles for MEFV-associated familial Mediterranean fever. 26% of case studies (426/1639 studies) with overlapping authors were confirmed to have shared study cohorts; our curation SOP guards against overcounting these cohorts. We additionally report analyses that leverage high-resolution variant frequencies and variant co-occurrence information obtained in our CLIA sequencing lab.

**Conclusions**

We have described a database of literature cases, allele frequencies, and variant co-occurrence that was obtained in the course of population-scale expanded carrier screening. We are now using this resource to supplement variant classification protocols and improve assessment of variant pathogenicity.



## PgmNr 1576: Broad Institute's genomic platform portfolio leads support the research community.

**Authors:**

T. Bowers; C. Cibulskis; J. Abreu; A. Hollinger; C. Walsh; M. Costello; N. Lennon

[View Session](#) [Add to Schedule](#)

**Affiliation:** Broad Institute, Cambridge, MA.

---

In 2017, the Genomics Platform (GP) at the Broad Institute created a new set of roles (Portfolio Leads) to better support both the germline and somatic research communities. The creation of these roles have allowed the platform to roll out new products and improve existing products with a focus on the specific features required to serve the scientific questions our communities want to address. One of the key drivers for germline portfolio is Broad's Medical and Populations Genetics (MPG) group. When researching the requirements for the new exome, MPG investigators expressed a need for better mitochondrial coverage. After working closely with MPG, our R&D team, and TWIST Biosciences, a 100 fold increase in the mitochondrial genome coverage was achieved while still maintaining abundant and even coverage across the rest of the custom exome design. Another offering, portfolio has been instrumental in developing are the new single-cell RNA-seq products. A collaboration with Aviv Regev's lab, has successfully scaled the SmartSeq2M protocol with full automation from library construction to sequencing. The Platform now has the capacity to library construct and sequence 16 plates a week. Thus, allowing more researchers access to sequence either single-cell or populations for their full-length transcript capture methods. A suite of long read sequencing products are being developed. These aim to provide improved structural variation calling in human whole genome sequencing. The GP was an early access site for PacBio's Sequel II instrument with higher yielding 8M SMRT cells and longer run times. In GP's hands, the Sequel II has delivered raw average read lengths of ~50 kb with 50% of reads being >140 kb. These new offerings will continue to enable the science of our research community.

## PgmNr 1577: The Federated European Genome-phenome Archive.

### Authors:

T.M. Keane <sup>1</sup>; D. Spalding <sup>1</sup>; P. Flicek <sup>1</sup>; J. Rambla <sup>2</sup>; A. Navarro <sup>2,3</sup>; R. Guigo <sup>2</sup>; H. Parkinson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 2) Centre for Genomic Regulation (CRG), Barcelona, Spain; 3) Universitat Pompeu Fabra (UPF), Barcelona, Spain

---

The European Genome-phenome Archive (EGA), a joint project between EMBL-EBI and The Centre for Genome Regulation, Barcelona (CRG), has been a public resource since 2008 for permanent secure archiving and sharing of all types of potentially identifiable genetic and phenotypic data resulting from biomedical research projects. Its aim is to provide access to data, to foster data re-use, to enable reproducibility and to speed up biomedical and translational research in line with the 'FAIR' principles (Findable, Accessible, Interoperable, and Reusable). Access to EGA data must be approved by a Data Access Committee (DAC) and data must be appropriately consented for sharing. We have recently seen the emergence of genomic data from large cohorts of human samples from both national and regional healthcare in addition to traditional research projects. Many countries now have nascent personalised medicine programmes meaning that human genomics is undergoing a step change from being a predominantly research-driven activity to one funded through healthcare. We envisage that a significant subset of this data will be made available for secondary research (e.g., Genomics England). However, genetic data generated in a healthcare context is subject to national laws and health data from one country may not be exportable outside regional or national jurisdictions. In coordination with ELIXIR and the GA4GH, the EGA is committed to the establishment of a federated and interoperable network of human data resources to ensure continued sharing of human genetic data for research.

We are establishing a federated EGA node network consisting of three distinct node types: Central EGA, Federated EGA nodes, and EGA Community nodes. Broadly, Central EGA nodes (currently EMBL-EBI and CRG) will continue to accept international data submissions, provide international data distribution, and helpdesk support. Federated EGA nodes will offer submission and archiving services for submitters within a particular jurisdiction to accommodate use-cases where data must remain within a jurisdiction, and provide international data access and helpdesk for requestors of such data. EGA Community nodes will not offer external data submission services and rather act as data distribution nodes for datasets. In our presentation, we will provide a more detailed description of the network and nodes, initial use-cases and partners, and an update on the project implementation.

## PgmNr 1578: Expanding the scientific scope of the GWAS Catalog to include targeted and exome array studies.

### Authors:

A. Buniello <sup>1,2</sup>; M. Cerezo <sup>1</sup>; M. Ghoussaini <sup>2,3</sup>; P. Hall <sup>4</sup>; L.W. Harris <sup>1</sup>; J. Hayhurst <sup>1,2</sup>; E. Lewis <sup>1</sup>; A. McMahon <sup>1</sup>; J. Morales <sup>1</sup>; E. Sollis <sup>1</sup>; D. Suveges <sup>1</sup>; P.F. Whetzel <sup>1</sup>; T. Burdett <sup>1</sup>; L. Hindorff <sup>4</sup>; I. Dunham <sup>1,2</sup>; P. Flicek <sup>1</sup>; F. Cunningham <sup>1</sup>; H. Parkinson <sup>1</sup>; JAL. MacArthur <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridge CB101SD, UK; 2) Open Targets, Wellcome Genome Campus, Hinxton, Cambridge CB101SD, UK; 3) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB101SD, UK; 4) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

---

The GWAS Catalog delivers a publicly available and manually curated collection of all published genome-wide association studies. It currently contains highly searchable results from approximately 7,000 GWAS and 4,000 publications. To increase the potential of the Catalog for downstream analyses and future research needs, we have recently extended its scientific scope.

New content includes a set of studies (prioritized by trait) using targeted genotyping arrays, such as the Metabochip, Immunochip, Oncoarray and Exome arrays. So far, a total of 150 targeted array studies and over 5,000 associations have been extracted and made available through the Catalog from 67 publications. In addition to curated data and metadata, full p-value summary statistics from 47 targeted and exome array studies have been made available through the Catalog. This expansion creates a more comprehensive and representative Catalog, particularly for immunologic, metabolic and oncologic phenotypes. For example, the most significant result reported in chronic inflammatory disease is a unique SNP-trait association in NOD2 (Ellinghaus, 2016) that would have been missed without the inclusion of targeted array data.

Curation of these studies is performed following our consistent curation procedure, including ancestry annotation and mapping of all traits to Experimental Factor Ontology (EFO) terms to support richer querying and integration of trait information. Targeted and exome array studies are identifiable in the search interface and in study pages by a “target” icon in the Catalog’s new web interface. The data download format and the RESTful API have been updated to include information on the targeted array type and new documentation is now available supporting this functionality.

We are working towards identifying and curating all published targeted and exome array studies. So far, through text-based searching, we have identified 250 additional publications based on targeted or exome array analysis from 2012 to 2019. To ensure systematic capture of all relevant publications, we are collaborating with PubTator (<https://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/PubTator/>) to develop a machine-learning based literature search.

These new data underpin research into common disease, enabling investigations to identify causal variants, understand disease mechanisms, and establish targets for novel therapies.

## PgmNr 1579: Increasing the availability of summary statistics in the GWAS Catalog.

### Authors:

JAL. MacArthur <sup>1</sup>; M. Cerezo <sup>1</sup>; A. Buniello <sup>1,2</sup>; J. Hayhurst <sup>1,2</sup>; P. Hall <sup>3</sup>; L.W. Harris <sup>1</sup>; L.A. Hindorff <sup>3</sup>; S. Jupp <sup>1</sup>; E. Lewis <sup>1</sup>; C. Malangone <sup>1</sup>; A. McMahon <sup>1</sup>; J. Morales <sup>1</sup>; E. Mountjoy <sup>2,4</sup>; E. Sollis <sup>1</sup>; D. Suveges <sup>1</sup>; O. Vrousseau <sup>1,2</sup>; P.L. Whetzel <sup>1</sup>; T. Burdett <sup>1</sup>; F. Cunningham <sup>1</sup>; P. Flicek <sup>1</sup>; H. Parkinson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 2) Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, UK; 3) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK

---

Recently many new methods have been developed which leverage summary statistics (SS) from genome-wide association studies (GWAS) to provide insights into the mechanisms of complex disease and predict disease risk. Having extended scope to meet this need, the NHGRI-EBI GWAS Catalog ([www.ebi.ac.uk/gwas](http://www.ebi.ac.uk/gwas)) contains over 6,000 full p-value SS datasets, making it the largest repository of GWAS SS. SS are available for 11% of all Catalog studies. The numbers are higher for recent studies, with 25% for studies published in 2018 and 32% for studies published in the first 3 months of 2019. SS are highly accessed, with over 13,000 SS downloads in the last 6 months. The most downloaded SS by trait are for hematological measurements, immune system disorders and metabolic disorders. By year the most highly accessed SS dataset is from 2016, indicating the importance of making available SS from older studies. Access to SS is provided from the GWAS Catalog FTP site ([www.ebi.ac.uk/gwas/downloads/summary-statistics](http://www.ebi.ac.uk/gwas/downloads/summary-statistics)), via API from a SS datastore ([www.ebi.ac.uk/gwas/summary-statistics/docs](http://www.ebi.ac.uk/gwas/summary-statistics/docs)) and via links within our query interface ([www.ebi.ac.uk/gwas/search](http://www.ebi.ac.uk/gwas/search)).

To ensure SS are consistent across studies and interoperable they are pre-processed and made available in a standard format. They are also harmonised, with respect to genome build and strand. We propose the GWAS Catalog format (Buniello et al., 2019; <https://www.ebi.ac.uk/gwas/docs/methods/summary-statistics>) as a standard format for reporting SS. We are seeking feedback from the community to ensure the standard format includes all data elements required for common analyses and is applicable to the full range of GWAS study designs and genotyping methods.

Despite the recent increase in publicly available SS through the GWAS Catalog, the majority of SS are not public, evidenced by >3,000 publications in the GWAS Catalog for which the SS are not freely available. Therefore, we are encouraging authors to submit their SS directly to the GWAS Catalog and have seen the rate of author submissions increase 3-fold over the last 6 months. To facilitate sharing and further increase the availability of SS we are developing a web-based author deposition interface.

The GWAS Catalog continues to engage with the community towards defining standards and incentives to promote and enable sharing of GWAS data. The broad availability of SS will vastly

extend the potential of GWAS for the needs of future research to improve disease understanding.

## PgmNr 1580: Genomics and Biobanking in African settings: The strides made and status quo.

### Authors:

D. Kezimbira; I. Sserwadda; S. Kyobe; E. Kigozi; F. Ashaba; H. Musinguzi; G. Mboowa; E. Katgirya; S. Mwesigwa; N. Lwanga; R. Kamulegeya

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Immunology and Molecular Biology, Makerere University, College of Health Sciences kampala, Upper Mulago Hill Road, Uganda ,

---

### Introduction

Genomics wields the power to define and revolutionize modern day precision medicine due to its high-resolution power applied in the diagnosis and treatment of infectious diseases compared to the empirical methods especially in the research world. Genomics has witnessed a rampant evolution in the developed world as a consequence of the Bio-Banks and Biorepositories. Fully fledged Bio-banks are very critical and a one-stop shop for preserving, processing, accession and long-term storage of all kinds of biological specimens; some initiatives have been introduced in some resource-limited countries in terms of infrastructure and capacity building, especially in Africa like the H3Africa and B3Africa. However, these efforts have been limited to very few African countries, leaving the majority unaccounted for.

### Methods

We applied the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guiding principle to examine the strides that have been taken to introduce Genomics and Bio-Banks in Africa using the PubMed NCBI database.

### Results

We discovered very limited evidence of efforts and Bio-Banking infrastructure to support the fast-evolving Genomics discipline in African settings.

### Conclusion

We propose various approaches and models that when adopted will realize the increased outreach of Bio-Banks, Biorepositories and Genomics, consequently checking on the burden of infectious diseases on the African continent.

## **PgmNr 1581: Compound heterozygotes with oculocutaneous albinism as a special case of outbreeding depression in an admixed Colombian population.**

### **Authors:**

W.A. Cardenas <sup>1</sup>; A.B. Conley <sup>3</sup>; I.K. Jordan <sup>3,4</sup>; M.C. Lattig <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Servicios Integrados de Genética, Universidad de los Andes, Bogotá, COL; 2) Servicios Integrados de Genética, Alianza Universidad de los Andes y Fundación Santa Fe de Bogotá, Bogotá, COL.; 3) Applied Bioinformatics Laboratory, Atlanta, GA; 4) School of Biological Science, Georgia Institute of Technology, Atlanta, GA

---

Autosomal recessive conditions are frequently associated with homozygous mutations within a common ancestral origin. However, an increasing number of compound heterozygotes have been documented, especially in admixed populations, most likely due to the elevated mating among populations that are occurring due to the ease of modern transportation. Oculocutaneous albinism type 1 and type 2 (OCA 1 and OCA2) are the most severe forms of albinism caused by mutations in TYR or OCA2 genes respectively. We evaluated the ancestral origin of mutations in these two genes in a sample of 32 individuals with either OCA1 or OCA2 previously described by our group. Haplotype analysis in regions containing these two genes was performed to determine the ancestral origins - European, African, Native American or Jewish. Genotyping was performed using the 24 OmniExpress Chip (Illumina) and Macrogen service (Seoul, Republic of Korea). Data was merged with the reference data panel from the 1000 genomes and Sephardic Jews using PLINK. Haplotype inference from genotype and local ancestry was performed using ShapIT2 and RFMix. MEGA was used to compare the probable haplotype obtained for the TYR and OCA2 genes in our samples with the haplotype in the reference's samples and the results were plotted using R software. Various Spanish haplotypes were found both in the TYR and OCA2 genes. Of special interest was a compound heterozygote individual that carries the G47D mutation in a Spanish haplotype and the 1379delTT deletion in an Amerindian haplotype exemplifying a case of outbreeding depression in admixed population, whereby chance two random individuals living in the same population and carriers of different mutations in the TYR gene mate and produce a child with oculocutaneous albinism.

## PgmNr 1582: The portrait of fully phased assembled diploid human genome.

**Authors:**

A. Fungtammasan; N. Thangaraj; B. Hannigan; N. Hill; C.S. Chin

[View Session](#) [Add to Schedule](#)

**Affiliation:** DNAnexus, MountainView, CA.

---

Personalized de novo genome assembly offers potential benefits in discovering complex structural variants and unique sequences with less reference bias as compared to standard whole-genome resequencing. However, the challenge of variant calling from the assembly-based method is the heterozygous regions, which obscures the variant calls from the unphased assembled genome. Besides calling variants, genomic phasing is also useful in putting the genetic variation into context. This is critical to study allele-specific regulation, compound heterozygous mutations, or population genomics. For these reasons, the de novo assembly and phasing have become increasingly important in the study of personalized genomics. To date, the phasing of the assembled human genome is challenging due to the low heterozygosity of the human population.

Here we present the fully phased diploid human assembly from single molecule sequencing data of Oxford Nanopore Technology (ONT) using a trio-binning assembly approach. The ultra-long ONT data of the HG002 individual was provided by Genome In a Bottle Consortium (GIAB) and UCSC. These raw long reads were fully segregated into paternal and maternal reads using parental specific k-mer sequences from Illumina data of the parents. Then, the paternal and maternal raw long read data were successfully assembled into two 3 Gbp genomes and polished to improve their overall accuracy. Consequently, we explore the benefits of fully phased assembled genomes in the context of variant discovery. We compare between diploid-assembly-based variant calling and mapping-based variant calling against GRCh38 using the GIAB truth set as the ground truth. The fully phased assembled human genome is publicly available on the DNAnexus platform and the GIAB Consortium ftp site.



## PgmNr 1583: Assembling scraps: Deriving host genotype and dietary phenotypes from metagenomic datasets.

### Authors:

S. Eckert <sup>1,2</sup>; F. Chen <sup>2</sup>; D.J. Liu <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics and Genomics at the Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA.; 2) Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA

---

There is much interest in the human microbiome due to its association with various disease, including, but not limited to type 2 diabetes, ulcerative colitis, and colorectal cancer. A common concern that is raised about the microbiome is that it is highly variable, which has led to a lack of reproducibility of some findings. Microbiome composition has been found to be heritable at levels ranging from 10% to 40%, however, many identified genetic associations have not been replicated. This may be due in part to insufficient sample size and lack of control between studies. From off-target reads in whole-genome shotgun (WGS) sequencing of the microbiome, we are capable of retrieving human genotypes. Recently, it has been shown that chip-quality genotype can be imputed from low-coverage ( $> 0.1x$ ) alignment using standard genotype imputation software, such as BEAGLE. We leveraged this knowledge and developed a robust pipeline to generate a large cohort with both host genotype and microbiome. From the Human Microbiome Project (HMP) as well as several other studies, we aligned metagenomic WGS reads against the human genome, filtered for only human reads, called variants, and imputed genotypes using 1000 Genomes Project data as a reference. From off-target reads, we generated, on average, a 6.8x coverage alignment. From our alignment, we called and imputed over 15 million common genetic variants. To validate our imputed genotypes, we assessed concordance between our imputed genotypes and the array genotype for the subset of the HMP for which genotype array data was available. Using these datasets, we will conduct host genetics x microbiome interaction analyses, which will help elucidate the genetic architecture for microbiome composition and diversity.

## **PgmNr 1584: The Alliance of Genome Resources: A next-generation comparative genomics data ecosystem for functionalizing the human genome.**

### **Authors:**

C. Bult <sup>1</sup>; P. Sternberg <sup>2</sup>; J.M. Cherry <sup>3</sup>; J. Blake <sup>1</sup>; M. Westerfield <sup>4</sup>; P. Thomas <sup>5</sup>; B. Calvi <sup>6</sup>; K. Howe <sup>7</sup>; C. Mungall <sup>8</sup>; M. Shimoyama <sup>9</sup>; N. Perrimon <sup>10</sup>; The Alliance of Genome Resources

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) The Jackson Laboratory, Bar Harbor, ME; 2) CalTech, Pasadena, CA; 3) Stanford University, Palo Alto, CA; 4) University of Oregon, Eugene, OR; 5) University of Southern California, Los Angeles, CA; 6) Indiana University, Bloomington, IN; 7) European Bioinformatics Institute, Hinxton, UK; 8) Lawrence Berkeley National Laboratory, Berkeley, CA; 9) Medical College of Wisconsin, Madison, WI; 10) Harvard University, Boston, MA

---

For decades, Model Organism Databases (MODs) and the Gene Ontology Consortium (GOC) have expedited discoveries about gene function and biological principles of human biology through expert curation and integration of heterogeneous genetic, genomic, and phenotypic data from disparate sources. Open access to the collected knowledge about a model organism and about gene function in human readable and computation-ready formats as provided by these resources would otherwise require hundreds of hours for an individual researcher to assemble on their own. Despite their demonstrated utility as core community data resources, the lack of unified data access mechanisms and user interfaces--even for data types shared in common among model organisms--have prevented these resources from realizing their full potential for functionalizing the human genome. To address this usability gap, the GO Consortium and six MODs (Saccharomyces Genome Database, WormBase, FlyBase, Zebrafish Information Network, Mouse Genome Database, Rat Genome Database) have joined together in to form a consortium, the Alliance of Genome Resources (the Alliance). Through our web portal ([alliancegenome.org](http://alliancegenome.org)) with common user interfaces and our shared application programming interfaces (APIs), the Alliance has implemented an advanced comparative genomics platform making it easier than ever for researchers to perform multi-species comparisons of biological annotations for orthologs of human genes for such properties as gene function, expression, interaction, phenotype, and disease. The Alliance resource is designed to facilitate identification of relevant animal models of human disease and to evaluate the disease significance of human genes and variants of unknown significance identified from large scale clinical sequencing initiatives.

## PgmNr 1585: Mapping tolerance of intraspecific variation across the human and mouse genomes.

**Authors:**

G. Powell <sup>1,2</sup>; M. Simon <sup>1</sup>; A. Mallon <sup>1</sup>; C. Lindgren <sup>2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) MRC Harwell Institute, Mammalian Genetics Unit.; 2) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK

---

Regional differences in population-wide variation across the genome can be explained in part by constraint due to purifying selection, and provide a framework for interpreting the functional importance of genomic elements. Here, we use a sliding window approach to quantify intraspecific constraint across the human and mouse genomes. We highlight the relative enrichment and depletion of genomic annotations across the most and least constrained regions of the genomes, and show constraint is closely correlated between protein-coding regions and cis-regulatory elements for both species. Furthermore, we compare constraint across syntenic regions, and for pathogenic variants in human with an orthologous position in the mouse. Our results have implications for inferring when the mouse may be sufficiently analogous to serve as a model for human disease.

## PgmNr 1586: A model organism matchmaking software platform for clinical genetics research.

### Authors:

S. Rogic<sup>1,2</sup>; M. Jacobson<sup>1,2</sup>; S. Tesar<sup>1,2</sup>; M. Belmadani<sup>1,2</sup>; P. Hieter<sup>1</sup>; P. Pavlidis<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada; 2) Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada

---

Clinician-scientists who identify novel disease genes or variants often wish to validate their findings in a model system, but identification of an appropriate model organism (MO) collaborator can be a challenge. To this end, we developed a web-based registry system for MO researchers, originally implemented as the central resource for the Canadian Rare Diseases Models and Mechanisms (RDMM) Network. Because the concept and implementation has broader applicability, here we describe our software platform and how it can be used to establish collaboration between MO researchers and human geneticists. The software is open source and is available at <https://github.com/PavlidisLab/modinvreg>.

The purpose of the Registry is twofold: 1) to provide a straightforward way for MO researchers to describe their research and relevant expertise at the gene level; and 2) to allow comprehensive searches of the registrants' information. Besides their contact information, affiliation and a brief research description, the researchers enter genes for each MO they use. Genes are organized in three tiers, based on the researcher's expertise: TIER1 and TIER2 genes are directly entered, while TIER3 genes are computationally inferred using Gene Ontology terms.

The Registry can be searched using researcher's name, research description keywords, or genes in their profile. Gene search can either be MO specific or, if the query gene is human, the orthology mapping is done automatically and search performed for all the orthologs.

The Registry currently supports multiple user-settable levels of privacy of MO researcher information. The default data privacy settings, user interface and documentation are all customizable on the application level to suit instance-specific needs.

As of mid-2019, the software is being adopted by other rare disease research networks. To enable federated interaction between these sites, we implemented an application programming interface (API) that enables remote querying and fetching of data between different software instances. As a result, a simple checkbox selection on the search page will include the search results from other instances, according to appropriate viewing permissions. The API also allows public data to be accessed and displayed by other applications and resources.

In conclusion, the RDMM Registry is a secure, customizable, portable, and linkable gene-based matchmaking platform that should be easily adoptable for other similar research collaborations.

## PgmNr 1587: Identifying very rare genetic variants in DNA microarray data using a population-specific reference panel for imputation.

### Authors:

M. Alver<sup>1,2</sup>; T.T. Sikka<sup>1,2</sup>; M. Kals<sup>1,3</sup>; K. Pärn<sup>3</sup>; T. Nikopensius<sup>1</sup>; P. Palta<sup>1,3</sup>; A. Metspalu<sup>1,2</sup>; R. Mägi<sup>1</sup>; T. Esko<sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia; 2) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia; 3) Institute of Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 4) Broad Institute, Cambridge, MA

While the ascertainment of individuals affected by rare clinically actionable disorders requires next generation sequencing (NGS) for accurate calling of carrier status, availability of these resources is not always feasible for population-wide screening due to cost. Estonian population-based biobank, that holds DNA microarray-derived genotype information for 52,274 individuals, has built a population-specific reference panel from NGS data (n=2,244) for imputation. We therefore aimed to investigate whether the utilization of a custom reference panel for imputing missing genotypes enables the ascertainment of very rare genetic variant (VRGV) carriers from microarrays.

We set to determine carrier status for 20 VRGVs linked with clinically actionable disorders that were identified in heterozygous state in NGS (allele counts between 1-7, n=34) from Illumina Global Screening Array-genotyped data of 33,277 individuals. To this end, six different approaches were applied: long-range haplotyping as well as basic (EAGLE/BEAGLE), read-aware phasing-based (SHAPEIT2/ BEAGLE) and non-pre-phased (IMPUTE2) imputation. As phasing of rare alleles is largely random, we additionally used allele switching. For that, we flipped the phased alleles in the reference panel to the other haplotype, and then imputed the variants of interest in a 20Mb window via basic and read-aware phasing imputation using created dummy haplotypes. Across six methods, 1079 putative VRGV carriers were identified and Sanger sequenced for validation. Carrier status was confirmed for 187 individuals for 16 VRGVs, yielding true positive rate of 18% across all methods. While 163 of 188 (87%) true carriers were identified via basic imputation, allele switching and long-range haplotyping facilitated to uniquely identify carriers for two (n=15) and four (n=5) AC=1 variants, respectively, not identified with any other approach, confirming random phasing of singletons.

Despite modest yield, these results demonstrate that microarray-based ascertainment of risk haplotypes using a custom population-specific reference panel for imputation provides a cost-effective strategy for identifying putative carriers affected with clinically actionable disorders at a population scale, when NGS is unavailable. While basic imputation allows to identify the majority of those with risk alleles, modification approaches are required for detection of singleton variants.

## PgmNr 1588: Quantifying phenotype similarity for complex harmonized disease cohorts.

### Authors:

D.M. Taylor <sup>1,2</sup>; L. Becker <sup>1</sup>; L. Qian <sup>1</sup>; N. Vasilevsky <sup>3</sup>; S. Essaid <sup>3</sup>; M. Haendel <sup>3,4</sup>; A. Kelman <sup>5</sup>; N. Singh <sup>5</sup>; P. Killada <sup>5</sup>; M.C. Kim <sup>5</sup>; B. Furner <sup>6</sup>; L. Graglia <sup>6</sup>; S. Volchenbaum <sup>7</sup>; A. Heath <sup>5</sup>; A. Resnick <sup>5</sup>; Gabriella Miller Kids First Data Resource Center

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical and Health Informatics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Department of Pediatrics, Perelman School of Medicine, The University of Pennsylvania, Philadelphia PA, USA; 3) The Center for Data Driven Discovery in Biomedicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 4) Translational and Integrative Sciences Lab, Oregon Health & Science University, Portland, OR, USA; 5) Oregon Health & Science University, Portland, OR, USA; 6) Center for Research Informatics, The University of Chicago, Chicago, IL, USA; 7) Department of Pediatrics, Biological Sciences Division, The University of Chicago, Chicago, IL, USA

---

The NIH Common Fund's Gabriella Miller Kids First Pediatric Research Program (KF) is a collaborative initiative focused on providing large-scale clinically-annotated genomic data for pediatric cancer and structural birth defect cohorts, including trio germline whole genome sequencing (WGS) and tumor WGS and RNA-seq. The KF Kids First Data Resource Center (KFDRC; <https://kidsfirstdrc.org/>) is charged with empowering collaborative research and discovery through integration of KF cohorts and external data. As of mid-2019, there are 33 GMKF unique disease cohorts funded through KF: 23 structural birth defect (SBD) and 10 pediatric cancer (PC) cohorts, with one mixed SBD & PC cohort. Approximately 7300 participants from 2300 SBD and PC families have been harmonized into the KFDRC since the launch of the portal in June 2018, making the KFDRC one of the largest pediatric data resources of its kind.

Currently, over 23,000 individuals (probands and families) have been funded by KF for sequencing and harmonization, representing approximately 7000 individual probands, all of whom have disease-specific clinical and phenotype data that must be harmonized against common classification sets for cohort analysis intra- and inter-operability.

We show how the Kids First Data Resource is working to harmonize clinical and outcome data across phenotypes and diseases using the Mondo Disease Ontology for disease classification, the Human Phenotype Ontology (HPO) for phenotype classification, the Uber-anatomy ontology (Uberon) for anatomical classification, and the National Cancer Institute's Thesaurus (NCIT), which classifies across many different levels and types of data in the cancer domain. Notably, we present work on quantification of individual and cohort-level phenotype content and pairwise similarities using metrics such as the Jaccard index. Quantifications are benchmarked using cohorts with co-occurring structural birth defects and against randomly generated null sets of phenotypes. Quantification of phenotype interoperability is useful for identifying not only cohort groups that may be more analytically interoperable in genetic analyses, but will also help inform contributing investigators as to the quality and interoperability of their own clinical phenotypic datasets with the rest of KF cohorts.

## PgmNr 1589: Identification of clinically relevant polyadenylation signals.

**Authors:**

H. Shiferaw; C. Hong; J. Johnston; L. Biesecker

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Institutes of Health, National Human Genome Research Institute, Bethesda, Maryland, USA

---

Polyadenylation is essential in maintaining nascent mRNA stability. Variants in polyadenylation signals (PAS) can result in reduced polyadenylation at the normal polyA sites and lead to reduced gene expression. Only 26 variants in PAS in 20 genes have been associated with mendelian disorders. We hypothesize that this is under-representative of this class of mutations. Here, we aimed to comprehensively identify clinically important PAS. PolyA sites were collected from the Alternative Polyadenylation Database (APADB) and the Polyadenylation Database Version 3 (PADB3) and examined for dominant polyA site usage activity (defined by > 50% of overall EST representation). We filtered for inclusion of polyA sites with only one canonical AATAAA PAS within 50 bases upstream. We also filtered for >=20x coverage in 90% of gnomAD samples and removed low complexity regions. Remaining sites from the two databases were intersected to produce the final list of PAS of interest. To understand constraint in the identified PAS, we compared the number of variants in the PAS of interest vs other control regions using gnomAD. We identified 3,242 canonical (AATAAA) PAS for further examination. Of these PAS, 1,727 were in genes in HGMD. Five of these 1,727 genes were among the previously reported 20 genes with disease-associated PAS variants in HGMD. The 3,242 PAS were at least 6 times more constrained than AATAAA hexamers occurring in intergenic regions, hexamers immediately upstream of the signal, and AAT or AAA triplets occurring in upstream 3' UTR regions. Interestingly, we observed position-specific substitution bias in the middle 4 positions compared to positions 1 and 6. A base-specific bias was also seen in middle positions, which favored guanine and cytosine substitutions. We identified 3,242 PAS of potential clinical relevance. The high constraint observed in these PAS supports their biological importance and potential clinical implications. We hypothesize that these sites harbor pathogenic variants for mendelian disorders. We are currently screening for variants in these 3,242 PAS in a cohort for clinical follow-up and validation.

## PgmNr 1590: Pretrel: PRedicting and ETHnicity and RELatedness from genotyping data - an R Package.

### Authors:

S. Tuna <sup>1,2</sup>; C. Penkett <sup>1,2</sup>; W. Astle <sup>3</sup>; K. Stirrups <sup>1,2</sup>; W. Ouwehand <sup>1,2,4,5</sup>; E. Turro <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Haematology, University of Cambridge, Cambridge, United Kingdom; 2) NIHR BioResource, Cambridge University Hospitals, Cambridge, United Kingdom; 3) MRC Biostatistics Unit, Cambridge Institute of Public Health, Cambridge, United Kingdom; 4) Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom; 5) NHS Blood and Transplant, Cambridge Biomedical Campus, Cambridge, United Kingdom.

---

Prediction of ethnicity and relatedness from genomic markers is a crucial step in Quality Control (QC) analyses for any large scale sequencing and genotyping study. By identifying any mislabelled samples at an early stage in the project, the researcher can investigate any potential issues in their laboratory/informatics pipelines, such as possible sample swaps or duplications, and to ensure the correct samples are included in the final dataset. Furthermore, by using an unrelated set, bias in population genetics parameters can be removed from the mainstream analysis. This allows for example for accurate calculation of Allele Frequencies (AF) for a particular mutation.

In the NIHR Bioresources , we have developed an R package: pretrel that supports initial SNP selection from the overall dataset, then calculates ethnicity and relatedness between samples and finishes by identifying a maximum set of unrelated samples. pretrel can be run on genotyping data obtained from a wide variety of platforms including Illumina sequencing data and microarrays. The full analysis suite scales to large sample size, and has been run on the GEL Whole-Genome Sequencing (WGS) data (~50,000 samples), NIHR BioResource WGS data, targeted NGS data for rare disease panels and data from Axiom arrays.

As a first step pretrel can select a reliable set of common SNPs from your dataset. The analysis also uses genotypes from a reference project such as 1000 Genomes project to infer ethnicities and part of the selection involves finding SNPs intersecting both the project data and reference samples. VCF files from the project and reference samples are then filtered for this SNP set whilst checking for data completeness, before being merged to one VCF file with high quality genotypes. The main statistical method used for predicting ethnicities is Principal Component Analysis (PCA) where the principal components are calculated using the reference set and the samples of interest are then projected onto the reference set using the GENESIS R package. By using these projections, we further assign ethnicities for each sample based on their similarity to the reference samples. Pairwise relatedness is then calculated with the GENESIS package, which starts with the ethnicity PCA loadings and then calculates the IBD ratios and a kinship score.



# PgmNr 1591: RSAT var-tools: An accessible and flexible framework to predict the impact of regulatory variants on transcription factor binding.

## Authors:

A. Medina Rivera <sup>1</sup>; W. Santana-Garcia <sup>1,2</sup>; M. Rocha-Acevedo <sup>1</sup>; L. Ramirez-Navarro <sup>2</sup>; Y. Mbouamboua <sup>3</sup>; D. Thieffry <sup>2</sup>; M. Thomas-Chollier <sup>2</sup>; B. Contreras-Moreira <sup>4,5</sup>; J. van Helden <sup>6</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Laboratorio Internacional de Investigación sobre el Genoma Humano, Universidad Nacional Autónoma de México, Querétaro, Querétaro, México, México; 2) Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France; 3) Fondation Congolaise pour la Recherche Médicale, Brazzaville, Republic of Congo; 4) Estación Experimental de Aula Dei-CSIC, Zaragoza, Spain; 5) Fundación ARAID, Zaragoza, Spain; 6) Aix-Marseille Univ, INSERM UMR S 1090, Theory and Approaches of Genome Complexity (TAGC), F-13288 Marseille, France

Only 6.4% out of the over 96,332 single nucleotide polymorphism (SNPs), from over 3,412 studies with reported trait associations of  $p < 1 \times 10^{-5}$  that have been curated in the NHGRI-EBI GWAS[1], are located in protein coding sequences with effects on protein structure. Overall, it is more easy to predict the functional impact of genetic variation in protein coding genes compared to variation in non-coding sequences, which generally do not have so well defined features such as codons or mRNA splice sites.

Noteworthy, non-coding DNA contains the gene regulatory information necessary for spatial and temporal gene expression patterns. Specifically, gene regulatory regions function by interacting with transcription factors (TFs) that bind to short, highly redundant DNA motifs. SNPs in gene regulatory regions that TF DNA binding sites have been implicated as causal variants in GWAS studies. However, there are numerous challenges (e.g. measure the effect in TF binding affinity) to unravel the impact of genetic variations in gene regulatory regions.

Here we describe var-tools, a set of programs within the Regulatory Sequence Analysis software suite [2], and present four case studies exemplifying their usage and applications. In brief, var-tools facilitate i) obtaining variation information, ii) interconversion of variation file formats, iii) retrieval of sequences surrounding variants, and iv) predicting the impact of the variation on the binding affinity of a given TF, estimated by matching the sequence with the annotated binding motif. The tools are available through a web interface that enables analysis of 5 Metazoa and 10 plant genomes, and can be further used in command-line with locally installed genomic sequences. Users can input personal variation and motif collections, providing flexibility in the analysis. In addition, haplotype phase information can be provided in the VCF file to estimate the joint effect of variants in close proximity on the same TF-binding event. Overall, RSAT var-tools provide a comprehensive resource to experienced and nonexpert users (accessible through a web interface) to analyze regulatory variants in several organisms.

## References

Annalisa Buniello, et al. *Nucleic Acids Research*, 47: D1005–D1012, 2019.

Ngathi Thuy Nguyen, et al. *Nucleic Acids Research*, 46: W209–W214, 2018.

## PgmNr 1592: IKAP: Identifying K mAJor cell Population groups in single-cell RNA-seq analysis.

**Authors:**

Y. Chen; A. Suresh; K. Singh; F. Seifuddin; M. Pirooznia

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bioinformatics and Computational Biology core, NHLBI, NIH, Bethesda, MD.

---

**Background**

In single-cell RNA-seq analysis, clustering cells into groups and differentiating cell groups by marker genes are two separate steps for investigating cell identity. However, results in clustering greatly affect the ability to differentiate between cell groups. This interdependency often creates a bottleneck in the analysis pipeline in which researchers repeat these two steps many times by trying different clustering parameters to identify a set of cell groups that are more distinguishing and biologically relevant.

**Findings**

To accelerate this process, we develop IKAP – an algorithm identifying major cell groups that improves differentiating by systematically tuning parameters for clustering. We demonstrate that, without specifying any parameters, IKAP successfully identifies major cell types such as T cells, B cells, NK cells, and monocytes in two peripheral blood mononuclear cell (PBMC) datasets and recovers major cell types in a previously published mouse cortex dataset. In addition, those major cell groups identified by IKAP present more distinguishing marker genes compared with cell groups generated by different combinations of clustering parameters. Finally, we show that cell subtypes can be further identified by re-running IKAP within identified major cell types so that a cell ontology prototype can be automatically constructed using single-cell RNA-seq data.

**Conclusions**

IKAP can speed up single-cell RNA-seq analysis and cell type recognition by identifying major cell groups with distinguishing marker genes and can facilitate cell ontology curation using single-cell RNA-seq data.

# PgmNr 1593: A robust clustering algorithm for identifying cell types from single-cell RNA sequencing data.

## Authors:

Liang <sup>1,2</sup>; Y. Hu <sup>3</sup>; M. Ranganathan <sup>1,2</sup>; S. Ganesh <sup>1,2</sup>; C. Yan <sup>3</sup>; D. D'Souza <sup>1,2</sup>; K. Xu <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Psychiatry, Yale School of Medicine, New Haven, CT; 2) Connecticut Veteran Healthcare System, West Haven, CT; 3) Center for Biomedical Information and Information Technology, National Cancer Institute, Bethesda, MD

---

Single-cell RNA sequencing (scRNA-seq) is a robust approach to identify cell types, cell status, and to discover novel genes for human diseases. Despite many computational methods and tools have been reported to analyze scRNA-seq data, accurately cell clustering remains challenging, which may substantially impact on downstream analysis. The challenges of cell assignment include data high dimensionality and instable clustering. In this study, we proposed a generalized linear model (GLM)-based cell mapping approach with cell-type “marker” genes curated from the literature to assign cell clusters to cell types. We first selected a reference gene panel based on the known signature genes in a given cell type and created a dummy matrix for gene expression of each cell type. We tested the association between gene expression in each cell and each cell type using GLM. We assigned one cell to one cell type by a predefined cut-off value. Then, the proportion of each assigned cell type was calculated for each cluster. The dominant proportions are used to assign one cluster to one cell type. For the cells falling between the clusters, we estimated a probability of the cell for each cluster and assigned cell into a cluster based on its probability using method based on Fuzzy c-means clustering. Briefly, we used the within-cluster cells for hyperparameter (weighting exponent) tuning in Fuzzy c-means clustering and chose the hyperparameter with highest prediction accuracy. Then, the probability of between-cluster cell belongs to each cluster was estimated by Fuzzy c-means. Finally, the cell type for each cluster was confirmed manually by cell type marker gene expressions mapping on the t-distributed Stochastic Neighbor Embedding (tSNE) plot. To test the performance of our method, we profiled four samples on the 10X Genomics platform and yielded a total of 15,973 human peripheral blood mononuclear cells and 21,430 genes. Our approach unambiguously deconvoluted the 15,973 cells among 21 clusters into eight cell subtypes: CD4+ T-cells (34.6%), IL7RCD4+ T-cells (8.4%), CD8+ T-cells (17.4%), B cells (13.2%), natural killer cells (12.3%), CD14+ monocytes (10.0%), FCGR3A monocytes (3.9%), and dendritic cells (DC) (0.3%). Our results demonstrated that GLM-based cell clustering is a robust approach for identifying cell types in scRNA-seq analysis.

## PgmNr 1594: Optimal design of single-cell RNA sequencing for cell-type-specific eQTL analysis.

### Authors:

I. Mandric<sup>1</sup>; T. Schwarz<sup>6</sup>; A. Majumdar<sup>5</sup>; R. Perez<sup>8,9,10,11</sup>; M. Subramaniam<sup>7,8,9,10,11</sup>; J. Ye<sup>8,9,10,11</sup>; B. Pasaniuc<sup>2,4,5,6</sup>; E. Haplerin<sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, University of California Los Angeles, Los Angeles, CA, USA; 2) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 3) Department of Anesthesiology and Perioperative Medicine, University of California Los Angeles, Los Angeles, CA, USA; 4) Department of Computational Medicine, University of California Los Angeles, Los Angeles, CA, USA; 5) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 6) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA, USA; 7) Biological and Medical Informatics Graduate Program, University of California, San Francisco, San Francisco, California, USA; 8) Institute for Human Genetics (IHG), University of California, San Francisco, San Francisco, California, USA; 9) Institute for Computational Health Sciences, University of California, San Francisco, San Francisco, California, USA; 10) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, USA; 11) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, California, USA

---

One of the main limitations of single-cell RNA-Seq is its high cost which prohibits population-scale analyses that aim to connect population level variation (e.g., genetics and/or disease status) with single-cell transcriptomics. The most common population-scale analysis is eQTL mapping, which is a widely-used tool in functional genomics used to identify putative mechanisms underlying the connections between the genotypes and the phenotypes mediated through gene expression. Single-cell RNA-Seq enables cell-type-specific eQTL mapping, i.e. finding associations between genetic variants and cell-type-specific gene expression. Achieving high levels of statistical power in these studies requires a large sample size, a large number of cells, and a high-coverage sequencing. However, with a limited budget one can not increase all the three quantities simultaneously. We demonstrate that cell-type-specific gene expression can be accurately inferred with low-coverage single-cell RNA sequencing given enough cells and individuals. For example, we show on a single-cell dataset (Segerstolpe et al., 2016) that by aggregating reads across cells within a cell type, it is possible to achieve a high level of the average Pearson R<sup>2</sup> between the low-coverage estimates and the ground truth cell-type-specific gene expression. The power of a study with sample size N is approximately the same as the power of a study with sample size R<sup>2</sup> \* N (the effective sample size, ESS) and the true cell-type-specific expression. Thus, under a constant budget, using low-coverage sequencing can increase the sample size and the power of an eQTL study. We also show that even taking into account all related costs such as the library preparation ones, using low-coverage single-cell sequencing can considerably decrease the cost of a cell-type-specific eQTL study without sacrificing its power. For example, we show that ESS of 50 can be achieved by sequencing 56 individuals with 33000 reads per cell (2750 cells per individual) which costs \$50,000. With our proposed approach of using low coverage, the same ESS can be achieved by sequencing 96

individuals at 1500 reads per cell (2500 cells per individual) and the total costs \$25,000. We also provide a practical methodology on designing cell-type-specific eQTL studies which maximizes statistical power. Our results provide a clear pathway for the design of efficient cell-type-specific association studies that are scalable to large populations.

## PgmNr 1595: Diagnosing pathogenic splicing variants using Exomiser.

### Authors:

D. Danis<sup>1</sup>; L.C. Carmody<sup>1</sup>; A. Hegde<sup>1</sup>; J.O.B. Jacobsen<sup>2</sup>; G. Valentini<sup>3</sup>; D. Smedley<sup>2</sup>; P.N. Robinson<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Jackson Laboratory for Genomic Medicine, Farmington, 06032 CT, USA.; 2) Genomics England, Queen Mary University of London, Dawson Hall, Charterhouse Square, London EC1M 6BQ, UK.; 3) Anacleto Lab Department of Computer Science, University of Milan, Via Celoria 18, 20133 Milano, Italy.

---

Clinical whole-exome sequencing (WXS) and whole-genome sequencing (WGS) are transforming how rare Mendelian disorders are being diagnosed. Although diagnostic yield is still low, it has the potential to be improved with more comprehensive support for integrative phenotype-genotype analysis and with better classification of variants that alter pre-mRNA splicing. Exomiser is a tool for annotation, filtering, and phenotype-driven prioritization of variants identified in WXS or WGS. Here, we demonstrate a method for improved identification of pathogenic splicing variants and integration into Exomiser to improve diagnostic yield.

Extensive review of published cases was performed to curate a collection of pathogenic variants (n=611) that cause splicing alterations leading to Mendelian diseases. We focused on variants at positions other than the canonical donor and acceptor dinucleotide sequences, such as those present in splicing *cis*-regulatory elements. The clinical data of the probands was recorded using Human Phenotype Ontology terms. The control set of splicing-neutral variants (n=657) was derived from ClinVar benign variants and from published variants where the absence of missplicing was demonstrated experimentally. A separate validation set of both pathogenic and benign variants was created in the same way. We developed schemes for scoring REF and ALT alleles of variants in different positions with respect to splice sites (e.g. exonic variant, variant in canonical splice site, etc.). The scoring schemes were based on the information content of the nucleotide sequence. Raw output of the schemes was then scaled between 0 and 1 in a way that maximized separation between pathogenic and benign cases. To evaluate the performance of the scoring schemes in simulations, we spiked pathogenic splicing variants into a VCF file from an unaffected individual and then analyzed the VCF together with the patient's phenotype. After the analysis, we measured how often the gene with pathogenic splicing variant was placed in any given rank among the full set of prioritized genes.

We will present current results of simulations, and demonstrate how to use the splicing prioritizer within the Exomiser framework for phenotype-driven analysis of WXS/WGS experiments.

# PgmNr 1596: Analysis of sequencing error profile across different next-generation sequencing platforms.

**Authors:**

J. Bai; X. Su; S. Li; L. Wu

[View Session](#) [Add to Schedule](#)

**Affiliation:** BerryOncology, Corporation, Beijing, China

---

**Background**

Next generation sequencing has revolutionized genome wide research and provided unprecedented amounts of data. Quality analysis is essential since a large number of erroneous basecalls were produced during sequencing procedures. Currently, raw base quality scores are widely used for quality analysis of next generation sequencing data. However, base quality scores are differently preset across different sequencers. Alternatively, sequencing of standard DNA and comparative analysis can be conducted in pipeline evaluations, but the conclusions of which would largely depend on data analysis strategies.

**Methods**

Exome sequencing data of six specimens were generated from three different platforms, including Illumina HiSeq2000, XTen, and NovaSeq6000 (two for each platform). Reads generated from same template molecules, which are often marked and removed by most analysis pipelines, were identified and defined as PCR duplication clusters. According to the origination of PCR duplicates, we defined loci that discordant within duplication clusters as erroneous basecalls. Total number of sequencing bases will be introduced to estimate the error rate of the sequencing run and under each Q score given by the sequencer.

**Results**

The overall sequencing error rates were 0.0444%/0.0493%, 0.0833%/0.0768%, and 0.0294%/0.0284%, for HiSeq2000, XTen, and NovaSeq6000, respectively. As expected, the basecalls with higher Q scores have magnificently lower error rates. Furthermore, using the erroneous substitution type as well as the context bases of the error loci, we analyzed the 'erroneous signatures' by non-negative matrix factorization. Interestingly, each of the three sequencing platforms tends to have a specifically dominant signature. For NovaSeq6000, the base mis-calling were often identified as T>G transversion at NpTpK contexts. Sequencing errors from HiSeq2000 are characterized by overrepresented of C>A and T>G transversion. Erroneous basecalls on XTen platform, on the other hand, are not showing any strong bias.

**Conclusions**

The tool we developed, ngsErrPro, defines actual sequencing errors from discordant bases within PCR duplication clusters. The results showing that even under the same quality score, error rates are different across sequencing platforms. Moreover, predominant erroneous signatures were identified for different sequencing platforms.

## PgmNr 1597: Linking pathogenic variants to expected disease phenotypes in electronic health records using UMLS.

### Authors:

A.L. Blumenfeld; J. Staples; A. Yadav; C. Gonzaga-Jauregui; L. Habegger; M. Cantor; J.G. Reid; Geisinger-Regeneron DiscovEHR Collaboration

[View Session](#) [Add to Schedule](#)

**Affiliation:** Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY

---

Large scale next-generation sequencing of unascertained population cohorts paired with patients' electronic health records (EHRs) provides an excellent resource for the study of gene-disease associations, as demonstrated by the DiscovEHR and UK Biobank collaborations. In these types of studies, a common question arises: does an individual carrying a known or expected pathogenic variant in a disease-associated gene exhibit the predicted disease phenotype? Answering this question for each pathogenic variant is challenging because the process is manual and time-consuming, often requiring an expert to perform a literature search on each variant and a review of the carrier's EHR. To address this problem, we developed a method that partially automates the process of screening an individual's EHR for diagnosis codes that can be attributed to variants in many of the most well studied human genes associated with disease.

We used the National Library of Medicine's UMLS (Uniform Medical Language System) to identify mappings between human disease genes and International Classification of Diseases (ICD) diagnoses. The UMLS is a compendium of ~200 controlled vocabularies from the biomedical sciences (e.g., ICD-10-CM, SNOMED CT, HPO) where similar concepts from the source ontologies are mapped to a common UMLS concept while retaining the ontological relationships. We represented a subset of the UMLS as a graph database and used it to find ICD-code derived phenotypes that can be conceptually linked to a gene or variant. The result is a set of gene-phenotype mappings that are not readily available from any publicly available ontology. These mappings specify a conceptual pathway between genes and phenotypes to accelerate follow-up analysis and frequently include secondary phenotypes that might otherwise be overlooked.

We applied our method to ~300k predicted pathogenic variants (in ~2400 disease genes we mapped to ICD phenotypes using UMLS) identified in DiscovEHR (N=92k) exome sequencing participants. For each variant, we were able to screen for expected phenotypes in the variant carrier's EHR in just a few seconds. This process identified a predicted ICD code in a carrier's EHR in ~35% of cases, which is appropriate considering the prevalence of reported recessive disease genes, disease penetrance, and the incomplete nature of de-identified EHRs. Our method provided a first-pass analysis of predicted pathogenic variants, potentially saving significant manual work.



## PgmNr 1598: Systematic gene set analyses using MAGMA.

**Authors:**

T.A. Boltz <sup>1</sup>; L.M. Olde Loohuis <sup>2</sup>; A.P.S. Ori <sup>2</sup>; R.A. Ophoff <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, California; 2) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, Los Angeles, California

---

A major challenge in genetics today is to understand the biology that underlies regions of the genome identified by GWAS. An approach to tackling this problem has been the development of gene-set analysis tools such as MAGMA (Multi-marker Analysis of GenoMic Annotation). While MAGMA outperforms other similar methods, there are still questions concerning how variables such as gene-set size, SNP-heritability (SNP- $h^2$ ), the sample size of the GWAS, or even the method of phenotyping the trait can influence the results of gene-set analysis. We set out to characterize the behavior of MAGMA (using real rather than simulated datasets) when different types of traits or gene-sets are used in various combinations as inputs to the software. We find that there are differences in resulting effect sizes and degrees of significance depending on whether the gene-set was manually curated or computationally derived, and that specific categories of traits (namely metabolic) are consistently significant for computationally derived sets. Furthermore, we found that continuous traits tend to have a higher proportion of significant results than binary traits, regardless of the source of the gene-set. When considering sample size of the GWAS or the SNP- $h^2$  of the trait, we found no strong correlations with degree of significance. However, we observed a strong correlation ( $R^2 > 0.5$ ) between gene-set size and the magnitude of the regression coefficient between heritability and degree of significance, regardless of the source of gene-set. Similarly, gene-set size and degree of significance among the computationally derived sets were also found to be strongly correlated (Pearson correlations  $> 0.7$ ), though the same was not observed for manually curated sets. Thus, we find that factors such as gene-set size, gene-set type, and trait type should be taken into consideration when interpreting the results of MAGMA.

## PgmNr 1599: Scalable probabilistic PCA for large-scale genetic variation data.

### Authors:

A.M. Chiu <sup>1</sup>; A. Agarwal <sup>2</sup>; M. Le <sup>3</sup>; E. Halperin <sup>1,3,4,5,6</sup>; S. Sankararaman <sup>1,3,4,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, California.; 2) Department of Computer Science, Indian Institute of Technology, Delhi, India.; 3) Department of Computer Science, University of California Los Angeles, Los Angeles, California.; 4) Department of Human Genetics, University of California Los Angeles, Los Angeles, California.; 5) Department of Anesthesiology and Perioperative Medicine, University of California Los Angeles, Los Angeles, California.; 6) Department of Computational Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California.

---

Inference of population structure is a key step in population genetic analyses with applications that include controlling for confounding in genome-wide association studies. The naïve approach for principal component analysis (PCA), one of the most widely used methods for inferring populating structure, estimates principal components (PCs) by computing a full singular value decomposition, resulting in runtimes unsuitable for large data sets.

Several solutions have been proposed for the efficient computation of PCs. One approach taken by two recent scalable implementations (FastPCA and FlashPCA2) takes advantage of the fact that typical applications of PCA in genetics only require computing a small number of PCs. Another approach utilizes the parallel computation infrastructure of the cloud, an approach that is often cost-prohibitive. Furthermore, these implementations lack a full probabilistic model, making them challenging to extend to other settings.

In this work, we describe ProPCA, a scalable method to compute the top PCs on genotype data. While PCA treats the PCs and the PC scores as fixed parameters, ProPCA imposes a prior on the PC scores. This formulation leads to an iterative Expectation Maximization (EM) algorithm for computing the PCs. ProPCA leverages the structure of genotype data so that each iteration of the EM algorithm can be computed in time that scales sub-linear in the number of individuals or SNPs.

In both simulated and real data, ProPCA is able to accurately infer the top PCs while scaling favorably with increasing sample size. We applied ProPCA to compute the top five PCs on genotype data from the UK Biobank, consisting of 488,363 individuals and 146,671 SNPs, in less than thirty minutes. Leveraging the population structure inferred by ProPCA within the White British individuals in the UK Biobank, we scanned for SNPs that are not well-modeled by the top PCs to identify several novel genome-wide signals of recent positive selection. Our scan recovers sixteen loci that are highly differentiated across the top five PCs that are likely signals of recent selection. While these loci include previously reported targets of selection, the larger sample size that we analyze here allows us to identify eleven novel signals including a missense mutation in *RPGRIP1L* and another in *TLR4*.

## PgmNr 1600: Cerberus: A laboratory information management framework in support of research and clinical operations.

### Authors:

B. Craig <sup>1,2</sup>; K. Duerr <sup>1,2</sup>; M. Hurley <sup>1,2</sup>; C. Ongaco <sup>1,2</sup>; A. Robinson <sup>1,2</sup>; S. Griffith <sup>1,2</sup>; M. Mawhinney <sup>1,2</sup>; K. Doheny <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Inherited Disease Research , Baltimore, Maryland; 2) Johns Hopkins Genomics, Baltimore, Maryland

---

The Center for Inherited Disease Research (CIDR) provides high quality next generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. The expansion of research and clinical genomic services from collaboration with other Johns Hopkins Genomics members have highlighted the need to create a new Laboratory Information Management System (LIMS) to support the breadth of current and future protocols. Previously we described the development of a MySQL database backed Java framework consisting of tools to reduce disruption to lab processing with module hot swapping, limit database schema complexity, and create a catalogue of reusable graphical user interfaces (GUI) to flexibly build and extend workflows for any protocol driven technology. Now named Cerberus, we present the first production use from this effort, the replacement of our legacy genotyping LIMS for all variations of Illumina Infinium assays. Lab managers create experiments in Cerberus by queuing samples from the in-house sample handling system and connecting these lists to job code templates that load wet-bench phases for user input (UI). Phase classes that are common between protocols can be reused for each job, reducing codebase complexity. All phase submission data is stored as XML metadata in the database reducing schema complexity and creating an audit of changes. Technicians scan one or more experiment entities that load UI for the next phase available to those entities. Custom validations are in place to help efficiently use lab automation resources, check expected reagent type suffixes, and prevent sample swapping. A robust comment tracking UI is easily accessible from any phase ensuring comments are not forgotten. Reports integral to downstream analysis have been implemented in a consolidated and efficient manner. Utility jobs restricted to administrative roles exist for the management of user access, validation of wet-bench inventory, and predefined data value modification. Multiple installations exist to aid in testing and transition of jobs from software developers to lab management and into production use. This model ensures that feature and unexpected behavior requests are adequately tested and approved before being deployed for production use. Having transitioned genotyping wet-bench operations to Cerberus use, we now target inclusion of research sequencing workflows.

# PgmNr 1601: A unified platform for scalable genomic data analysis.

**Authors:**

H. Davidge

[View Session](#) [Add to Schedule](#)**Affiliation:** Databricks, San Francisco, CA.

---

As the growth of genomic data continues to accelerate, single node tools are unable to keep pace. Distributed computing platforms, which provide high-level APIs to parallelize operations across thousands of CPU cores, allow practitioners to scale their analyses as data accumulates. Apache Spark has already gained traction as the engine that underlies popular tools like GATK4, Hail, and ADAM. This talk explores solutions that we have built on top of Spark and the Databricks platform to further scale common genomic query patterns.

A framework for efficiently assigning short reads to different nodes in a cluster has allowed us to distribute single node variant calling tools with roughly linear efficiency. With this method, we are able to use the GATK4 HaplotypeCaller to call variants on a 30x coverage whole genome sample in under 10 minutes for less than \$1 in AWS costs while producing results concordant with the reference single node implementation. When coupled with BWA, we can align and call variants in 24 minutes for under \$3.

We also describe techniques for accelerating ad-hoc analyses and joint variant calling. By extending the query optimizer built into Spark SQL, we achieve speedups of 10-100x on common query patterns such as genomic interval joins. Optimized readers for BGEN and VCF files enable these queries to run at biobank scale over genomics file formats or general purpose big data file formats like Apache Parquet. Our joint variant calling pipeline scales nearly linearly with the number of cores available and has successfully called cohorts of several thousand whole genomes.

# PgmNr 1602: High-throughput genetic variant classification for inherited cancer gene panels through an artificial intelligence inference engine.

## Authors:

F.M. De La Vega <sup>1,2</sup>; S. Nohzadeh-Malakshah <sup>1</sup>; E. Frise <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Fabric Genomics, Oakland, California.; 2) Dept. of Biomedical Data Science, Stanford University School of Medicine, Stanford, California.

---

A growing number of clinical labs are implementing cancer risk screening tests that sequence panels of cancer genes ranging from a few to over a hundred genes. This growth is driven by both the reduction of sequencing costs and the availability of reimbursement pathways for such tests. However, an important part of the cost involves the assessment of variant pathogenicity by trained clinical geneticists. The ACMG and AMP developed evidence-based guidelines to standardize variant assessment defining 28 criteria for supporting evidence of pathogenicity, which are then combined to classify a variant as either pathogenic, likely-pathogenic, benign, likely-benign, or uncertain significance. Although widely adopted in clinical interpretation of variants this process has remained largely manual and time-consuming. We recently developed an artificial intelligence-based forward-chaining inference engine implementing the ACMG-AMP criteria that takes as input annotated variants and codified gene-condition curation to automate the classification of variants. Here we analyzed with our engine 9,539 variants previously classified in a large Japanese survey of breast cancer patients and controls (7,051 cases and 11,241 controls), and for 50,000 individuals publicly released by Color Genomics. In addition, we analyzed variants from a cohort of 2,853 cancer patients of diverse histology whose genomes were sequenced as part of the International Cancer Genome Consortium. With this data we evaluated the performance of our classification inference engine for a 15-gene hereditary cancer risk panel we previously curated. We show up to 74% concordant classification of the variants previously classified by Color Genomics and in the Japanese cohort. Importantly, our engine produces zero misclassifications. Unclassified variants are annotated with the resolved criteria, enabling rapid manual classification. In addition, we are able to classify up to 46% of variants previously deemed VUS in these databases. On the ICGC donors we observe most variants benign, with just a few pathogenic, and <10% unclassified in need of manual review. Population screening for cancer risk by sequencing of cancer gene panels is actively being discussed by healthcare stakeholders, but cost is still posing a challenge. We demonstrate that our methods can allow clinical labs to scale-up in such tests while reducing processing effort and cost, potentially enabling wider adoption of cancer risk testing.

## PgmNr 1603: Modeling the ACMG/AMP guidelines as a quantitative approach.

### Authors:

Fan <sup>1,2</sup>; J.D. Wilkinson <sup>3</sup>; J.A. Schubert <sup>4</sup>; M. Tariq <sup>4</sup>; A. Sridhar <sup>4</sup>; S.D. Colan <sup>7</sup>; L.A. Sleeper <sup>7</sup>; L. Shi <sup>10</sup>; C.E. Canter <sup>9</sup>; D.T. Hsu <sup>8</sup>; N. Bansal <sup>8</sup>; S.A. Webber <sup>6</sup>; D.A. Dodd <sup>6</sup>; M.D. Everitt <sup>11</sup>; P.F. Kantor <sup>15</sup>; J.W. Rossano <sup>14</sup>; E. Pahl <sup>16</sup>; P. Rusconi <sup>17</sup>; L.J. Addonizio <sup>2</sup>; J.A. Towbin <sup>13</sup>; A.K. Lal <sup>12</sup>; J.L. Jefferies <sup>5</sup>; B.J. Aronow <sup>5</sup>; P.J. Dexheimer <sup>5</sup>; L.J. Martin <sup>5</sup>; E.M. Miller <sup>5</sup>; S.E. Lipshultz <sup>3</sup>; S.M. Ware <sup>4</sup>; Y. Shen <sup>1</sup>; W.K. Chung <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Systems Biology, Columbia University, New York, NY; 2) Pediatrics, Columbia University Medical Center, New York, NY; 3) Pediatrics, Wayne State University School of Medicine and Children's Hospital of Michigan, Detroit, IN; 4) Pediatrics, Indiana University School of Medicine, Indianapolis, IN; 5) Pediatrics, Cincinnati's Children's Hospital Medical Center, Cincinnati, OH; 6) Pediatrics, Vanderbilt University School of Medicine and Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, TN; 7) Pediatrics, Boston Children's Hospital, Boston, MA; 8) Pediatrics, Albert Einstein College of Medicine and Children's Hospital at Montefiore, Bronx, NY; 9) Pediatrics, Washington University School of Medicine, St. Louis, MO; 10) Pediatrics, New England Research Institute, Watertown, MA; 11) Pediatrics, Children's Hospital Colorado and University of Colorado, Aurora, CO; 12) Pediatrics, Primary Children's Hospital, Salt Lake City, UT; 13) Pediatrics, Le Bonheur Children's Hospital, Memphis, TN; 14) Pediatrics, The Children's Hospital of Philadelphia, Philadelphia, PA; 15) Pediatrics, Stollery Children's Hospital and University of Alberta, Edmonton, AB; 16) Pediatrics, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 17) Pediatrics, Miller School of Medicine University of Miami, Miami, FL

**BACKGROUND:** The American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology standardized variant interpretation and classify many variants as variants of uncertain significance (VUS). VUS can cause confusion and a diagnostic dilemma for patients and providers. Recently a Naïve Bayes model was proposed to estimate a probability of pathogenicity based on several questionable assumptions: independent evidence and 10% prior probability of pathogenicity. We have sought to improve based on a quantitative model.

**METHODS:** We first quantified two pieces of evidence as exemplars: 1) PVS1 was presented as a probability of a null variant based on gene isoform expression in a relevant tissue; 2) PM2 was presented using minor allele frequency (MAF) in gnomAD exome dataset. Second, we summarized ACMG evidence using principal component analysis (PCA) and estimated the probability of pathogenicity in a case-control cohort. Cases (N=333) were participants of European ancestry enrolled in the Pediatric Cardiomyopathy Registry. Controls (N=11,736) were unaffected parents of European ancestry in SPARK (Simons Foundation Powering Autism Research for Knowledge). We compared our results with the previous Naïve Bayes model using 37 known cardiomyopathy genes in the case-control cohort.

**RESULTS:** In this dataset, 29% of the null variants had <30% probability of being truly loss-of-function, and would change a piece of very strong evidence. The first principal component (PC1) is correlated with pathogenicity score. The loadings of the PC1 indicated PM2 (MAF) and BP6 (benign records) explained most of the variance in assigning pathogenicity. We compared our scores with the Bayesian probabilities and the ACMG rules. For pathogenic and likely pathogenic definitions, the three

methods had nearly identical classification performance (112 estimated risk variants with 90% probability). For VUS, quantitative methods could better communicate pathogenicity likelihood and prioritize variants for validation. Our PCA approach provided a slightly higher positive predictive value of pathogenicity (72%) compared with the Naïve Bayes model (66%) given a similar number of estimated risk variants (N=138).

**CONCLUSIONS:** Quantitative modeling of ACMG guidelines offers an improved understanding and greater detail about variant pathogenicity. PCA reduces evidence correlation and outperforms the Naïve Bayes model. The improvement is not significant given a small dataset.

## **PgmNr 1604: RefSeq Functional Elements: An annotated dataset of validated non-genic elements for variant interpretation and functional discovery in the human genome.**

### **Authors:**

C. Farrell; T. Murphy; RefSeq Curation and Development Team

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA

---

The human genome contains many non-genic elements that play roles in gene regulation, chromosome organization, recombination, repair or DNA replication. Human disease can result from sequence variation in those elements, with many genome-wide association studies indicating disease-associated variation in non-coding regions. The locations of gene regulatory elements can be predicted from several large-scale epigenomic mapping projects, but those data are not generally visible in traditional genome annotation, are difficult to interpret in the absence of specialized research knowledge or customized displays, and do not always show function when tested experimentally. NCBI has therefore introduced a more accessible dataset, RefSeq Functional Elements ([www.ncbi.nlm.nih.gov/refseq/functionalelements/](http://www.ncbi.nlm.nih.gov/refseq/functionalelements/)), which are annotated on the human genome alongside conventional genes. This curated dataset, which is restricted to known elements from published experimental data, includes richly annotated RefSeq records and accompanying descriptive records in the Gene database ([www.ncbi.nlm.nih.gov/gene/](http://www.ncbi.nlm.nih.gov/gene/)). The dataset includes known enhancers, silencers, recombination regions, and other non-genic regions with experimentally-validated function. As of NCBI's Updated Annotation Release 109.20190125 on the GRCh38.p12 genome assembly, the dataset includes over 3.7K GeneIDs and 8.8K feature annotations, with further growth expected for future NCBI annotation releases. The dataset is publicly available for FTP download ([ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo\\_sapiens/](ftp://ftp.ncbi.nlm.nih.gov/genomes/Homo_sapiens/)). Feature annotation can be visualized in the 'Biological regions' track available in NCBI browsers, including the Genome Data Viewer. This data track is particularly useful when viewed alongside other available tracks, such as variation tracks from dbSNP, ClinVar or dbVar, or study-specific custom tracks or track hubs. These non-genic annotations provide insights into non-coding genome function. They are valuable for basic discovery of gene regulatory regions, interpretation of non-coding variants, or as known positive controls for genome-wide studies aimed at discovering additional elements. This presentation will provide further information on how to use and access the dataset, and specific examples of functional determination will be presented. This work was supported by the Intramural Research Program of the NIH, National Library of Medicine.



## PgmNr 1605: A framework for mapping GWAS loci to causal genes for common disease and traits.

### Authors:

V. Forgetta <sup>1</sup>; N.A. Vulpesu <sup>2</sup>; J.A. Morris <sup>3,4</sup>; M.S. Hogan <sup>2</sup>; S. Grinek <sup>2</sup>; A.K. Loomis <sup>5</sup>; C.L. Hyde <sup>6</sup>; E. Fauman <sup>7</sup>; J.B. Richards <sup>1,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital; 2) Institute for Systems Genetics and Dept. of Pathology, NYU School of Medicine; 3) New York Genome Center; 4) Department of Biology, New York University; 5) Emerging Science & Innovation, Pfizer Worldwide Research, Development and Medical; 6) Early Clinical Development, Pfizer Worldwide Research, Development and Medical; 7) Internal Medicine Research Unit, Pfizer Worldwide Research, Development and Medical; 8) Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill University

---

A barrier to translate GWAS findings into clinical insights is mapping associated loci to causal genes for common disease. We use a positive control approach to identify genomic features that provide strong enrichment for causal positive control genes and a probabilistic method to evaluate whether a specific gene in an associated locus is causal.

Causal positive control genes for common diseases were obtained from Mendelian causes using OMIM and from the gene targets of approved medications. We selected 10 diseases — type 2 diabetes, atrial fibrillation, bone density, rheumatoid arthritis, diastolic and systolic blood pressure, height, low-density lipoprotein and triglycerides (combined as lipids) and inflammatory bowel disease — because all have large-scale GWAS findings, enabling testing of methods. GWAS loci were statistically finemapped to define a set of putatively causal SNPs (cSNPs). The genomic features mapped from each locus to one or more genes included those within 100 kb of a cSNP or closest to or overlapping a cSNP, or containing a deleterious coding cSNP, or a cSNP that maps to a cell-relevant open chromatin peak. For each trait, we trained logistic regression models using leave-one-out cross validation, where a single trait model was trained using all other traits using causal positive control gene status as the dependent variable.

We identified 488 positive control genes, where 380 and 108 are from Mendelian causes or medication targets, respectively. We finemapped 8,061 loci ranging from 91 to 3,286 for rheumatoid arthritis and height, respectively. Genes with predicted causal probability of >1% showed enrichment for positive control genes, with odds ratios ranging from 7 (95% confidence interval (CI): 2-20) for hypertension to 42 (95% CI: 18-90) for type 2 diabetes. Genes with causal probability of >5% had higher enrichment, with odds of 25, 95% CI: 3-103, and 160, 95% CI: 35-589 for hypertension and type 2 diabetes, respectively. However, these estimates are less certain due to the lower number of positive control genes per disease. The best predictors were cSNPs that overlapped open chromatin peaks (16, 95% CI 10-22) and those that were deleterious coding variants (82, 95% CI: 35-129).

Human GWAS evaluated with cell-specific genomic features can identify known causal genes across common diseases and traits. These methods formally test the strength of human genetic evidence

supporting a causal role for genes in human disease and traits.

## PgmNr 1606: When gold standards don't shine: Quality control of large genome datasets using genome fingerprints.

### Authors:

G. Glusman; M. Robinson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute for Systems Biology, Seattle, Washington.

---

The 1000 Genomes Project (TGP) is a foundational resource to modern human biomedicine, serving as a standard reference for human genetic variation. Three new versions of the TGP dataset were since released, expressed relative to the current version of the human reference sequence (GRCh38), and partially validated by benchmarking one genome (NA12878) against the reference for the same genome from the Genome In A Bottle Consortium.

We used our ultrafast genome fingerprinting method [1] to evaluate all four versions of the TGP: 1) the original cohort of 2504 individuals, relative to GRCh37; 2) the same 2504 individuals, 'lifted over' to GRCh38; 3) an expanded cohort of 2548 individuals, natively mapped onto GRCh38 and calling only biallelic SNVs; and 4) the same 2548 individuals on GRCh38, calling biallelic SNVs and indels. We also evaluated the smaller, supplemental cohorts of related individuals that have been released.

Our analyses revealed discrepancies of several different types. The increase of 44 genomes resulted from adding 45 new genomes, but also removing one. Of the 45 additional genomes, 34 have various unannotated relationships. Eight genomes from the original cohort, all from the ACB population, show changes so extensive that they are difficult to explain by differences between the reference sequences alone; eight other ACB genomes show less difference, but still substantially more than genomes outside of ACB. These differences include decreased SNV count, decreased heterozygosity, and poor genotype concordance with the prior version of the same genome. Finally, two sex-discordant genomes have nearly identical autosomes.

Our observations suggest that the quality of processes as complex as sequence mapping and variant calling, applied to a large number of samples, cannot be extrapolated from comparison of a single sample against a gold standard. When detailed comparison cannot be applied to all samples, a rapid, approximate evaluation of the kind provided by genome fingerprinting can identify additional, unexpected quality issues, and support the goal of providing the community with resources meeting a high quality standard. As the size of whole genome cohorts reach into the hundreds of thousands (e.g., the current TOPMed project) and beyond, rapid, comprehensive quality testing will become an important part of 'best practice' quality control.

1. Ultrafast Comparison of Personal Genomes via Precomputed Genome Fingerprints. *Front Genet.* 2017;8:136.

## PgmNr 1607: Detecting copy number variation on low coverage whole genome sequencing.

### Authors:

L.S. Hon; J. van den Akker; D. De Sloover; C.L. Neben; A.Y. Zhou; Z. Mahkovec; J. Ginsberg; J. Lock; S. Topper

[View Session](#) [Add to Schedule](#)

**Affiliation:** Color Genomics, Inc., Burlingame, California.

---

### Introduction

Next generation sequencing has become an integral tool in the study of human genetics, with whole genome sequencing (WGS) typically targeting 30X coverage. To support population-scale studies, low coverage WGS (lcWGS, <5X coverage) can be used to survey genomes in a cost-effective manner and is not limited by pre-determined probe design as in array CGH and FISH. Copy number variation (CNV) is an important component of genomic diversity and has been linked to a number of diseases in a dosage-sensitive manner. Here, we assess the feasibility and accuracy of using lcWGS to detect CNVs and explore CNV across a set of multi-allelic genes.

### Methods and Results

Laboratory procedures were performed at the Color laboratory, with DNA was extracted from blood or saliva samples and sequenced using the NovaSeq 6000 instrument, at ~1X coverage. Using a bioinformatics pipeline developed for clinical detection of structural variants in a targeted NGS panel, we tuned read depth based calling (CNVkit) for use with lcWGS. We utilized the resulting data in two ways. First, we assessed detection performance using a set of 137 clinical structural variants that had been previously confirmed present by array CGH, MLPA, or Sanger sequencing. Using lcWGS, we were able to reliably call and detect deletions as small as 20 kb and deletions greater than 100 kb; duplications had lower sensitivity, with reliable detection achieved at over a megabase. Notably, in one clinical sample a CNV near *NBN* was initially called from the targeted NGS panel data as a 3 MB copy number gain, but lcWGS and array CGH detected a chr8 trisomy. Given the sensitivity in detecting these larger events, we further characterized rare, larger structural variants and compared them against those found in gnomAD-SV. Second, we explored CNV across a set of multi-allelic genes using set of 949 clinical samples. We observed unique genomic structures in the *AMY1* region across a subset of 149 samples with high copy number. We also compared samples with extreme copy number within genes such as *LPA* to explore the relationship of dosage with available phenotypic data.

### Conclusion

lcWGS provides a rich data type with a number of unique characteristics. For CNV detection, it can allow for greater resolution in multi-allelic gene regions with high copy number and reliable detection of large and rare structural events, with further research and validation required to understand its utility for clinical use.

## PgmNr 1608: Multisample metrics for genome-wide single-position-resolution sequence quality assessment.

### Authors:

N. Johnson <sup>1</sup>; M.A. Bekritsky <sup>1</sup>; C. Colombo <sup>2</sup>; B.Z. Jia <sup>1</sup>; S.S. Ajay <sup>1</sup>; M.A. Eberle <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA; 2) Illumina Cambridge Ltd., Illumina Centre, 19 Granta Park, Great Abington, Cambridge, Cambridgeshire, CB21 6DF, UK

---

In a clinical setting, it is essential to understand sequence quality issues affecting each locus, both to evaluate variants called and to understand the probability of failing to call a variant that actually exists in the patient. While resources such as the Genome in a Bottle consortium and Platinum Genomes provide ground truth data for labs to sequence and quantify the accuracy of variant calls compared to an external standard, they do not cover most reference positions and are unlikely to cover the loci of pathogenic variants, which are rare by definition. Therefore it is highly desirable to understand a priori which loci have sufficient quality to be reliably called. Genome-wide measures of sequencing quality, such as the percentage of basecalls exceeding a certain quality threshold or the percentage of sites covered to at least a certain depth, may be of limited relevance at a specific locus of interest. On the other hand, locus-specific information from a single sample has little power to expose pipeline and platform biases affecting a site.

We have developed a software tool to aggregate sequence-quality metrics (e.g. read depth, base quality and mismatch rate) from many WGS samples. We applied this to groups of samples processed with consistent sample preparation, sequencing platform, and software pipeline. Based on these sequence-quality metrics, we identified deviations from normal sequencing quality at each position of the genome. By drawing on information from up to 150 samples, we were able to identify anomalies in the data with much greater robustness than would be possible with a single sample.

We give an overview of the means by which multi-sample sequence quality metrics are generated and discuss the designation of outlier regions with respect to those metrics. We report on the prevalence of genomic regions with highly consistent sequencing behavior, and compare our list of confident regions to other resources and to validated pathogenic variant calls. We further investigate how machine-learning methods and automated segmentation of the genome can be used to extend this work and potentially recover additional regions of reliable variant callability along with other annotations of interest.

## PgmNr 1609: Medical Genomics Japan Database (MGeND): 2019 update.

### Authors:

M. Kamada; M. Nakatsui; R. Kojima; E. Uchino; Y. Okuno

[View Session](#) [Add to Schedule](#)

**Affiliation:** Graduate School of Medicine, Kyoto University, Kyoto, Japan

---

To accelerate the implementation of genomic medicine in clinical practice, it is required to accumulate and share associations between genotypes and clinical phenotypes. Several databases containing genetic information and their clinical significance have already been constructed. However, the genomic information stored in those databases has been primarily obtained from US and European populations, and it is reported that genes and genotypes associated with the risk of onset of several diseases can vary between ethnic groups. Moreover, since some diseases are known to be triggers for other diseases, interpretation of variants across diseases is necessary to elucidate variants and diseases with unknown mechanisms.

For this reason, we developed a database, Medical Genomics Japan Database 'MGeND' (<https://mgend.med.kyoto-u.ac.jp/>) that provides disease-related genomic variants and corresponding clinical information in Japanese individuals. The first version of MGeND was released in March 2018 with genomic variations collected from Japanese research groups in the fields of “cancer,” “rare/intractable disease,” “dementia,” “infectious disease,” and “hearing loss”.

The number of data in MGeND has been increasing since it has been released; as of May 2019, 21260 SNVs/CNVs/Fusions, 22772 susceptibility variants for diseases identified using GWAS analysis, and 6195 HLA allele frequencies are published.

In addition, the target disease field has been further expanded. We are particularly focusing on polygenomic diseases, such as type 1 diabetes and narcolepsy, and genomic and clinical information has begun to be collected.

MGeND has disease-specific viewers to provide appropriate information for each disease field. In dementia, it is known that the genotype of the Apolipoprotein E (ApoE) gene is involved in the onset risk. Thus, we developed a viewer that can confirm the distribution of ApoE gene's genotypes based on selected phenotype in collaboration with research group of dementia.

Analysis of the registered data of MGeND has identified several variants common to different diseases. It is expected that investigating these variants will assist in clarifying the underlying onset mechanisms for each disease.

In the presentation, we will introduce the latest version of MGeND, and future prospects.

## PgmNr 1610: Development of semantic APIs for biomedical databases standardized in Med2RDF.

### Authors:

T. Katayama <sup>1</sup>; M. Kamada <sup>2</sup>; S. Kawashima <sup>1</sup>; M. Nakatsui <sup>2</sup>; R. Kojima <sup>2</sup>; Y. Okuno <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Database Center for Life Science, Joint Support-Center for Data Science Research, Research Organization of Information and Systems, Japan; 2) Department of Biomedical Data Intelligence, Graduate School of Medicine, Kyoto University, Japan

---

In the previous ASHG, we reported the Med2RDF project (<http://med2rdf.org>) which aggregated major biomedical databases as Resource Description Framework (RDF) based on the Semantic Web technologies. The resulting resource has a lot of potential for driving clinical genome medicine by developing data analysis methods such as artificial intelligence (AI) and machine learning (ML) technologies. However, the usage of this integrated resource can vary depending on the user's purpose, and also writing an effective "SPARQL Protocol and RDF Query Language (SPARQL)" query requires an in-depth understanding of the data models of each dataset. Therefore, it is demanded to provide reference SPARQL queries for typical cases as a set of APIs so that it can be easily used for data extraction and application development.

In this work, we stored RDF data in our SPARQL endpoint and introduced the SPARQLList application (<https://github.com/dbcls/sparqlist>) to provide REST APIs for executing complex SPARQL queries and subsequent data transformation at once. With SPARQLList, we provide APIs which extract biomedical facts and relations of the entities including genes, variants, diseases, drugs, and interactions from the integrated RDF datasets. These APIs can be used for visualizing data on a Web application which consumes JSON data. For example, at the National Bioscience Database Center (NBDC) and the Database Center for Life Science (DBCLS), we develop the TogoVar database (<https://togovar.biosciencedbc.jp>) for accumulating comprehensive Japanese genomic variations, in which we show annotation data obtained from the SPARQLList APIs using the TogoStanza framework (<http://togostanza.org>).

Further, these APIs are expected to contribute for accelerating AI and ML analysis by providing a relevant subset of the entire graph as a training set for the method to develop. Our group in the Kyoto University is developing methods for predicting pathogenicity of variants and relations of variation-drug to help experts curating variants of uncertain significance (VUSs). Then the curated interpretations will be incorporated into our Medical genomics Japan Variant Database (MGeND; <https://mgend.med.kyoto-u.ac.jp>). During the course of development, it is required to refine the training set for AI/ML methods. It can be easily accelerated with SPARQLList because each API is documented in the Markdown-formatted text and can be forked and edited to make a customized API for this purpose.

## **PgmNr 1611: Identifying clinically relevant results in 145,000 exomes: Experiences and lessons learned from population genomic screening in the Geisinger MyCode® Community Health Initiative.**

### **Authors:**

M.A. Kelly <sup>1</sup>; J.B. Leader <sup>1</sup>; T.N. Person <sup>1</sup>; K.E. Wain <sup>1</sup>; M.T. Oetjens <sup>1</sup>; R.G.C. Regeneron Genetics Center <sup>2</sup>; D.H. Ledbetter <sup>1</sup>; C.L. Martin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Geisinger, Danville, PA.; 2) Regeneron, Tarrytown, NY.

---

The Geisinger MyCode® Community Health Initiative is a biorepository with >240,000 consented patient-participants; exome sequencing and linked electronic health record (EHR) data are currently available for >145,000 participants, known as the DiscovEHR cohort. Participants consent to receive clinically relevant results after confirmation in a CLIA certified/CAP accredited laboratory. Reportable results are limited to pathogenic and likely pathogenic variants in 61 genes reviewed for actionability using internal criteria and definitions proposed by the CDC and the ACMG secondary findings recommendations.

Timely identification of clinically relevant results from a dataset at this scale requires an efficient and systematic filtration pipeline. To meet this need, we shifted from the standard patient-centered approach that focuses analyses on a single patient and developed a variant-first method utilizing both bioinformatics-based filtration and manual review to collate a list of variants across our cohort. In the first 92,456 participants, we identified 73,983 variants passing variant-quality thresholds in our genes of interest. Additional filtering based on population frequency, predicted impact, and ClinVar classification and star status reduced the number to 801 variants in 1,978 participants. These variants then underwent manual review by a variant scientist, further reducing the number selected for clinical confirmation by 34% to 527 in 1,139 participants, resulting in a clinical confirmation cost savings of 42%.

Variant annotations from external sources are versioned and stored in a database that seamlessly recognizes changes as annotations evolve, facilitating identification of variants requiring re-review. Technical hurdles included harmonizing variant nomenclature across annotation sources. Interpretive challenges arose around the variant-first approach, including identification of inaccurate variant calls, complex variants, and compound heterozygous individuals.

Large-scale population genomic screening poses new challenges compared to current patient-centered interpretation pipelines. We have developed a scalable, variant-first method for timely identification of clinically relevant results and a dynamic data management strategy that accounts for the evolving nature of genomic knowledge. By combining bioinformatics-based filtration with a focused variant scientist review, we have increased the specificity and the cost-effectiveness of our approach.



## PgmNr 1612: Implementing a scalable sequencing analysis pipeline for clinical whole-genome samples using pipeline accelerators.

### Authors:

Z. Khan; J. Farek; E. Venner; F. Sedlazec; N. Siddiqui; C. Sanchez; T. Wu; S. Kalla; V. Yi; J. Traynelis; D. Muzny; R. Gibbs

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

---

**Abstract:** Decreasing costs in next-generation sequencing (NGS) and the increasing importance of non-coding regions in clinical interpretation have prompted a shift from whole-exome sequencing to whole-genome sequencing (WGS) in clinical sequencing projects. The Baylor College of Medicine Human Genome Sequencing Center (BCM-HGSC) sequences and analyzes over 2,000 human whole-genome samples per month from multiple projects and collaborators, with a growing number of these samples as part of clinical studies. Adapting WGS analysis to clinical projects introduces additional analysis requirements, such as shorter turn-around times and the reliable detection of variants with known clinical significance.

We present a sequencing analysis pipeline for analyzing whole-genome clinical samples sequenced on the Illumina NovaSeq platform from FASTQ sequences to alignment, variant calling, and variant annotation. This pipeline is modeled from NHLBI TOPMed analysis specifications, using the GRCh37 human reference genome, with optimizations for computational scalability. We utilize software with low computational resource requirements in various pipeline steps, including xAtlas, a lightweight SNV and small indel variant caller developed at BCM-HGSC. Further improvements being evaluated in scalability are the use of Sentieon NGS data processing software or Illumina's DRAGEN platform. Sentieon's highly optimized BAM and CRAM-processing produces alignments that are nearly identical to those processed by more conventional sequencing analysis software, while providing a two-fold to eight-fold decrease in processing time for BAM and CRAM-processing pipeline steps. DRAGEN Platform uses high configurable FPGA to provide hardware-accelerated implementations of genome analysis.

We hope to show comparisons on WGS runs of various validation samples which will include numerous variants. Once complete we wish to show results describing variant recovery and computation performance improvements due to the accelerators used.

Our pipeline is implemented as a Snakemake workflow, which allows flexible control of pipeline and execution parameters. Clinically significant variants, including those from ACMG and OMIM panels, were confirmed to be detectable by analysis in validation samples.

## PgmNr 1613: Joint taxa/gene analysis of metagenomic sequencing data.

### Authors:

J. Kim <sup>1</sup>; S. Jiang <sup>2</sup>; G. Xiao <sup>1,3,4</sup>; Y. Xie <sup>1,3,4</sup>; A. Koh <sup>3,5,6</sup>; X. Zhan <sup>1,3,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Quantitative Biomedical Research Center, Department of Population and Data Sciences, University of Texas Southwestern Medical Center, Dallas, TX; 2) Department of Statistical Science, Southern Methodist University, Dallas, TX; 3) Harold C. Simmons Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX; 4) Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX; 5) Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, TX; 6) Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX; 7) Center for Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX

---

In microbiome research, metagenomic sequencing generates enormous amounts of data. These data are typically classified into taxa for taxonomy analysis, or into genes for functional analysis. However, a joint analysis where the reads are classified into taxa-specific genes is often overlooked.

To enable the analysis of this biologically meaningful feature, we developed a novel bioinformatic method, accompanied by a software toolkit, MetaPrism, to analyze sequence reads for a set of joint taxa/gene analyses: 1) classify sequence reads and estimate the abundances for taxa-specific genes; 2) tabularize and visualize taxa-specific gene abundances; 3) compare the abundances between groups, and 4) build prediction models for clinical outcome.

We illustrate these functions using a published microbiome metagenomics dataset from patients treated with immune checkpoint inhibitor therapy and showed the joint features can serve as potential biomarkers to predict therapeutic responses.

MetaPrism is freely available at <https://github.com/jiwoongbio/MetaPrism>.

## PgmNr 1614: Database and supercomputer resources for human data at DDBJ.

### Authors:

Y. Kodama <sup>1</sup>; A. Fukuda <sup>1</sup>; T. Hachiya <sup>2</sup>; C. Hata <sup>1</sup>; J. Mashima <sup>1</sup>; T. Fujisawa <sup>1</sup>; M. Kawashima <sup>3</sup>; L. Toyooka <sup>3</sup>; N. Mitsunashi <sup>3</sup>; O. Ogasawara <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformation and DDBJ Center, National Institute of Genetics, Mishima, Shizuoka, Shizuoka, Japan; 2) Genome Analytics Japan Inc., Tokyo 162-0067, Japan; 3) National Bioscience Database Center, Japan Science and Technology Agency, Tokyo 102-8666, Japan

As a centralized repository, the Japanese Genotype-phenotype Archive (JGA, <https://www.ddbj.nig.ac.jp/jga>) provides access to raw and processed personal genomes from array-based and next-generation sequencing platforms, and related phenotype data, in collaboration with the National Bioscience Database Center (NBDC) of the Japan Science and Technology Agency (<https://humandbs.biosciencedbc.jp>). NBDC has established data sharing guidelines and policies, and its Data Access Committee (DAC) permits submission and access requests to JGA. Summaries of available JGA studies are listed at <https://ddbj.nig.ac.jp/jga/viewer/view/studies> (100 studies, 236,593 samples and 72 terabytes of data files as of 28 May 2019).

The JGA system is connected to our secure supercomputer through a high-speed network so that users can smoothly download and analyze data with pre-installed bioinformatics tools on the supercomputer.

We have improved our database and supercomputer resources by implementing the Global Alliance for Genomics and Health (GA4GH) standard ontologies and tools. 1) Data use restrictions (policies) of the JGA datasets have been tagged with the Data Use Ontology codes (DUO, <https://github.com/EBISpot/DUO>) to allow searching use conditions such as 'datasets available for cancer research'. 2) The standard variant-call pipeline is made available in Common Workflow Language (CWL) (<https://github.com/ddbj/human-reseq>) so that users can reproduce calling results by running the same workflow on different computers.

To promote group collaboration using pre-publication research data, we also started a new service "DDBJ Group Cloud (DGC)" in February, 2017. One example is the "AMED Genome Group Sharing Database (AGD)" for sharing human genomes among researchers funded by AMED (Japan Agency for Medical Research and Development). AGD and JGA use the identical data model, and AGD data can be easily transferred to JGA upon data publication.

In the poster, we present recent developments of our database and supercomputer systems for archiving, sharing and analysis of human data.

## PgmNr 1615: Surfing petabytes of genomic data from your laptop.

### Authors:

A. Kotlar; K. Karczewski; T. Poterba; J. Goldstein; D. King; P. Schultz; C. Vittal; A. Wang; C. Churchhouse; B. Neale; C. Seed

[View Session](#) [Add to Schedule](#)

**Affiliation:** Massachusetts General Hospital and Broad Institute of MIT and Harvard, Boston, MA.

---

As terabyte-scale human genetic datasets become increasingly ubiquitous, advanced programming and technical skills are becoming necessary for participation in genomic research. Analysis at such scales exceeds the capabilities of all but the largest computational clusters and makes interactive data exploration impractical. In order to empower scientists with a broader range of computational backgrounds, we present a simplified genomic analysis framework built on scalable software. Recent projects like Hail and Bystro have addressed this challenge of scale in two different ways. Hail is a powerful toolkit that enables a wide range of analyses on arbitrarily-large datasets, used on projects like gnomAD to generate public summary statistics. However, it is relatively complex, and is optimized for processing entire datasets at the cost of lower efficiency on some individual tasks, like variant/sample annotation and filtering. In contrast, Bystro provides efficient variant/sample annotations, as well as millisecond natural-language queries in a simple web interface, but is less flexible.

Here we introduce Bystro Analyze, which combines the strengths of both methods and makes genomics at petabyte scales available in an easy-to-use, modern web application. It enables researchers to perform tasks such as variant/sample quality control, natural-language filtering, and association tests on datasets at the scale of UK Biobank and gnomAD, without requiring programming experience. Most importantly, we introduce a searchable analysis exchange, where users can share custom analysis pipelines for others' benefit. These editable widgets enable non-programmers to build and share complex analyses at scale. For instance, one could perform ancestry PCA based on gnomAD loadings in a few clicks with this system.

To demonstrate its abilities, we used Bystro Analyze to re-analyze all 125,748 whole exomes and 15,708 whole genomes from gnomAD. Annotation was completed with roughly 500-fold greater efficiency than previously described. The natural-language search engine was then used to perform a number of tasks with nearly interactive performance. For instance, identifying the number of high-confidence predicted loss-of-function mutations in the gnomAD dataset took roughly 1 second. In short, Bystro Analyze dramatically simplifies genomic analysis at large scales, bringing into reach for a diversity of scientists tasks previously only possible at the most resource-rich institutions.

## PgmNr 1616: A flexible simulation toolkit for designing and evaluating ChIP-sequencing experiments.

### Authors:

M. Lamkin <sup>1</sup>; A. Zheng <sup>1</sup>; Y. Qiu <sup>1,2</sup>; K. Ren <sup>3</sup>; A. Goren <sup>1</sup>; M. Gymrek <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of California San Diego, La Jolla, CA USA; 2) Carnegie Mellon University, Pittsburgh, PA USA; 3) Massachusetts Institute of Technology, Cambridge, MA USA

---

Chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq) is a widely used technology for genome-wide mapping of the location of DNA-associated proteins, such as transcription factors (TFs), histone modifications (HMs), and chromatin regulators (CRs). Dozens of methods have been developed for quantitatively analyzing ChIP-seq data, including peak callers (e.g., MACS2, HOMER) and differential binding tools (e.g. diffbind, DESeq2). A major challenge in training and evaluating these methods as well as interpreting their results is a lack of reliable ground truth data: in most cases, the actual location and strength of binding sites is not known and cannot be reliably measured using orthogonal experimental techniques. Accurate simulation of ChIP-seq data can mitigate this challenge, but existing frameworks are either too cumbersome to use or do not capture important sources of variation present in real data such as pulldown non-specificity or fragment length variability. Here, we present Tulip (Toolkit for simULating IP-sequencing data), a flexible toolkit for rapidly simulating ChIP-seq data based on realistic statistical models of key steps of the ChIP-seq process. Tulip simulates four key steps of ChIP-sequencing: Shearing, Pulldown, PCR, and Sequencing. Tulip consists of two modules, learn and simreads. The learn module infers model parameters such as shearing fragment length distribution, antibody specificity, and PCR duplicate rates, from existing data (Peak and BAM file). These parameters can then be used as input to simreads, which simulates ChIP-seq reads in FASTQ format based on specified model parameters. We demonstrate the utility of Tulip with two use cases: (1) measuring power to detect binding sites under a range of experimental conditions and (2) analyzing the effects of spike-in normalization controls. Tulip is implemented using C++ with multithreading capabilities and is currently available on github. Overall, Tulip is an efficient, versatile ChIP-Seq simulation framework with the capabilities to generate realistic datasets over a flexible range of experimental conditions.

## PgmNr 1617: Machine learning models for classification of small sample size RNA sequencing data.

**Authors:**

S. Listopad; T.M. Norden-Krichmar

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of California, Irvine, CA, USA

---

**Background:** Most studies are limited to small numbers of samples due to the high costs of recruitment, sequencing, data storage, and analysis. This results in many data sets with a low number of samples and large number of features, making classification challenging for machine learning (ML). For example, our study has RNAseq data from 36 individuals with alcoholic hepatitis (AH) and from 20 healthy controls. Our overall aim is to establish the best ML model for classification of gene expression counts to correctly distinguish AH from healthy samples.

**Methods:** We examined four types of ML models for classification of the RNAseq samples: Logistic Regression, Gaussian Naïve Bayes, k Nearest Neighbors (kNN), and Support Vector Machines (SVM). We used the Python scikit-learn implementation of these models with the default settings. The samples were obtained from the Southern California Alcoholic Hepatitis Consortium, RNA sequenced with 2x100 paired-end Illumina HiSeq, followed by standard RNAseq quality control. The Tuxedo pipeline generated distribution normalized counts for every sample. We performed 10-fold cross validation with feature selection of the top differentially expressed genes (DEGs) using CuffDiff. The top DEGs were filtered for  $\text{abs}(\log_2(\text{fold change})) \geq 1.0$ ,  $\text{FPKM} \geq 1$ , and significance at FDR-adjusted  $p\text{-value} \leq 0.05$ . The ML classifiers were then trained and validated with the top ( $p=10, 20, 35, 50, 100, 250, 500$ ) DEGs. Accuracy was used as a performance metric.

**Results:** Our initial testing indicated that all four models required at least 50 features to achieve classification accuracy of 90% or higher. As the number of features increased, the performance increased until it plateaued around 500 features. Gaussian Naïve Bayes was the best performing classifier at 98% accuracy with a large number of features ( $p = 250$  and  $500$ ). Regularized logistic regression and regularized linear SVM classifiers also performed well at 95% ( $p=250$  and  $500$ ). The kNN algorithm ( $k=7$ ) performed the worst across all feature sets, but still achieved classification accuracy of 92% ( $p=500$ ).

**Conclusion:** Preliminary results for our small sample size RNAseq dataset indicated that Gaussian Naïve Bayes was the best performing classifier given a large number of features. With a smaller number of features, classification accuracy near 90% was reached across all models. Transforming the features and tuning each individual model can obtain further improvements in accuracy.

# PgmNr 1618: DeepRepeat: De novo identification of tandem repeats on electric signal data from Oxford Nanopore sequencing.

## Authors:

Q. Liu <sup>1</sup>; K. Wang <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

---

**Background:** Tandem repeats represent a common type of genomic sequence variation in human genome. Several repeat expansions have been associated with different types of human diseases. These repeats can range from hundreds of bp to thousands of bp, and are much longer than the read length of short-read sequencing platforms. Long-read sequencing on the PacBio and Oxford Nanopore platform have been used to detect repeat expansions due to the ability to generate longer reads than repeat regions. For example, we previously developed RepeatHMM, which uses a hidden Markov model on long reads to quantify the number of repeats in known repeat regions. However, the error rates of base calls in repeat regions are higher than other regions due to the low complexity, which poses a challenge for repeat quantification from sequence data alone. Several recent studies showed that signals from Nanopore sequencing may be used to identify repeats with sophisticated machine-learning techniques.

**Methods:** We developed a computational method called DeepRepeat, which uses deep learning to estimate self-similarity of Nanopore signals of adjacent subsequences with repeat length, and then automatically detect repeat regions in long reads. To do that, long reads are aligned with a reference genome, and then the portions in long reads which are aligned with candidate repeat regions are identified. The signals distribution of events in the identified regions and adjacent events are used as input of DeepRepeat to make the predictions on whether the events are repeat units in a repeat region, outside a repeat region, insertions or deletions in a repeat region.

**Results:** From analysis of real data, we demonstrated that regions with tandem repeats show highly characteristics signal intensity patterns compared to neighboring regions. We next evaluated DeepRepeat on several whole-genome sequencing data sets and amplicon sequencing data sets generated by Oxford Nanopore sequencing. We compared the results with those obtained from analyzing sequences of long reads directly. Our analysis demonstrated that DeepRepeat can improve the quantification of repeat counts in tandem repeat regions, and can perform de novo detection of novel repeat expansions.

**Conclusion:** Our evaluation suggests that Oxford Nanopore sequencing together with DeepRepeat provide a useful tool to identify tandem repeats at a genomic scale, and may facilitate the discovery of repeat expansions associated with human diseases.

## PgmNr 1619: Entanglement mapping: A model-free approach to detecting interactions among predictive features.

### Authors:

J. Malley<sup>1</sup>; D. Kiser<sup>1,2</sup>; J. Sabourin<sup>1</sup>; A. Musolf<sup>1</sup>; J.E. Bailey-Wilson<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Human Genome Research Institute, Rockville, MD.; 2) Renown Institute for Health Innovation, Desert Research Institute, Reno, Nevada

---

In many genome-wide association studies (GWAS), the proportion of variation in disease risk or quantitative trait measurement that can be attributed to individual significantly-associated genetic variants has tended to be small. One possible reason is that most GWAS look at only single risk factors or independent features, and do not examine potential interactions among all such features. That is, individual features, such as genotypes and environmental measures, may be only weakly predictive in themselves, but strongly predictive when part of a subgroup of other features. We provide a scheme for identifying such *entangled communities* of interacting features and demonstrate its performance on simulated data.

This scheme for locating entangled features requires only a single pass through the full (potentially large) list of features. The method removes or randomizes a feature and determines which of the remaining features are also down-weighted in importance (using any reasonable feature importance method) when compared to the analysis with all features. All the features in the identified subgroup are *jointly* necessary for good prediction, but individually may be only weakly predictive. For each down-weighted subgroup, this entire set of features must be included for good prediction. Simulations show the algorithm can detect: (1) higher order interactions of more than two features, (2) interactions that don't follow the multiplicative model, (3) multiple sets of interacting features, and (4) interactions in the presence of noise features and main effects.

This is a model-free, parameter-free scheme for detecting jointly predictive features and works with any learning machine.



## PgmNr 1620: ClinGen Linked Data Hub: Scalable infrastructure for aggregation of diverse types of variant information to support pathogenicity assessment.

### Authors:

A. Milosavljevic <sup>1</sup>; S. Subramanian <sup>1</sup>; A.R. Jackson <sup>1</sup>; R.Y. Patel <sup>1</sup>; S. Paithankar <sup>1</sup>; M.A. Iacocca <sup>3</sup>; C. Preston <sup>3</sup>; M.W. Wright <sup>3</sup>; H. Costa <sup>3</sup>; B. Wulf <sup>3</sup>; H. Wand <sup>3</sup>; C. Thaxton <sup>4</sup>; S. Madhavan <sup>5</sup>; A.I. Su <sup>6</sup>; C.D. Bustamante <sup>3</sup>; S. Plon <sup>1,2</sup>; Clinical Genome Resource Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 2) Cancer Genetics Clinical and Research Programs, Texas Children's Hospital, Houston, TX; 3) Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, CA; 4) Department of Genetics, University of North Carolina, Chapel Hill, NC; 5) Department of Oncology, Georgetown University, Washington DC; 6) Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA

---

Variant pathogenicity assessment requires evaluation of numerous data types such as functional, population, case-level and in-silico predictors. The ClinGen Linked Data Hub (LDH) aggregates structured data from a diverse set of sources, ensuring variant curators have access to current evidence. LDH is a scalable and extensible infrastructure to facilitate efficient access to collated information about any subject such as genes, variants and others to support curation efforts. Using newly developed W3C Linked Data technologies, LDH enables contributors to link information about variants registered in the ClinGen Allele Registry (CAR) and store a subset of relevant structured information within the LDH itself. LDH tracks provenance of the information and provides a permanent reference for inclusion as supporting evidence in variant interpretations. Select data and links are stored in LDH as “quotes” indicating that specific information (i.e. “*Linked data*”) about the variant (i.e. the “*Subject*”) was provided by a specific source at a specific time (“*provenance*”). The “quote content” is modeled using the ClinGen/GA4GH community standards.

LDH is implemented using the Genboree platform, which also provides cloud-hosted tools for modeling, validation, and hosting of detailed information which may be linked through LDH. It is accessible via a user interface linked through the CAR and also via APIs for use by ClinGen curation tools including the Variant Curation Interface (VCI), the Pathogenicity Calculator as well as by the tools developed outside of ClinGen. LDH can/will support notification capabilities over all tools, thus facilitating resolution of discrepancies due to stale information, and keep variant knowledge up to date. LDH integrates with the ClinGen community annotations available through Hypothes.is, and also with the WikiData Gene Wiki project. Here, we present the beta version of the LDH resource linking information from 40 genes and over a million variants for use by the VCI. We demonstrate tools for modeling and linking layers of variant evidence from Genboree-hosted Functional Data Repository (FDRRepo), Case-level Data Repository (CDRepo) and other external data sources, tracking provenance and provision of permanent references. Finally, we explore the future potential of the novel approaches for exchanging genomic information first implemented by LDH to empower clinical

genomics and genetics and also to open new roads to discovery in genomics.

## PgmNr 1621: Evaluation of bioinformatic pipelines performances in detecting germline variants.

**Authors:**

R. Minillo; M. Cervato; M. Caraciolo; G. Carvalho; R. Reis; T. Almeida; J. Filho

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genomika-Einstein, São Paulo, São Paulo, Brazil

---

Bioinformatic pipelines have become crucial components in interpreting next generation sequencing (NGS) data. However, there is high variability between pipelines performances currently used in routine clinical genetic diagnosis. In order to minimize this issue we developed a web platform for costumed annotation and analysis of NGS data called Varstation. We evaluated the performance of six bioinformatic pipelines suitable for Varstation ([www.varstation.com](http://www.varstation.com)) and also used by Genomika Einstein laboratory for routine clinical diagnosis. We sequenced sample NA12878 with comercial kits used in our laboratory and we compared the list of variants found using six pipelines that were available at Varstation with gold standard variants of sample NA12878. We used FASTQ files as well as other entry files and performed all bioinformatic steps at Varstation. All pipelines analyzed have mapping, aligning reads to reference genome, variant calling and variant annotation steps. In each step, quality measurements were performed in order to guarantee a good quality of the data. The main germline pipelines available at Varstation were evaluated according to good international guidelines for variant calling. Our analyzis showed that these pipelines reached precision of 99%, sensitivity higher than 95% and accuracy higher than 96%. We report the performance of bioinformatic pipelines available at the web platform Varstation and our results show high precision, sensitivity and accuracy in identifying variants of sample NA12878. This analyzis is important for checking the performance of these tools and should be implemented in every center that analyzes NGS data.

## PgmNr 1622: Unsupervised deep representation learning for genetic variant-based clustering of individuals.

### Authors:

M. Oh <sup>1</sup>; E. Cosgun <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Virginia Tech, Blacksburg, VA.; 2) Microsoft AI & Research, Redmond, WA

---

Clustering of individuals based on their genetic variation is a fundamental step towards identifying novel disease subtypes and precisely tailored treatments as well as geographic population structures. The next-generation sequencing technology allows researchers to capture accurate genome sequences in reads. The optimization in secondary analysis such as Microsoft Genomics service substantially accelerates reads-to-variants workflow identifying genome-wide variants at scale. However, the extreme high-dimensionality of the variant data and the relatively small amount of available data points limit its utility for unsupervised learning in the tertiary analysis. Here, we propose to utilize Deep Embedded Clustering (DEC), an unsupervised deep learning approach that transforms the high-dimensional genetic variants into low-dimensional latent representation and that iteratively refines the learned representation to be suitable for clustering analysis. More specifically, we pre-trained deep autoencoder model to get initial representation. Then, the weights and biases of the encoder network have been additionally updated by feeding clustering loss as an objective which forces ambiguous sample points to be nudged toward the most probable centroid in the latent space based on Kullback-Leibler divergence between the student's t-distribution and the target distribution. We compared our approach with baseline models which are equipped with widely-used dimensionality reduction algorithms, such as Principal Component Analysis (PCA), Multi-dimensional Scaling (MDS), Locally-Linear Embedding (LLE), and Isomap, and the following clustering algorithms, including k-Means, Gaussian Mixture Model, and Spectral Clustering. To evaluate clustering performance, we used Microsoft Genomics service to call variants of 448 individuals in the 1000 genome project and clustered their origins (African, British, Chinese, Colombian, and Finnish) in an unsupervised manner. Our approach successfully recovered the population structure of different origin and surpassed the best performing baseline method in terms of clustering accuracy (increased by 28%). Potential applications of our approach include identification of novel disease subtypes based on genetic variation and characterizing patient groups to optimize medical treatments.

## PgmNr 1623: OpenCRAVAT: A customizable annotation and prioritization pipeline for genes and variants.

### Authors:

K. Pagel <sup>1</sup>; L. Zheng <sup>1,2</sup>; R. Kim <sup>3</sup>; K. Moad <sup>3</sup>; M. Ryan <sup>3</sup>; R. Karchin <sup>1,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Institute for Computational Medicine, The Johns Hopkins University, Baltimore, MD; 2) McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD; 3) In Silico Solutions, Falls Church, VA; 4) Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD; 5) Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD; 6) Department of Computer Science, The Johns Hopkins University, Baltimore, MD

---

The modern genetics researcher is confronted with hundreds of published methods to interpret variants. There are databases of genes and variants, phenotype-genotype relationships, algorithms that score and rank genes, and in silico variant effect prediction tools. Because variant prioritization is a multi-factorial problem, a welcome development in the field has been the emergence of decision support frameworks, which make it easier for users to integrate many resources in an interactive environment. Current decision support frameworks are limited by closed proprietary architectures, access to a restricted set of tools, lack of customizability, web dependencies that expose protected data, and limited scalability.

OpenCRAVAT is a new open source, scalable decision support system to support variant and gene prioritization, relevant to inherited disease and cancer. It offers a dynamic biologist-friendly GUI, allowing users to: install with a single command, easily download tools from an extensive resource catalog, create customized pipelines, run jobs at speeds that exceed current variant annotation API services, and explore results in a richly detailed viewing environment. OpenCRAVAT is distinguished from similar tools by the amount and diversity of data resources and computational prediction methods available, which span germline, somatic, common, rare, coding and non-coding variants. We have designed the OpenCRAVAT resource catalog to be open and modular to maximize community and developer involvement, and as a result the catalog is being actively developed and growing larger every month.

We will present several case studies to illustrate the design of custom workflows to prioritize causal variants. OpenCRAVAT is freely available for non-profit use at <https://opencravat.org>

## PgmNr 1624: Integrative multi-omic approach to identifying functional amino acids.

### Authors:

M. Palafox <sup>1</sup>; K. Backus <sup>3,4</sup>; V. Arboleda <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, David Geffen School of Medicine at University of California, Los Angeles; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at University of California, Los Angeles; 3) Department of Biological Chemistry, David Geffen School of Medicine at University of California, Los Angeles; 4) Department of Chemistry and Biochemistry, University of California, Los Angeles

---

Each of the 20 common amino acids have unique side chains that confer chemical property differences. This diversity contributes to the challenges presented in the clinical interpretation of missense variants from whole exome sequencing data. Various computational methods have been developed to predict missense deleteriousness. However, the predictive power of these tools is somewhat limited, particularly for poorly characterized genes. We hypothesize that further integration of protein-level information with these predictive algorithms will allow us to better interpret the potential effects of missense mutations. Here we investigate the relatedness between protein composition and gene intolerance to change. We used the non-redundant UniProtKB H. sapiens proteome as our reference proteome, quantified each amino acid's abundance per protein and normalized by protein length, and used pLI gene constraint scores provided by gnomAD. We found that the abundance of certain amino acids is significantly associated with higher odds that the gene is constrained (pLI > 0.9). When proteins are binned into three equal-sized groups based on normalized amino acid abundance (low, medium, or high), we found that abundance groups with the highest odds ratio for the occurrence of gene constraint were low leucine (OR 1.98, CI 95% 1.82-2.15), high serine (OR 1.67, CI 95% 1.54-1.81), and high aspartic acid (OR 1.66, CI 95% 1.52-1.80). Constrained genes have lower proportions of leucine and cysteine (mean difference -0.012 and -0.005, respectively) and higher serine and aspartic acid (mean difference 0.006 and 0.005, respectively) compared to non-constrained proteins. By probing questions like the one presented here through the integration of multi-omic datasets, we hope to identify important associations between features corresponding to different levels of the central dogma, and ultimately advance our understanding of functional variation.

## PgmNr 1625: LocusFocus: A web-based tool to prioritize genes and tissues for study from genome-wide association studies.

### Authors:

N. Panjwani <sup>1</sup>; F. Wang <sup>2</sup>; A. Bao <sup>1</sup>; S. Mastromatteo <sup>1,3</sup>; J. Gong <sup>1,3</sup>; L. Sun <sup>2,4</sup>; J.M. Rommens <sup>1,5</sup>; L.J. Strug <sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 3) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada

The majority of genome-wide association studies (GWAS) have identified trait-associated loci that fall within noncoding regions. Understanding how associated loci impact disease requires significant investigation since they can be extensive and encompass several genes. Integration of GWAS association results with expression quantitative trait loci (eQTL) can demonstrate colocalization of these independent sources of data and generate hypotheses about the responsible mechanism, gene and tissue of origin to guide functional investigation. Tools to visualize GWAS with eQTL data along with formal evaluation of statistical significance of colocalization across genes and tissues are needed.

We developed the locusFocus tool to facilitate the visualization and formal statistical testing of the colocalization of GWAS and eQTL data. Several approaches to statistical testing of colocalization of GWAS and eQTLs have been proposed, including our own *Simple Sum* (SS) approach (Gong et al 2019) that provides advantages over previously published methods with ability to detect colocalization in the presence of high LD and/or allelic heterogeneity without requirement for specific causal variation assumptions. The locusFocus tool was developed to visualize GWAS summary statistics with GTEx or user-specified eQTL p-values at a given locus in one web-based interactive plot. eQTL p-values from GTEx are obtained from Ensembl via API calls, and positions of genes from GENCODE v19. Pairwise linkage disequilibrium for a user-defined SNP can be displayed for user-selected populations from the 1000 Genomes Project for the GWAS summary statistics. Statistical evaluation for colocalization using the SS for all local genes and user-selected GTEx tissues is carried out and displayed in an interactive heatmap, demonstrating the gene(s) and tissue(s) that are most correlated with the GWAS evidence.

In summary, locusFocus facilitates the standard hypothesis-generating data-integration analyses following GWAS by prioritizing tissue(s) and gene(s). LocusFocus and its full documentation are available at <https://locusfocus.research.sickkids.ca> (Jun. 6, 2019).

## PgmNr 1626: Detecting transcriptional regulatory modules using machine learning.

### Authors:

N. Patel; W.S. Bush

[View Session](#) [Add to Schedule](#)

**Affiliation:** Cleveland Institute for Computational Biology, Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, United States, Case Western Reserve Univ, Cleveland, Ohio.

---

Binding of different transcription factors (TFs) in a gene's regulatory region is a pivotal step for the process of transcription. TFs can exert significantly different influence over a gene's expression via the nature of their binding. Although TF binding models such as additive sequential binding of individual TFs and non-additive binding of multiple interacting TFs ("regulosome") have been described, genome-wide characterization of these models has been limited. In this study, we have developed a novel machine learning based approach to estimate the impact of linear additive binding and non-linear binding of individual TFs and TF regulosomes respectively on gene expression. We used elastic net regularized linear regression (ENET) and multilayer perceptron (MLP) based prediction models to predict gene expression for 11,778 protein coding genes. These models were informed using inputs from multiple TF based transcriptional regulatory mechanisms such as TF-TF cooperativity, TF-target gene (TG) binding affinity and gene co-regulation. We aggregated ENET effect estimates and MLP layer weights across all genes to determine the influence of 149 individual TFs and their regulosomes respectively on gene expression in lymphoblast cells. We computed interaction scores using an algorithm called neural interaction detection (NID) for all the possible combinations of TFs to capture their non-linear interaction effects. TFs with the highest effect estimates included the crucial initiation factors (*TBP*, *TAF1* and *POL3RG*), as expected since they are often the first TFs to recognize and bind the transcription start sites (TSS). We also detected 10,391 multi-way TF regulosome complexes binding in the regulatory region (50kb upstream/downstream) of the genes that significantly influenced their expression. Of these, 6,195 were pairwise interactions with the most influential regulosomes (NID scores > 0.1) containing chromatin remodeling TFs such as *CBX5* and *ATF2* and TFs important for hematopoiesis namely *SMAD5* and *POU2F2*. Lastly, we found important regulosome modules regulating genes enriched for terms like Alzheimer's disease (*BHLHE40*, *ELF1*, *SMARCA5*, *CTCF*, *SMARCA5*, and *CREM*, *MEF2B*) and MAPK signaling (*BCL3*, *STAT1*, *TAF1*, *ZNF217*, and *ATF2*, *NFIC*), some of which were novel. Thus, we have devised a framework to efficiently detect and characterize different TF modules that may help us to understand their influence on gene expression as well as on specific phenotypes of interest.



## PgmNr 1627: Enhanced interpretation of genome-wide regulatory data using consensus set of enhancer-target gene assignments.

### Authors:

T. Qin <sup>1</sup>; R.G. Cavalcante <sup>2</sup>; C.T. Lee <sup>3</sup>; P. Orchard <sup>1</sup>; S. Patil <sup>1</sup>; M.A. Sartor <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MICHIGAN; 2) Biomedical Research Core Facilities, Epigenomics Core, University of Michigan, Ann Arbor, MICHIGAN; 3) Department of Biostatistics, University of Michigan, Ann Arbor, MICHIGAN

---

**Background:** Revealing the gene targets from distal regulatory elements is critical to interpret genome-wide study results, as “cis-regulome” is highly enriched for disease-susceptibility loci. However, this is hampered by limited knowledge about the association between enhancers and their target genes. High resolution experiment/association-derived enhancer-gene links are restricted to a small set of cell types, and their reliability and generalization remain elusive due to the lack of a systematic evaluation. **Methods:** By cross-linking different combinations of enhancer-defining methods across cell types (UCSC ChromHMM tracks, ENCODE DNase hypersensitive sites [DHSs], distal DHSs with correlated promoter DHSs, and FANTOM5 regions) to their target genes using spatial and *in silico* approaches (ChIA-PET links, FANTOM5 links, correlation between promoter and distal DHS signals, and enhancer-gene pairs within the same CTCF loops), we generated ~2,800 human genome-wide distal **Enhancer** (>5kb from a transcription start site [TSS]) to **Target** gene locus **Definitions** (**EnTDef**). We also extended EnTDefs by adding links between non-enhancer distal regions and the nearest TSS to increase the genome coverage up to 90%. **Results:** Using functional annotation and enrichment analysis on 87 independent ChIP-seq datasets of 34 transcription factors (TFs) among ENCODE tier-1 cell lines, we illustrated how the 741 top ranked EnTDefs significantly outperformed simply linking distal regions to the genes with the nearest TSS (FDR < 0.05). Interestingly, enhancers defined by ENCODE DHSs, FANTOM5, and/or DHS signal correlations performed better than those by ChromHMM tracks, and the ChIA-PET links surpassed other enhancer-gene linking methods. The performance of the top 10 EnTDefs on ChIP-seq data from ENCODE non-tier-1 cell lines was close to that from the tier-1 evaluation cell lines, indicating that the selected EnTDefs can be generalized to other cell types. Significantly, the consensus set of EnTDefs across cell types showed comparable results to those generated using the cell-type-specific ChIA-PET links, suggesting the power of our approach to provide cell-type-specific genome-wide interpretation. Therefore, our top ranked EnTDefs provide a valuable resource to interpret cis-regulome data from a variety of cell types.

## PgmNr 1628: SNP ReMap: A tool for updating GWAS summary data across genome builds.

### Authors:

N.W. Rayner <sup>1,2,3</sup>; N.R. Robertson <sup>1,2</sup>; M.I. McCarthy <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Churchill Hospital, Oxford, OX3 7LJ, UK; 3) Department of Human Genetics, Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK; 4) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LJ, UK

Genome Wide Association Studies (GWAS) have now been performed for more than a decade and have been incredibly successful in identifying disease associated loci, in part due to the widespread sharing of GWAS summary statistics allowing large, well powered meta-analyses to be undertaken.

The data from many of these individual GWAS studies remains available today however they are often not aligned to the latest genome build and in some cases can be several releases older, a problem when attempting to reuse these data sets in current analyses.

Imputation and reanalysis using the relevant genome build would be the best option to resolve this but often is not possible for a variety of reasons, such as an inability to access the original genotype data or contact the original analysis team.

Whilst options do exist to allow the bulk update of variant positions across genome builds they generally do not provide information on changes to the strand orientation of the variants, information critical to the incorporation of the summary data into any analyses.

To address this problem we have extended to GWAS summary data the concept we previously developed for directly genotyped array data, namely that of utilising the mapping of variant flanking sequences to a reference genome to derive their positions on multiple genome builds.

In this instance we have used the fasta sequences for every dbSNP variant from build 151 to provide the variant flanking sequence data. These have been aligned to the reference genomes using BLAT utilising settings designed to maximise the number potential alignments. To ensure the widest compatibility we have used all the genome builds relevant to the GWAS time period from NCBI35 to GRCh38. The alignments for all variants are then parsed to determine the best match to the genome in terms of both length and concordance. From this the position and strand orientation are derived for all aligned variants. Variants not matching the genome to the required concordance are noted as are variants exhibiting no discernable match. Variants showing multiple high quality matches are also recorded along with their alternate positions.

To date we have successfully used the program to update multiple data sets as part of our latest meta-analysis efforts and at the moment have been able to update on average more than 98% of the

variants successfully for both position and strand. We plan to further optimise this to increase the number of successful alignments.

## PgmNr 1629: Integration of best practice RNA-Seq workflows into cloud-based translational analysis platform.

**Authors:**

M. Rickles-Young; J. Tsuji; B. Granger; A. Macbeth; C. Cibulskis; T. Bowers; N. Lennon

[View Session](#) [Add to Schedule](#)

**Affiliation:** Broad Institute of MIT and Harvard, Cambridge, MA 02141

---

The Broad Institute has a long history in genomic sequencing and in the development of tools for researchers to analyze these data. With improvements in technology and reductions in cost, the rate of sequence generation is increasing, which necessitates a platform to scale the associated analyses. We also need to be able to apply our best practice methods across a range of complex workflows to support the breadth of science among our users. This challenge is what spurred the creation of the Translational Analysis Group (TAG) within the Genomics Platform at the Broad Institute. Over the past two years, our group has developed and maintained over 30 validated, version-controlled workflows and has run over 20,000 analyses on Terra, the Broad Institute's cloud-based analysis platform. Until recently, we have mainly focused on supporting germline and somatic variant analyses on whole genome and exome libraries, however, there is high demand to integrate RNA-sequencing (RNA-seq) into the analyses. In this presentation, we introduce our new RNA-seq workflows for bulk and single-cell RNA experiments. Our suite of RNA-seq workflows starts with mapping RNA reads to a reference genome and then profiles gene and the isoform expression. The bulk RNA-seq outputs can be used as inputs for the downstream workflows to perform differential expression and RNA variant calling analysis. For evaluating the workflows, we benchmarked with publicly available datasets such as GTEx to check the expression and the RNA variant calls against the matched exome. The development of our RNA-seq analysis capabilities increases the scope of projects, both internal and external, for which TAG can provide analysis services with the reproducibility, scalable resources, and version control necessary for consistency in studies which extend over a long period of time, such as clinical trials.

## PgmNr 1630: Interpretable clinical genomics with a likelihood ratio paradigm.

### Authors:

P.N. Robinson<sup>1</sup>; V. Ravanmehr<sup>1</sup>; J.O.B. Jacobsen<sup>2</sup>; A. Zhang<sup>1</sup>; D. Danis<sup>1</sup>; L.C. Carmody<sup>1</sup>; J. McMurry<sup>3</sup>; M. Haendel<sup>3</sup>; D. Smedley<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA.; 2) Genomics England, Queen Mary University of London, Dawson Hall, Charterhouse Square, London EC1M 6BQ, UK.; 3) Oregon State University, USA

---

In medical genetics and related fields, Human Phenotype Ontology (HPO) analysis has become one of the standard methods for diagnostics. Current algorithms such as Exomiser and Phenomizer use a variety of semantic and statistical approaches to prioritize the typically long lists of genes with candidate pathogenic variants, but do not provide robust estimates of the strength of the predictions beyond the placement in a ranked list, nor do they provide measures of how much any individual phenotypic observation has contributed to the prioritization result. Here, we present an approach to genomic diagnostics that exploits the clinical likelihood ratio framework to provide an estimate of the posttest probability of candidate diagnoses, the likelihood ratio for each observed HPO phenotype, and the predicted pathogenicity of observed variants. These tools will help clinicians interpret the contribution of each individual phenotypic abnormality. The odds ratio for the variants provides a measure of the probability of the gene to harbor rare, predicted pathogenic variants in the general population.

We curated a collection of 236 detailed clinical case reports from a wide range of Mendelian diseases, capturing clinical abnormalities as HPO terms and recording the disease-causing variants and the diagnosis. We spiked the variant into an otherwise “normal” VCF file and then ran LIRICAL. The correct diagnosis was placed at rank 1 in 178 of 236 cases (75%), within the top three candidates in 92% and within the top ten in 97%. The performance of Exomiser on the same dataset was 51% at rank 1, 56% within the top 3 ranks, and 83% within the top 10 ranks. On a set of solved 108 cases from the 100,000 Genomes Project, LIRICAL ranked 50/108 as the top hit, and put 82/108 in the top 5. This compares to 76/108 and 98/108 for Exomiser 12.0.0., suggesting that the performance of genomic diagnostic algorithms can be dataset dependent. By setting a minimum threshold of 75% posterior probability, LIRICAL was effective at limiting the number of candidates that require detailed manual review by diagnosticians. Additionally, LIRICAL provides a graphical output of the contribution of each query HPO term to the diagnosis. In summary, LIRICAL provides clinically interpretable results for phenotype-driven genomic diagnostics with performance comparable to current state of the art tools.

## **PgmNr 1631: Disease-specific enhancement for biological networks improves prioritizing candidate disease genes.**

### **Authors:**

P. Ruan <sup>1</sup>; S. Wang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Statistics, The George Washington University, Washington, D.C., USA; 2) Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY USA

---

Biological networks-based strategies have been proven to be useful in prioritizing candidate genes in various diseases. A number of comprehensive human gene networks have been developed, such as HumanNet, GIANT and STRING. However, these comprehensive human gene networks are not disease-specific and cannot be used to represent true gene interactions for a specific disease. Here, utilizing an established human gene network, the HumanNet, we propose a framework that enhances the comprehensive gene network for a specific disease using disease-specific omics data through a diffusion process. This diffusion process uses a doubly stochastic transition matrix generated from disease-specific omics data. The enhanced disease-specific gene network thus has weak edges from the original network being removed and disease-specific gene connections being enhanced. Through a random walk, the enhanced disease-specific network helps prioritize candidate disease genes for a specific disease. Applications to gene expression data of multiple cancers from The Cancer Genome Atlas (TCGA) project suggest that more prioritized candidate genes by the enhanced disease-specific networks are cancer-related than those by the original general human gene network as well as those selected using omics analysis only.

## PgmNr 1632: Improved detection of allele-specific expression via personalised mapping approaches.

### Authors:

A. Saukkonen <sup>1,2</sup>; H. Kilpinen <sup>2,3</sup>; A. Hodgkinson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, King's College London, London, United Kingdom; 2) UCL Great Ormond Street Institute of Child Health, 30 Guilford Street, London, WC1N 1EH, United Kingdom; 3) Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, CB10 1SD, United Kingdom

---

A key feature of the genetic architecture of complex traits is that many variants associated with disease are in non-coding regions and overlap significantly with variants modulating gene expression, thus making a strong case for the role of gene regulation in disease. On a population scale, variants influencing gene expression are detected through expression quantitative trait loci (eQTL) mapping, but this is not possible when sample sizes are small, or variants are rare. In these cases, detection of an imbalance in the expression of two alleles within a gene (allele-specific expression, ASE) is potentially more informative about genes that are undergoing genetic regulation in *cis*. However, this approach is susceptible to false positives due to technical factors affecting RNA-seq data such as read mapping bias. We have developed a pipeline that minimises errors in ASE detection through multiple steps of filtering, and incorporating isoform-aware parental genome mapping and haplotype structure resolution. We apply this pipeline to simulated RNA-seq and whole-genome sequencing data and show that it dramatically improves the accuracy of ASE calls, particularly in regions surrounding indels and rare variants. We also use the pipeline on real data from the HipSci project ([www.hipsci.org](http://www.hipsci.org)), where we see better coverage around heterozygous sites and reduced reference allele bias, thus facilitating more accurate detection of ASE events. We anticipate our pipeline to be useful when sample size is small, particularly for rare diseases.

## PgmNr 1633: Efficient and rapid clustering of identity-by-descent tracts in biobank-scale datasets.

### Authors:

R. Shemirani <sup>1</sup>; G.M. Belbin <sup>2</sup>; E.E. Kenny <sup>2</sup>; C.R. Gignoux <sup>3,4</sup>; J.L. Ambite <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Information Sciences Institute, University of Southern California, Marina del Rey, CA; 2) Center for Population Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Colorado Center for Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO; 4) Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO

---

Community detection for clustering of identity-by-descent (IBD) segments along the genome offers a powerful approach for rare disease mapping in massive population-based biobanks. However, when developing methods that scale efficiently, several challenges need to be addressed. First, methods that detect IBD in massive datasets use approximate matching techniques and are more prone to higher error rates. Second, IBD sharing in biobank-sized datasets poses a scale challenge, with the potential for billions of IBD segments to be present in modern datasets, such as the UKBiobank. This intensifies the need to rely on robust clustering techniques able to accurately identify IBD patterns and account for error at scale.

To remedy this situation we have developed a clustering method that employs robust community detection algorithms to scalably and accurately cluster IBD-haplotypes. iCURL (Identity-by-descent ClUstering of Related Loci) uses an efficient implementation of Markov Clustering in sliding windows along the genome to characterize overall patterns of relatedness and modularity diverse population settings. It also uses parallel computing libraries, available on scientific computational clusters, to further improve its performance. We also describe a novel community simulation technique to simulate realistic patterns of IBD clusters determined from real data in order to measure the performance of each algorithm and optimize its parameters. These simulations help us understand the effects of different windowing methods utilized by IBD mapping tools to reduce runtime and memory use, while calibrating clustering thresholds to improve statistical power. We compare iCURL with DASH-CC, a greedy algorithm designed for scalability. For the PAGE Study dataset of ~52,000 individuals with 15 million segments shared by IBD, iCURL took 132 minutes, compared to 466 minutes for DASH-CC, while both running on a single thread. Using a less restricted sliding technique, iCURL runtime can be reduced to under 5 minutes in the same dataset. We describe our strategy for running iCURL on UK Biobank to cluster IBD segments estimated from our ultra-rapid IBD method iLASH. iCURL allows tuning of the clustering inflation parameter that provides the ability to look for both fine-grained and coarse-grained communities sharing IBD. In summary, iCURL enables the investigation of IBD clusters, communities, and modularity with applications for IBD mapping in modern, large-scale datasets.



## PgmNr 1634: Leveraging natural genetic variation to discover DNA motifs mediating transcription factor binding and function.

**Authors:**

Z. Shen; M. Hoeksema; Z. Ouyang; C. Glass

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of California, San Diego, La Jolla, California.

---

Genome-wide association studies have identified thousands of disease-associated genetic variants. The majority of variants fall into non-coding regions, making the interpretation difficult. Many causal variants influence the binding of transcription factors (TFs) by altering TF binding sites, also called motifs. TF binding can induce the activation of enhancers or promoters and ultimately regulate gene expression. One locus often has more than one altered TF motif, and a genetic variant can alter multiple motifs simultaneously. Prioritizing functional motifs is important but remains challenging. Currently, there is a lack of tools that can identify causal motif alteration for differential TF function.

Here we introduce a statistical framework, MAGGIE (Motif Alteration Genome-wide to Globally Investigate Elements), to discover DNA motifs mediating TF binding, chromatin accessibility, and enhancer activity. We leverage the measurements of TF functions corresponding to at least two genotypes such as different alleles, individuals, or mouse strains. For loci with differential function, the locus-wise motif alterations between genotypes are aggregated and tested against the null hypothesis, which expects a random alteration for motifs that do not contribute to the function of interest.

By simulating genetic variation, we demonstrate that MAGGIE is able to identify motifs more precisely than conventional motif discovery tools. To validate its ability to reveal causal motif alterations responsible for changes in TF function, we tested MAGGIE on allele-specific binding sites computed from ENCODE data. Many TFs solely have their corresponding motifs as significant output by MAGGIE while other tools report many more motifs that are enriched but likely not required for TF binding. Finally, we applied MAGGIE to uncover motifs necessary for activation and repression of inflammatory gene enhancers in macrophages by leveraging H3K27Ac signals from five mouse strains. MAGGIE reports p65 as necessary for activated enhancers and p50 for repressed enhancers even though both TFs belong to the nuclear factor-kappa B (NF- $\kappa$ B) family and their motifs are indistinguishable by other tools. This relationship is validated by experimental assays measuring genome-wide binding of these TFs. Enhancers co-bound by p65 and p50 are activated in response to inflammatory signals, while enhancers bound by p50 alone are repressed.

## PgmNr 1635: CAVERN: Computational and visualization environment for RNA-seq analyses.

### Authors:

A.C. Shetty; R.S. Adkins; A. Chatterjee; C.L. McCracken; T. Hodges; H.H. Creasy; M. Giglio; A. Mahurkar; O. White

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland

---

**Background:** Over the last decade, there has been an exponential increase in the application of next-generation sequencing technologies to generate transcriptomic datasets. RNA-sequencing is being used increasingly to study gene expression profile in samples and to assess the differential gene expression between conditions in biological research. **Motivation:** The analytical approach of transcriptomic data involves multiple steps. First, the alignment of sequencing reads to a reference genome for each sample. Then quantification of gene expression for multiple samples belonging to 2+ conditions that are compared using statistical approaches to determine differential expression between conditions. This analytical approach requires intermittent checkpoints to control data quality, bioinformatics expertise, large computational resources, and the ability to interact with downstream results. While there are well developed tools for individual steps, few exist that integrate multiple steps into a comprehensive analytical pipeline to provide researchers with an end-to-end solution to quickly and easily process RNA-sequencing data. **Methods:** We developed CAVERN as an analytical resource to facilitate the end-to-end analyses of transcriptomics datasets. CAVERN comprises of a user-friendly interface GROTTTO and an interactive sandbox ALCOVE. GROTTTO allows a user to provide relevant information regarding input files for multiple samples, reference genome sequence and annotation, and to select from multiple analytical modules that support different aspects of the pipeline. A user can modify parameters that influence different analytical steps. GROTTTO enables pipeline creation, monitoring, and the ability to download relevant results. GROTTTO provides a quick overview of the results via a sampling of descriptive illustrations and tables. Alternatively, a user can access relevant data and results within ALCOVE, a R-based Shiny-server application. ALCOVE provides researchers with an interactive environment to view pertinent results, modify analytical features, and make inferences about their biological research questions. CAVERN has been implemented using the Docker framework to enable quick setup for use across multiple operating systems. **Conclusions:** Overall, we were able to develop a user-friendly computational and visualization environment for transcriptomic analyses that is available to the scientific community at <https://github.com/adkinsrs/ergatis-docker-recipes.git>.

## PgmNr 1636: Clinical actionability curation tools and knowledge repository of the ClinGen consortium.

### Authors:

S. Subramanian <sup>1</sup>; R.Y. Patel <sup>1</sup>; N. Shah <sup>1</sup>; A.R. Jackson <sup>1</sup>; S. Paithankar <sup>1</sup>; M.J. Gilmore <sup>2</sup>; J.E. Hunter <sup>2</sup>; K. Lee <sup>3</sup>; K. Mittendorf <sup>2</sup>; B. Powell <sup>3</sup>; K. Wallace <sup>3</sup>; E.M. Webber <sup>2</sup>; K.A.B. Goddard <sup>2</sup>; A. Milosavljevic <sup>1</sup>; Clinical Genome Resource Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Center for Health Research, Kaiser Permanente Northwest, Portland, OR; 3) Department of Genetics, University of North Carolina, Chapel Hill, NC

---

The ClinGen Adult and Pediatric Actionability Working Groups (AWG) assess the clinical implications of human genes and genetic disorders using a standardized protocol. The product of the AWG is evidence-based summary reports that define a known ability to intervene and avert a poor outcome due to a previously unsuspected high risk of disease. These reports of curated evidence provide a resource to aid decision makers as they determine best practices regarding secondary findings. To facilitate generation and dissemination of clinical actionability curation knowledge in both human and machine readable formats, we have developed the following informatics tools: (1) the Actionability Curation Interface (ACI), which is a web-based curation portal, (2) a knowledge repository of curation summaries and reports and (3) REST-APIs to programmatically access structured content.

ACI has been developed to streamline the curation of actionability knowledge in both adult and pediatric contexts and capture the curation workflow, which includes searching literature for evidence, documenting the references, performing an early survey to determine if the topic is actionable and if so, associating sets of gene-disease pairs with outcome-intervention pairs, applying semi-quantitative scores by experts and generating consensus scores, making assertions on the clinical actionability of genetic variations and abstracting data into a structured format. The application provides a transparent and systematic way to access supporting evidence and previous summary report versions, thus facilitating critical evaluation of current as well as prior recommendations in light of evolving evidence.

The ACI and associated services are developed using Genboree Stack, following the microservices architecture and implemented as Redmine Ruby-on-Rails plugins. The application includes role-based access control, allows tracking version histories and a web interface for browsing, filtering, searching and export of curated knowledge. REST-APIs enable retrieval of reports and summaries of individual and consensus scores and structured representation of curated evidence using the SEPIO ontology framework. As of June 2019, 92 genetic disorders have been evaluated and released in the adult context and 18 in the pediatric context. The summary reports and scores are publicly accessible through [actionability.clinicalgenome.org](https://actionability.clinicalgenome.org) and are integrated with other curation efforts of ClinGen through [clinicalgenome.org](https://clinicalgenome.org).

# PgmNr 1637: Memory bandwidth-aware seeding for genome sequencing.

**Authors:**

A. Subramaniyan; J. Wadden; K. Goliya; X. Wu; S. Narayanasamy; D. Blaauw; R. Das

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of Michigan-Ann Arbor, Ann Arbor, Michigan.

---

The process of identifying the locations of short substrings of a read in a candidate reference genome, known as seeding, is one of the most time-consuming step in whole genome secondary analysis. For example, in the widely used read alignment software BWA-MEM, seeding consumes 56% of the overall runtime (measured on 50x coverage reads from Illumina Platinum Genomes). In addition, seeding is also a performance bottleneck in several metagenomic classification tools such as Centrifuge.

Seeding is commonly performed using a highly-compressed index structure called the FM-index. While FM-index is highly space efficient (4.3 GB for a human genome), standard seeding algorithms using FM-index are memory bandwidth limited. Our key observation is that this trade-off is not optimal for modern systems which have availability of much large main memory (192 GB on Amazon c5n.18xlarge instance).

FM-index based seeding requires character-by-character processing of a read. Thus, each base-pair in every seed of every read issues a separate lookup into the 4.3 GB FM-index with little to no memory locality. On average, a 101 pb short read requires 108 KB of data from main memory and a complete human genome 158.4 TB of data in 2.72 T DRAM accesses, resulting in a severe memory bandwidth bottleneck.

Our proposed seeding algorithm is based on a new data structure called Enumerated Radix Tree (ERT) designed to addresses this bandwidth bottleneck: it trades-off a larger memory footprint for a significant reduction in memory bandwidth. Instead of using one monolithic large structure for representing all suffixes of the reference genome, like the FM-index, ERT groups together suffixes that share the same K-length prefix and enumerates all such prefixes in a K-mer index table with 1B entries. The suffixes for each prefix are then represented using a radix tree that naturally allows processing of multiple characters using a single lookup. Radix tree nodes that are likely to be accessed sequentially are placed close together in memory to reduce memory accesses during tree traversal. As a result, ERT enables multi-character lookup per memory access while the data structure is sufficiently compact to fit within 64GB of main memory.

Using ERT, we significantly improve the performance of seeding in BWA-MEM by 3.7x and its bandwidth efficiency by 7.57x on a 72-thread AWS c5n.18xlarge instance, while guaranteeing an identical (exact bit-equivalent) output as BWA-MEM (a de facto standard).

## PgmNr 1638: I-pute: A fast genotype imputation method based on positional Burrows-Wheeler transform.

### Authors:

K. Tang <sup>1</sup>; A. Naseri <sup>2</sup>; D. Zhi <sup>2</sup>; S. Zhang <sup>1</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, University of Central Florida, Orlando, Florida.; 2) School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, Texas.

---

Missing genotype data is a major barrier to genome-wide association studies. Modern imputation methods typically leverage a large reference haplotype panel. At this scale, fast methods with low-memory use and high accuracy are needed. We introduce a new computational tool, I-pute, using matching haplotype blocks in the positional Burrows-Wheeler transform (PBWT) matrix in the panel to impute missing data, without external references or complicated statistics methods. PBWT sorts similar haplotypes into blocks of matches at each site. It is very likely that a missing haplotype value remains the same as its neighbor haplotypes at the same site. We can impute a missing haplotype if it is within a block of haplotype matches and the surrounding neighbors have the same value. We compared our approach with Beagle on UK BioBank (UKB) SNP data and 1000G sequencing data with no reference. The results show I-pute runs 2 times faster than Beagle and is at least 10% more accurate on UKB SNP data and at least 3% more accurate on 1000G sequencing data. Non-reference based I-pute provides a fast imputation with high accuracy while avoiding using or choosing reference.

## PgmNr 1639: A bioinformatics-to-lab bench workflow to identify and characterize tandem repeats.

**Authors:**

D. Vandenberg; E. Domico

[View Session](#) [Add to Schedule](#)

**Affiliation:** Biobehavioral Health, Institute for the Neurosciences, Penn State Univ, University Park, Pennsylvania.

---

The “missing heritability” problem for complex traits is proving to be difficult to address. One possible reason is that the most common type of variant assayed (SNP) does not capture the genetic information of nearby variants of different classes. We have chosen to focus on Simple Tandem Repeats (STRs) based on recent results showing that STRs in or near the *AVPR1A* and *CACNA1C* genes are associated with phenotypes that are not associated with nearby SNPs (Landefeld, *et al.*, 2018; Song, *et al.*, 2018). Thus, characterization of repeats in the genome could provide insight into the genetic basis of disease, and it provides a platform for long-term undergraduate student engagement at both the bioinformatic and laboratory levels in a more real-world setting than is often accomplished in short summer internships.

Computational tools have been developed to enable researchers to identify important sequence elements with a focus on short tandem repeats in this project. A workflow was created using “etandem,” (Rice, *et al.*, 2000) to scan through sequence data to identify tandem repeats. It is run within Galaxy (Afgan *et al.*, 2018), a bioinformatics web interface hosted at Penn State, which aids in documenting the work. Resulting STRs were visualized using the UCSC Genome Browser and a candidate STR was chosen for wet-lab analysis by PCR to detect variation in the number of repeats. Pilot tests were done using an STR in the 5′-flank of *SLC6A3* (*DAT1*).

## PgmNr 1640: Searching efficiently through (genomic) sequences with vantage point trees.

**Authors:**

J. Vankerschaver; R. Kern; S.J. Kern; P. Zahemszky; C. Mueller

[View Session](#) [Add to Schedule](#)

**Affiliation:** Enthought Ltd., Cambridge, United Kingdom

---

Many tasks in computational biology involve determining sequence homology: given one or more DNA, RNA, or protein sequences, find the most similar sequence in a database of known sequences. Examples of this include determining the function of an unknown gene in a gene catalog, or assigning taxonomy information to a bacterial amplicon by finding the closest matching sequence in a database of sequences and taxonomies such as RefSeq.

From a computational perspective, to be able to process large volumes of data in a reasonable amount of time, this lookup must be fast as well as accurate. Standard tools such as Blast are able to provide fast lookups at the expense of some accuracy, whereas more accurate assignments can be obtained with tools such as glsearch or ggsearch (part of the fasta3 package), which perform a slower yet more accurate local-to-local or global-to-global alignment.

We will present a way of indexing databases of sequence data (DNA, RNA, protein, or arbitrary strings) into a vantage point tree, a datastructure which allows fast yet accurate query lookups by building a tree of 'near' and 'far' sequences. We will discuss the theoretical foundations of vantage point tree search, outlining how they achieve the best of both worlds: returning the most accurate match, without any approximations, with lookup time logarithmic in the size of the database. We will compare the behavior of our algorithm against standard tools such as Blast, ggsearch, or the RDP classifier, and we will discuss our reference implementation, showing in particular how a vantage-point tree can be stored efficiently as an array of offsets. This results in a simple, cache-friendly algorithm that can easily be ported across languages and platforms.

Most of our work is centered around databases of 16S amplicon data, for which global alignment distance is most appropriate, but vantage point trees can be used with other notions of distance metric as well. We will finish with a brief discussion of other distance functions and their use in computational biology.

## PgmNr 1641: Scandium: Sequencing coverage annotation database with uniformity metric.

### Authors:

K. Walker; P. Huang; D. Kalra; Q. Wang; R. Gibbs; D.M. Muzny

[View Session](#) [Add to Schedule](#)

**Affiliation:** Baylor College of Medicine, Human Genome Sequencing Center, Houston, Texas

---

With next-generation sequencing (NGS) technologies becoming routine in biological research and clinical diagnostics, the need for effective tools for quality control, sequence analysis, and annotation are essential to manage an active NGS pipeline and to fully capitalize on the resulting data. While many bioinformatic tools provide post alignment quality control metrics, none report the details of sequence coverage annotated by SNPs, gene, transcript or CDS (coding sequence). Those analyses are particularly useful in the clinical setting for applying stringent quality cutoffs before reporting disease-causing alleles.

To address this need, **Scandium** was developed in C as in-house open-source software. **Scandium** delivers comprehensive sequence QC and coverage metrics, including a newly developed WGS coverage uniformity metric. This metric measures the deviation of sequencing coverage from the central tendency. A low uniformity value alerts users to data variability, whereas a high uniformity value implies a uniform dataset, providing confidence in downstream analysis. Therefore, this metric reveals areas of the genome that are under-represented, with greater efficiency than by measuring average coverage.

For annotation, **Scandium** uses a backend MySQL database populated with publicly available annotation resources with all HGNC approved genes, but users can instead provide custom annotation via a bed file. **Scandium** calculates 'percentage of coverage' for each annotation type in NGS data which users can further customize by specifying cutoffs based on mapping quality, read quality or depth of coverage.

Finally, to visualize these metrics, an additional module creates user-friendly plots for uniformity across specific chromosomes and additional plots to highlight CDS or gene coverage as well as various exome-based and WGS-based coverage distributions.

**Scandium** employs OpenMP, a C/C++ multi-threading API, for faster processing of sequenced data. For a 35X WGS sample Scandium uses only 12 GB of memory and 1 hour of computation time, which is suitable for large sequencing centers and individual laboratories with limited computational resources.

**Scandium** has been used for clinical applications at the HGSC-CL to develop coverage metric paradigms across clinically reportable genes and to assess pathogenic variants at various coverage models.



## PgmNr 1642: A new annotation tool BioSplicer improves variant effect prediction of splice region variants using splicing context-specific alignment.

### Authors:

Wang <sup>1</sup>; D. Watson <sup>1</sup>; A. Nesbitt <sup>1</sup>; M. Harr <sup>1</sup>; C. Chung <sup>1</sup>; F. Mafra <sup>1</sup>; K. Stumm <sup>1</sup>; A. Wenocur <sup>1</sup>; L. Tian <sup>1</sup>; H. Hakonarson <sup>1,2,3</sup>; A. Santani <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pathology, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Interpretation of variants called by next generation sequencing data requires accurate annotations of variants, especially for their predicted biological effects. Multiple variant annotation tools have been developed for this purpose and are utilized by research and clinical laboratories. In order to assess the performance of annotation tools, we designed a comparison study between five commonly used programs. We analyzed annotations generated by Alissa (v5.1.3, Agilent), VEP (release 96, Ensembl), VarSeq (v2.1.0, Golden Helix), SnpEff (v4.3T) and ANNOVAR (wANNOVAR, 2019May) for >200,000 variants with matching transcripts from the HGMD database annotations (HGMD\_pro\_2019.1). In reviewing the discrepancies, we found that annotations associated with splice site regions as one of the major sources of inconsistency among these tools. Further analysis showed that alignment of indel variants around intron-exon boundaries is the cause of most of these errors. Discrepancies were observed in all six sources indicating that this is a persistent problem for all tools. Although it is appropriate at the variant calling step to utilize the left-alignment rule for genomic representation of indels, annotation algorithms should be adjusted relative to both the donor/acceptor and intron/exon directions to ensure that the annotations correctly predict the impact on splicing. Our study showed that 50% of the indel variants in and around the splice site regions predicted an incorrect impact in all the annotation tools. Therefore, we developed a new tool "BioSplicer" to automatically re-annotate splicing region indels with biologically correct alignments in order to predict the splicing effect with greater detail and accuracy. This tool can be integrated into a clinical diagnosis pipeline before the mainstream annotation tools are utilized to correct the alignment errors. With BioSplicer installed, all variants affected by this issue were reanalyzed and 78% of them had different annotations. Importantly, 55% of corrected annotations demonstrated category changes, such as a change from a predicted intron-skip to a frameshift and an intron-skip to an in-frame deletion. We expect this tool to facilitate more accurate and reliable variant annotation and disease diagnosis.

## PgmNr 1643: Whole genome sequencing for precision medicine.

### Authors:

Q. Wang; D. Kalra; K. Walker; H. Doddapaneni; J. Hu; R. Raj; Z. Momin; C. Nessner; C. Kovar; S. Dugan-Perez; J. Santibanez; N. Osuji; Z. Khan; J. Farek; S. Kalla; E... Venner; Y. Han; H. Dinh; R. Gibbs; D.M. Muzny

[View Session](#) [Add to Schedule](#)

**Affiliation:** Baylor College of Medicine, Human Genome Sequencing Center, Houston, TX.

High-throughput nucleotide sequencing has revolutionized genomic research and reshaped applications in clinical diagnostics. The NovaSeq 6000 platform expanded these opportunities by providing unprecedented capacity and the opportunity for routine clinical whole genome sequencing (WGS). In acceleration to the growing field of precision medicine, the Human Genome Sequencing Center Clinical Laboratory (HGSC-CL) has validated a germline WGS protocol in line with CLIA and CAP molecular requirements. The pipeline is built on more than a decade of experience with WGS of personal genomes, resulting in many process improvements. The development of a new pooling strategy to increase consistency across multiplexed samples, where sequencing libraries are first pooled based upon qPCR quantification and then re-pooled based on the first “calibration” sequencing results, has improved the coefficient of variation (CV) for an 81-plex pool from 17% to 6.4%. We also increased our sequencing yield, generating an average of 3.4 Tbases for each S4 flow cell with 27 WGS samples to yield an average 37x coverage, which exceeds the vendor’s specification. For clinical requirements, we determined that, at 35x average coverage, over 95% of clinically relevant genes (OMIM) are fully covered at 15x. The WGS pipeline now includes more than one hundred quality assessment metrics, including tests for sequencing coverage, sample integrity and representation of expected variants. Sequence coverage routinely exceeds 95% of the genome covered at 20x with a minimum of  $86 \times 10^9$  mapped, aligned bases with Q20 or higher.

An HGSC-CL workflow management system, HgV, is employed for all primary and secondary genome sequence analyses, executing base calling, mapping, merging, variant calling, post-processing and QC metric collection for all single-lane and multiplexed samples. The deliverables include alignment to either human reference GRCh37 or GRCh38, depending on project needs and a coverage analysis with annotation by SNPs, gene, transcript or coding sequence analyzed by Scandium. The pipeline also automates the first step of clinical interpretation, by using the in-house developed tools Cassandra and Mendelian Afterburner to annotate and filter the GRCh37 variants, reducing the review burden of the board-certified clinical geneticist team.

This WGS pipeline will be utilized for clinical WGS programs, including the NIH’s Fetal Genomes Sequencing project, and the National All of Us Research Program.

## **PgmNr 1644: MARRVEL 2.0 (Model organism Aggregated Resources for Rare Variant ExpLoration): A web-tool for human genetics and cross-species data collection.**

**Authors:**

J. Wang; S.Y. Kim; D. Mao; S. Pasupuleti; S. Yamamoto; H. Bellen; Z. Liu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Baylor College of Medicine, Houston, TX.

---

Genes and variants of unknown significance are one of the major challenges in understanding the cause of genetic disorders. There currently exists many tools that are useful for assessing the significance of genes and variants to establish a hypothesis that can be experimentally tested. However, the compilation of these dispersed data usually demands either a lot of time or bioinformatics skills. MARRVEL (Model organism Aggregated Resources for Rare Variant ExpLoration) provides a simple, easy-to-use tool for non-computational users interested in gathering data dispersed throughout dozens of tools and databases across the world-wide-web.

MARRVEL2.0 is a complete upgrade from MARRVEL1.0 which has accumulated more than 1600 users per month since 2017. The overall design, features, and backend have all been upgraded to improve user experience, expand functionality, and ensure the ability to stay up to date. The design of the tool emphasizes assistance with data interpretation and ease of navigation. The new features include commonly used pathogenicity prediction algorithms such as CADD and PolyPhen as well as an expanded set of model organism data such as phenotype and gene ontology ribbons that enable at-a-glance comparison across species.

## PgmNr 1645: Model-based autoencoders for imputing discrete single-cell RNA-seq data.

### Authors:

Z. Wei <sup>1</sup>; T. Tian <sup>1</sup>; R. Min <sup>2</sup>; H. Hakonarson <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) New Jersey Institute of Technology, Princeton Junction, New Jersey.; 2) NEC Laboratories America, New Jersey; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 4) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

---

Deep neural networks have been widely applied for missing data imputation. However, most existing studies have been focused on imputing continuous data, while discrete data imputation is under-explored. Discrete data is common in real world, especially in research areas of bioinformatics, genetics, and biochemistry. In particular, large amounts of recent genomic data are discrete count data generated from single-cell RNA sequencing (scRNA-seq) technology. Most scRNA-seq studies produce a discrete matrix with prevailing 'false' zero count observations (missing values). To make downstream analyses more effective, imputation, which recovers the missing values, is often conducted as the first step in pre-processing scRNA-seq data. In this paper, we propose a novel Zero-Inflated Negative Binominal (ZINB) model-based autoencoder for imputing discrete scRNA-seq data. The novelties of our method are twofold. First, in addition to optimizing the ZINB likelihood, we propose to explicitly model the dropout events that cause missing values by using the Gumbel-Softmax distribution. Second, the zero-inflated reconstruction is further optimized with respect to the raw count matrix. Extensive experiments on simulation datasets demonstrate that the zero-inflated reconstruction significantly improves imputation accuracy. Real data experiments show that the proposed imputation can enhance separating different cell types and improve the accuracy of differential expression analysis.

## PgmNr 1646: A rapid, accurate approach to inferring pedigrees in endogamous populations.

### Authors:

C. Williams <sup>1,2</sup>; B. Scelza <sup>3</sup>; J.A. Shortt <sup>1</sup>; E. Lange <sup>1</sup>; N. Rafaels <sup>1</sup>; K. Crooks <sup>1</sup>; K.C. Barnes <sup>1</sup>; C. Gignoux <sup>1</sup>; B.M. Henn <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Colorado Anschutz Medical Campus, Aurora, CO.; 2) University of California Davis, Davis, CA; 3) University of California Los Angeles, Los Angeles, CA

---

Despite the availability of multiple software packages for inferring relationships among individuals and reconstructing pedigrees, accurate assignment of relatives from genetic data remains a challenging problem. Certain algorithms are trained only on urban European-descent families, which are 'outbred' compared to many other global populations. Moreover, relationship categories can be difficult to distinguish between relationships with the same kinship coefficient (e.g., half-sibships versus avuncular) without external information. Furthermore, several published packages cannot accommodate endogamous populations where there may be reticulations within a pedigree, or may rapidly become computationally intractable. We designed a simple, rapid algorithm which initially uses only high-confidence 1st degree relationships to seed a machine learning step based on the number of identity by descent segments and a novel summary statistic called a haplotype score. We tested our approach in a sample of ~700 individuals from northern Namibia, sampled from an endogamous population. Due to the culture of concurrent (non-marital) relationships in this population, there is a very high proportion of half-sibships as well as cryptic relatedness. Further, we apply our algorithm to the Colorado Center for Personalized Medicine Biobank at the University of Colorado to detect unknown relationships in a multi-ethnic dataset with sparser training data. Our method accurately identifies all categories in 1st-3rd degree relationships, including differentiating half-sibships from avuncular relationships. Accurate reconstruction of pedigrees holds promise for tracing allele frequency trajectories, improved phasing, and other tangible population genetic and genomic questions.

## PgmNr 1647: *De novo* assembly and characterization of a breast cancer reference sample using multiple whole genome sequencing technologies.

### Authors:

C. Xiao <sup>1</sup>; W. Xiao <sup>2</sup>; Somatic Mutation Working Group of the SEQC2 Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NIH, 45 Center Drive, Bethesda, MD 20894; 2) CDRH/FDA, 10903 New Hampshire Avenue, Silver Spring, MD 20993

---

The advancements of DNA sequencing technology and genome assembly provide an unprecedented opportunity to comprehensively study the instability of cancer genome, and accurately pinpoint the full spectrum of the underlined genomics changes that contribute to cancer development, thereby uncovering important genetic markers for cancer diagnostic and drug development. Here we present a framework that combines multiple sequencing technologies, including Illumina short reads, 10X Genomics linked reads, Pacbio long reads, and Hi-C (high-throughput chromosome conformation capture) reads, to reconstruct the genomes of a paired human triple negative breast cancer (TNBC) cell line and a matched normal cell line. The reference-grade *de novo* assembly of normal sample consists of 1,715 scaffolds with 2.9 Gb, of which 2.69 Gb is from top 50 scaffolds, and is the most contiguous so far with a scaffold N50 size of 69.96 Mb, in comparison to 67.79 Mb for GRCh38, and 44.85 Mb for AK1, respectively. Conversely, the *de novo* assembly of the tumor sample comprises 2,808 scaffolds with a total basepairs of 2.85 Gb and a scaffold N50 size of 27.57 Mb. We have identified a set of high confidence heterozygous sites using short reads mapping to the normal genome, then utilized the long reads to perform phasing using multiple phasing methods, the accumulated length of the phased blocks is 2.23 Gb. The qualities of the *de novo* assemblies have been carefully assessed with BUSCO, QUAST, AUGUSTUS, and GRCh38 alignments. Among 50,052 RefSeq NM transcripts, 49,467 transcripts (98.82%) have been mapped on to the *de novo* assembly of the normal sample with at least 95% alignment coverage, indicating that the high quality of the assembled normal genome has been achieved. We have also mapped all the sequencing reads to the assembly of the normal sample, and in combination with alignments from tumor assembly to normal assembly mapping, we have identified a list of personal-genome based somatic events including SNVs, indels, and structural variations. The well-characterized tumor-normal pair genomes from this work will enable us to provide more insights into understanding the tumor biology of breast cancer.

## PgmNr 1648: LFMD: A new likelihood-based method to detect low-frequency mutations without molecular tags.

### Authors:

R. Ye<sup>1,2,7</sup>; J. Ruan<sup>2,7</sup>; X. Zhuang<sup>3</sup>; Y. Qi<sup>2,7</sup>; Y. An<sup>2,7</sup>; J. Xu<sup>2,7</sup>; T. Mak<sup>4</sup>; X. Liu<sup>2,7</sup>; H. Yang<sup>2,6,7</sup>; X. Xu<sup>2,7</sup>; L. Baum<sup>1,4,5</sup>; C. Nie<sup>2,7</sup>; P.C. Sham<sup>1,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; 4) Center for Genomic Sciences, The University of Hong Kong, Hong Kong, China; 5) State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China; 6) James D. Watson Institute of Genome Sciences, Hangzhou, China; 7) China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

---

As next generation sequencing (NGS) and liquid biopsy become more prevalent in clinical and research area, especially cancer diagnosis, targeted therapy guidance and disease surveillance, there is an increasing need for better methods to reduce cost and to improve sensitivity and specificity. Since the error rate of NGS is around 1%, it is difficult to identify mutations with frequency lower than 1% accurately and efficiently because of low Signal-to-Noise Ratio (SNR). Here we propose a likelihood-based approach, low-frequency mutation detector (LFMD), combining the advantages of duplex sequencing (DS) and bottleneck sequencing system (BotSeqS) to maximize utilization of duplicate sequenced reads. Compared with DS, the new method achieves higher sensitivity (improved ~16%), higher specificity and lower cost (reduced ~70%) without involving additional experimental steps, customized adapters and molecular tags. In addition, this method can also be used to improve sensitivity and specificity of other variant calling algorithms by replacing a step in traditional NGS analysis: removing polymerase chain reaction (PCR) duplication. Thus, LFMD can be a promising method used in genomic research and clinical fields.

## PgmNr 1649: Statistical analysis of spatial expression pattern in spatially resolved transcriptomic studies.

### Authors:

J. Zhu <sup>1</sup>; S. Sun <sup>1</sup>; X. Zhou <sup>1,2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI

---

Recent development of various spatially resolved transcriptomic techniques have enabled gene expression profiling with spatial information. Identifying genes that display spatial expression pattern in these studies is an important first step towards characterizing the spatial transcriptomic landscape. However, as we will show here, existing approaches for detecting spatially expressed genes often fail to control for type I error, can suffer from substantial power loss, and/or are computationally inefficient for even moderate sized data sets. Here, we developed a new method, SPARK (Spatial Pattern Recognition via Kernels), for statistical analysis of spatial expression pattern that addresses these previous limitations. SPARK directly models count data generated from various spatial resolved transcriptomic techniques. With a new penalized quasi-likelihood based algorithm, SPARK is scalable to data sets with tens of thousands of genes measured on tens of thousands of spatial locations. Importantly, SPARK relies on newly developed statistical formulas for hypothesis testing, producing well-calibrated  $p$ -values and high power. We illustrate the benefits of SPARK through extensive simulations and in-depth analysis of four published data sets. In the real data applications, SPARK is 1.5-10 times more powerful than existing approaches and identifies new genes that reveal the importance of neuronal migration in the formation of the olfactory system as well as the importance of immune system and cytoskeleton in tumor progression and metastasis.



## PgmNr 1650: CentroFlye: Assembling centromeres with long error-prone reads.

### Authors:

A. Bzikadze <sup>1</sup>; P.A. Pevzner <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Graduate Program in Bioinformatics and Systems Biology, University of California, San Diego, CA, USA; 2) Department of Computer Science and Engineering, University of California, San Diego, CA, USA

---

Long-read technologies (such as Pacific Biosciences and Oxford Nanopore) have greatly increased the contiguity of genome assemblies as compared to short-read technologies. However, although the existing long-read assemblers, such as Falcon, Miniasm, Flye, HINGE, Canu, Marvel, and wtdbg2, have been used in many sequencing projects, they typically fail to resolve long segmental duplication and long tandem repeats. We focus on the latter challenge – assembling long tandem repeats, and specifically on centromere assembly, the problem that was viewed as intractable until recently. Centromeres represent longest tandem repeats in human genome (also known as *satellite DNA*) that are often repeated thousands of times with extensive variations in copy numbers in the human population. Centromeres are the biggest gaps in the current human genome assembly. As the result, studies of associations between sequence variations and genetic diseases currently ignore  $\approx 3\%$  of the human genome. Although variations in centromeres have been linked to cancer and infertility, they still represent the “dark matter of the human genome” that has withstood all previous attempts to develop an automated tool for centromere assembly. Although long read technologies facilitated partial closing of some centromere gaps in the human assembly, no software tool for centromere reconstruction has been described yet and it remains unclear how accurate are the centromere reconstructions that resulted from previous semi-manual efforts. To assemble centromere and other “dark matter of the human genome”, the recently established Telomere2Telomere consortium aims to generate the first complete assembly of the human genome. We describe centroFlye, the first centromere assembly tool, apply it for assembling the human X chromosome using long error-prone reads, and evaluate the accuracy of the resulting assembly. Our analysis opens a possibility to automatically close the remaining multi-megabase gaps in the human genomes.

# PgmNr 1651: Value of the variations in noisy reads: Towards assembling human centromeres.

## Authors:

Y. Suzuki; S. Morishita

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, Japan

---

We applied single molecule real-time (SMRT, or PacBio) sequencing, which is capable to observe methylation status along the long reads, for observing personal diploid methylomes, CpG methylome pairs of homologous chromosomes that are distinguishable with respect to phased heterozygous variants (PHVs). The task has been considered challenging due to scarcity of PHVs (<1% within genes) and high error rate of sequencing reads which was as high as 15%. We demonstrated, however, the phasing error rate of CpG sites can be reduced to 1% if the positions of PHVs are given. With this method, we examined *GNAS* complex locus known for a combination of maternally, paternally, or biallelically expressed isoforms, and observed allele-specific methylation pattern almost perfectly reflecting their respective allele-specific expression status, demonstrating the merit of elucidating comprehensive personal diploid methylomes and transcriptomes[1].

The success of the method indicated that, to distinguish highly homologous regions in a genome, we could utilize variations much less frequent than sequencing error rate, if the variant sites are known beforehand. Thus, we extended the logic to calculation of read overlaps, designing an optimal scoring scheme under the given variation model. Then, we applied it to *de novo* assembly task to see whether it would be able to resolve higher order repeats (HOR) arrays making up human centromeres, which has been considered one of the *un-sequenceable* parts of the genomes. We assembled the chromosome X where >99% identical 2 kbp HORs are tandemly arranged >1000 times (each unit is characterized by a specific combination of 12 alpha-satellite monomers). We elucidated structures of several regions of >100 kbp. Comparison of the results from several publicly available PacBio datasets informed us the evolutionary patterns of centromeric regions. Therefore, we concluded that not only generating longer reads, which may span and thus trivialize the whole repeats, but also handling small variations within repeats in a proper theoretical way is another promising road to complete genomes.

## References:

[1] Suzuki, Yuta, et al. 'A statistical method for observing personal diploid methylomes and transcriptomes with single-molecule real-time sequencing.' *Genes* 9.9 (2018): 460.

## PgmNr 1652: Behind the scenes of submission processing at ClinVar.

### Authors:

M.J. Landrum; S. Chitipiralla; G. Brown; C. Chen; B. Gu; J. Hart; D. Hoffman; W. Jang; K. Kaur; C. Liu; V. Lyoshin; Z. Maddipatla; R. Maiti; J. Mitchell; G. Riley; W. Shi; G. Zhou; D.R. Maglott; J.B. Holmes; B.L. Kattman

[View Session](#) [Add to Schedule](#)

**Affiliation:** NCBI/NLM/NIH, Bethesda, Maryland.

---

ClinVar is a public database of submitted interpretations of variants and their relationships to disease, maintained by the National Center for Biotechnology Information (NCBI) at the National Institutes of Health (NIH). Many kinds of data are submitted to ClinVar by many types of organizations. Thus the ClinVar team uses both automated and manual steps to validate, aggregate, and add value to submitted data. Automated validation includes checking consistency of variant and gene data; validating database identifiers for conditions; and validating the required fields of clinical significance, affected status, allele origin, and collection method. ClinVar curators manually review submissions, e.g. to confirm the submitter's intent for rare combinations of data; to request the date that the interpretation was last evaluated; and to check consistency of data in comments and structured data. The curators communicate with ClinVar submitters in an iterative process to ensure that each submission can be processed and that it represents the data as the submitter intends. Each submitted record is assigned an SCV accession number, which is used for retrieval and to provide updates for the record. Once processed, submitted data are aggregated such that a user can review all data submitted for a variant (the VCV record) or a variant-condition pair (the RCV record) together. ClinVar adds value to RCV and VCV records by providing an identifier for the variant or set of variants that was interpreted (Variation ID); calculating HGVS expressions on genomic DNA, cDNA, and protein; calculating genomic location on GRCh37 and GRCh38; calculating a predicted molecular consequence; annotating the gene(s) that colocalizes with each variant; annotating allele frequency; calculating an aggregate clinical significance and indicating consensus or conflict; calculating a review status; and providing links to other databases. Upcoming improvements to submission processing include pre-validation of file submissions in the ClinVar Submission Portal. This feature will give the submitter quick feedback on the validity of variant descriptions when the submission file is uploaded. Validation for other types of data will be added in later phases. Validation of data upon upload will allow submissions to be processed faster. This will make submitted data publicly available in ClinVar sooner, so that clinical laboratories and other users have access to accurate and up-to-date information.

## PgmNr 1653: Accurate and high-resolution copy number variant detection in clinical germline screening.

**Authors:**

J. Li; K. Haas; S. Hong

[View Session](#) [Add to Schedule](#)

**Affiliation:** Myriad Women's Health, South San Francisco, CA 94080

---

**Introduction:** Structural variants, especially copy number variants (CNVs), can cause genomic disorders. CNVs can bring genomic coordinates that are distant in the reference genome into contact in a sample genome, forming “breakpoints.” Determining exact CNV breakpoint sequences (physical deletion or duplication boundaries) across individuals is crucial for clinical germline screening and associating genotype to phenotype. For targeted sequencing, the majority of the current tools rely only on the depth of coverage, which reports approximate genomic coordinates only. Here, we report a hierarchical approach for detecting high-resolution CNVs that integrates read-depth signal and reads that align in split fashion to discontinuous loci in the reference genome (split-reads) containing breakpoints.

**Methods:** Our approach first applied a Hidden Markov model (HMM) CNV caller that calls CNVs by observing sequencing depth, which reflects the copy number, for ~100 base-pair segments tiled along the targeted region of interest (ROI). CNV breakpoints were determined by the probe coordinates that targeted the CNV region in the HMM CNV caller. To generate the high-resolution CNV breakpoints, we developed a CNV breakpoint annotator to identify split reads that have soft-clipped bases on the side on which the depth-based CNV is located and replaced the CNV breakpoints with breakpoints defined by supporting split-reads.

**Results:** We first validated our HMM CNV caller and it achieved 100% concordance, emitting positive calls for 50 orthogonally confirmed CNVs and negative calls for 685 samples not known to have CNVs in the ROI. We then validated our breakpoint annotator by comparing the performance to open-source CNV caller LUMPY on 1,700 random selected in-house patient samples. We identified 310 CNVs in these samples through the HMM CNV caller. Our approach refined breakpoints for 70 (22.6%) CNVs, and 67 were confirmed by LUMPY. Confirmation analysis showed that 69 of the 70 breakpoints identified by our approach were accurate. Together our hierarchical approach showed high CNV calling accuracy and also provided precise breakpoints.

**Conclusions:** The approach described here for detecting high-resolution CNVs ensured accurate calling of important CNVs and reduced the burden of manually refining the CNV breakpoints in the downstream variant interpretation. With split reads, the accuracy of calls increased resolution from 100-300bp to 1-3bp.

## PgmNr 1654: Simple ClinVar: An interactive web server to explore and retrieve gene and disease variants aggregated in ClinVar database.

### Authors:

E. Perez-Palma <sup>1</sup>; M. Gramm <sup>2</sup>; P. May <sup>3</sup>; P. Nuernberg <sup>2</sup>; D. Lal <sup>1,2,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Medicine Institute, Cleveland Clinic, Cleveland, USA; 2) Cologne Center for Genomics, University of Cologne, Cologne, Germany; 3) Luxembourg Centre for Systems Biomedicine, University Luxembourg, Esch-sur-Alzette, Luxembourg; 4) Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, USA; 5) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

---

Clinical genetic testing has exponentially expanded in recent years, leading to an overwhelming amount of patient variants with high variability in pathogenicity and heterogeneous phenotypes. A large part of this information is aggregated in public databases such as ClinVar. However, the ability to explore this rich resource and answer general questions such as “How many genes inside ClinVar are associated with a specific disease?”, “In which part of the protein are patient variants located?” or “has a ClinVar variant been observed in the general population” is limited and requires advanced bioinformatics processing. Here, we present Simple ClinVar (<http://simple-clinvar.broadinstitute.org/>) a web-server application that is able to provide variant, gene, and disease level summary statistics based on the entire ClinVar database in a dynamic and user-friendly platform. The ClinVar database is downloaded from the ClinVar ftp site and processed internally to produce a pre-filtered ClinVar file for the user to explore on the web-server. The pre-filtering step is designed to reduce the complexity of ClinVar entries as well as to provide fast access to high-quality entries. Interactive summary statistics, variant mapping, and visualization were developed with the Shiny framework of R studio software. App deployment, hosting, and update is performed with Google Cloud services. At the time of abstract submission, our web server holds information for 499,389 ClinVar entries. Summary statistics are available for 18,523 genes and 11,265 unique phenotypes. Overall, our web application is able to interactively answer basic questions regarding genetic variation and its known relationships to disease. By typing a disease term of interest, the user can identify in seconds the genes and phenotypes most frequently reported to ClinVar. Subsets of variants can then be further explored, filtered, and visualized in the corresponding protein sequences. In addition, variants reports are connected to external resources such as the genome aggregation database (gnomAD) and commonly used variant interpretation scores (e.g. CADD, Polyphen). Finally, we enabled a 'submission' feature that allows the user to map their own variants on top of ClinVar reports. Our website will follow ClinVar monthly releases and provide easy access to ClinVar resources to a broader audience including basic and clinical scientists.

# PgmNr 1655: Improving the informativeness of Mendelian disease pathogenicity scores for common diseases and complex traits.

## Authors:

S.S. Kim <sup>1,2,3</sup>; K. Dey <sup>2,3</sup>; O. Weissbrod <sup>2,3</sup>; C. Marquez-Luna <sup>4</sup>; S. Gazal <sup>2,3</sup>; A.L. Price <sup>2,3,5</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

---

Despite considerable progress on pathogenicity scores prioritizing both coding and non-coding variants for Mendelian disease (Eilbeck et al. 2017 Nat Rev Genet), little is known about the utility of these pathogenicity scores for common disease. We sought to answer two questions: (1) how informative are Mendelian disease pathogenicity scores for common disease, and (2) how can we improve their informativeness for common disease?

To answer the first question, we applied stratified LD score regression with the baseline-LD model v2.1 (86 genomic annotations; Gazal et al. 2017 Nat Genet) to assess the informativeness of new annotations based on top prioritized variants from various pathogenicity scores (e.g. CADD, ReMM, Eigen) across 41 independent common diseases and complex traits (average  $N=320K$ ), meta-analyzing the results across traits. We considered two metrics: enrichment = (% of heritability)/(% of SNPs) and  $\tau^*$  = proportionate change in per-SNP heritability per 1 s.d. change in annotation value conditional on baseline-LD model annotations. Many pathogenicity score annotations were highly informative, e.g. enrichment=19.5 ( $P<1e-20$ ) and  $\tau^*=0.82$  ( $P<7e-19$ ) for the top 0.5% of SNPs from Regulatory Mendelian Mutation score (ReMM; Smedley et al. 2016 AJHG), while others were conditionally uninformative, e.g. enrichment=5.7 ( $P<2e-12$ ) and  $\tau^*=0.07$  ( $P=0.04$ ) for the top 1% of SNPs from Eigen-PC (Ionita-Laza et al. 2016 Nat Genet); proportions of top SNPs were optimized on a coarse grid to maximize informativeness.

To answer the second question, we developed a variant reprioritization framework to predict/impute and reprioritize existing pathogenicity scores using a supervised machine learning approach, gradient boosting. We used annotations from the baseline-LD model as features and subsets of scored variants as training data, reprioritizing pathogenicity scores without using disease data. Gradient boosting attained high predictive accuracy (AUPRC=0.74-0.97) and substantially increased informativeness for common disease both for previously informative scores, e.g. enrichment=17.0 ( $P<1e-20$ ) and  $\tau^*=1.17$  ( $P<1e-20$ ) for the top 1% of SNPs from boosted ReMM (conditional on baseline-LD + existing ReMM annotations), and for previously uninformative scores, e.g. enrichment=6.3 ( $P<4e-32$ ) and  $\tau^*=0.67$  ( $P<9e-7$ ) for the top 5% of SNPs from boosted Eigen-PC. The newly reprioritized scores have high potential to improve gene discovery and fine-mapping for common disease.

## PgmNr 1656: Polygenic risk score ensembles across methods and diseases increase disease prediction accuracy.

### Authors:

S. Kulm; J. Mezey; O. Elemento

[View Session](#) [Add to Schedule](#)

**Affiliation:** Physiology, Biophysics and Systems Biology, Weill Cornell Medicine, New York, NY

---

Predicting an individual's risk to a disease may lead to more precise prescription dosages, screening schedules and lifestyle adjustments – changes that improve disease treatment. The current genetics-based disease prediction paradigm is the polygenic risk score (PRS). Multiple methods have emerged to produce PRSs that limit various sources of noise. Each of these methods has been evaluated upon only a few diseases and compared against only a few other competing methods, and never within any ensemble. This work evaluates a total of 11 methods on 16 different diseases. Scores were generated from GWAS summary statistics and evaluated on the UK Biobank. There was no clear superior method on all traits, however the simpler methods (clumping and LDpred) had the top average performance. Analysis of each trait revealed a wide distribution in predictability measures. Ranking traits by full model AUC, including covariates such as sex and age along with the PRS, showed age-related traits such as gout and psoriasis are the most predictable (AUCs = 0.79, 0.71). Ranking by AUC Improvement, the difference between the covariate only and full model, showed that autoimmune traits such as celiac and inflammatory bowel disease received the greatest benefit from adding PRS (AUC Improvement = 0.17, 0.14). To utilize the 1,792 scores produced among all parameters, elasticNet models were generated that ensemble together either various models or diseases. The disease ensemble model showed poor performance with only 6 diseases having increased accuracy. Simple decision tree models were also generated on multiple diseases yet showed only a slight increase in accuracy. The methods ensemble model showed strong performance with 14 diseases exhibiting increased accuracy, and with 6 traits seeing AUC improvements greater than 0.1. This increased performance indicates that individual methods are not capturing a disease's full genetic architecture, whereas, ensembles are able to partially overcome this complexity - illustrating their clear utility in disease prediction. The disease ensembles suggest that as ensembles get more advanced, disease prediction accuracy will increase further.



## PgmNr 1657: Systematic comparison of different evidence sources for predicting GWAS effector genes.

### Authors:

B. Alexander <sup>1</sup>; M. Costanzo <sup>1</sup>; M. Duby <sup>1</sup>; J. Massung <sup>1</sup>; J. Flannick <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Institute of Harvard and MIT, Cambridge, Massachusetts; 2) Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts

---

While genome wide association studies (GWAS) have identified an ever-increasing number of complex trait-associated genomic loci, gleaned biological insight from these associations has proven difficult. A complete mechanistic understanding of a GWAS association describes a causal variant, an effector gene that mediates its effects, a tissue or cell type of action, and an impaired biological process. Many analytical techniques and datasets are often used to heuristically prioritize variants or genes; e.g., observation of an epigenomic annotation overlapping a variant, an expression quantitative trait locus (eQTL) coincident with the disease association, or enrichment of a biological process for genetic associations. A natural question is therefore: how consistent are these evidence sources with one another, and how are prioritizations affected when they are viewed alongside one another? Here, we systematically compare >15 sources of evidence for effector genes (including MAGMA, MetaXcan, DEPICT, eCAVIAR, and model organism phenotypes) across >50 complex traits. We present a visualization framework, FOCUS (Find Optimal Computational Support), which provides a standardized means to compare any set of methods or datasets that prioritize variants, genes, and/or tissues (or relationships among them). Application of FOCUS immediately emphasizes the limitations of coincident eQTL analysis to prioritize genes: while for many loci, promising candidate effector genes are the target of a coincident eQTL, in most cases (a) many other genes within the same locus are also the target of a coincident eQTL; (b) the eQTL also exists across many tissues with no evidence of disease-relevance; and (c) three different eQTL co-localization methods produce significantly different results. At the type 2 diabetes (T2D) *GPSM1* locus, for example, *DNLZ* initially seems a promising effector gene, based on a rare variant association and a coincident eQTL (predicted by eCAVIAR) in the pancreas. However, two other colocalization methods applied to the same data (COLOC and MetaXcan) do not replicate this prediction. These results suggest that current methods for prioritizing effector genes can be misleading when viewed in isolation or when only results for candidate genes are examined. FOCUS is available for use via a public web portal for T2D and related traits ([type2diabetesgenetics.org](http://type2diabetesgenetics.org)), with plans to extend it to more complex diseases in the next year.



## **PgmNr 1658: Recurrency approaches using Random Forests to identify genetic risk factors while controlling family-wise false positive rates.**

### **Authors:**

J.E. Bailey-Wilson; J. Sabourin; A.M. Musolf; E.R. Holzinger; J.D. Malley

[View Session](#) [Add to Schedule](#)

**Affiliation:** Computational and Statistical Genomics Branch, NIH/NHGRI, Baltimore, MD

---

Non-parametric machine learning methods are robust to complex models such as are hypothesized for complex genetic diseases and traits. Machine learning methods such as Random Forests (RF) can identify complex effects. We have developed an RF-based variable selection method called relative recurrency variable importance metric (r2VIM) to address hurdles in using RF to identify variables (features) that increase risk of complex phenotypes in genome-wide analyses.

Previously, we have shown that r2VIM improves false positive control and power to identify important features compared to RF, especially when non-linear effects contribute to the phenotype. The relative recurrency approach estimates the probability that a feature is important. Here, we use permutation of trait status to control family-wise error rates in variable selection to achieve a selection criterion that better scales with the number of SNPs evaluated and compare performance of several schemes for maximizing power to detect causal variants in simulated data. These data were simulated using GWAsimulator and the 1000 genomes European (EUR) haplotypes to create a set of subjects with linkage disequilibrium structure similar to the actual EUR population, pruned to match the SNP density of a typical 1M SNP chip and creating a trait due to: (1) one interaction between 2 SNPs with no main effects on maps of 10,000 to 100,000 SNPs and (2) both interaction effects and independent main effects. The effects of stepwise procedures on power are also being evaluated. Power is affected by multiple factors, including effect size of the causal features, sample size, number of trees and other parameters of the RF analysis.

This recurrency approach has the potential to elucidate novel biological pathways which could improve both treatment and prediction of complex human diseases.

## PgmNr 1659: Identifying distinguishing gene-level and protein-level features associated with gain-of-function and loss-of-function mutations.

### Authors:

C.S. Bayrak <sup>1</sup>; A. Schlessinger <sup>2</sup>; Y. Itan <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

---

Gain-of-function (GOF) and loss-of-function (LOF) mutations in the same gene may result in different phenotypes/diseases and hence require different treatments. Therefore, identifying the functional consequence of mutations is an important step for understanding disease mechanisms. Currently the only way to do so is performing experimental methods which is costly and time consuming. Although, there are numerous computational tools predicting the pathogenicity/deleteriousness of genetic variants (i.e., CADD, SIFT, PolyPhen-2) that help prioritizing causal variants, these methods cannot predict whether the variant results in gain-of-function (GOF, enhanced protein product) or loss-of-function (LOF, reduced protein product). To address this need, we searched for discriminating gene-level and protein-level features of GOF and LOF mutations. We generated the first database of all currently known disease-causing GOF and LOF mutations using natural language processing (NLP) on the available abstracts of the Human Gene Mutation Database (HGMD). We identified 1,209 GOF mutations in 302 genes and 11,452 LOF mutations in 1,628 genes. We then annotated all GOF and LOF mutations with gene-level and protein-level features including gene ontology terms, mode of inheritance, mutation position, amino acid changes, protein domain, secondary structure, solvent accessibility, and disordered regions. The results indicate significant discriminative power of some features, including mode of inheritance, mutation type, and protein domains, to distinguish GOF from LOF and neutral variants, and suggest a machine learning method to efficiently classify mutations by functional consequence.

## PgmNr 1660: Interactive, shareable plots of GWAS data with LocusZoom.

### Authors:

A.P. Boughton<sup>1,2</sup>; R.P. Welch<sup>1,2</sup>; D. Taliun<sup>1,2</sup>; S.A. Gagliano Taliun<sup>1,2</sup>; P. VandeHaar<sup>1,2</sup>; G. Abecasis<sup>1,2</sup>; M. Boehnke<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA; 2) Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA

---

Genome-wide association studies (GWAS) are a powerful tool for identifying trait-associated loci. It has long been recognized that GWAS results require context for interpretation. Since 2010, LocusZoom has provided a means to visualize GWAS results overlaid with linkage disequilibrium, recombination rate, and nearby genes, among other annotations. Our web-based service ([locuszoom.org](http://locuszoom.org)) has been used to generate more than 700,000 static plots of user-provided summary statistics. To support more dynamic and exploratory workflows, we have extended LocusZoom from a service that generates static PDF plots into an interactive tool that relies on standard web technologies. Users are now able to interact with results by panning and zooming, or adding and removing annotation tracks for context. We generate a Manhattan plot of the results, which makes it easy to quickly jump to significant variants of interest. The user can also zoom into arbitrary genomic regions, identified by gene name, genomic coordinates, or variant rsID. Multiple GWAS result sets can be visualized together for comparison. Annotations can be customized to include result sets such as NHGRI-EBI GWAS catalog and prior PheWAS results (such as those from analysis of UK Biobank and other rich datasets) or to include genomic features (such as chromatin accessibility or enhancer tracks). These new visualization options can be used in multiple ways: researchers have a choice to upload and explore their data on our remote server (with enhanced functionality) or, alternatively, visualize local files without sending any data ("LocalZoom"). If data are uploaded, plots can be shared publicly (or only with selected collaborators); both options also support exporting plots as SVG images. We are currently extending LocusZoom to natively support on-the-fly analyses that are based on GWAS summary statistics, such as credible set determination or gene-based burden tests. We expect to add additional analysis types and improved access for public datasets in the future, based on user input.

## PgmNr 1661: Analysis of the n-of-one problem in human genetics.

### Authors:

W. Charnig<sup>1,2,3</sup>; A.B. Wilfert<sup>1</sup>; D.F. Conrad<sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Washington University in St. Louis, St. Louis, Missouri.; 2) Department of Neurology, Washington University in St. Louis, St. Louis, Missouri.; 3) Department of Psychiatry, Washington University in St. Louis, St. Louis, Missouri.; 4) Oregon National Primate Research Center, Oregon Health and Science University, Portland, Oregon

---

Assessing the significance of rare variants in a single exome remains challenging even with the expansion of allele frequency information in large population databases. We previously developed a statistical inference system, PSAP (population sampling probability), to evaluate the significance of a genotype observed in a single patient based on *in silico* variant annotations. These p-values are calculated by simulating gene-specific null models and are able to prioritize 39% of disease genotypes as the most damaging unit in over 5 million simulated cases.

In our current study, we describe an analytic approach to calculating these gene-specific null models, introduce the hemizygous inheritance model, and integrate both gender and ethnicity factors. We compare the calibration and performance of PSAPs derived from different variant scoring methods, including FATHMM (Functional Analysis through Hidden Markov Models), MetaLR (Meta Logistic Regression), DANN (Deleterious Annotation of genetic variants using Neural Networks), fitCons (Fitness Consequence), REVEL (Rare Exome Variant Ensemble Learner), and CADD (Combined Annotation Dependent Depletion). We explore how using ExAC (the Exome Aggregation Consortium) and gnomAD (the Genome Aggregation Database), which contains twice sample size of ExAC, to parameterize our null models influences the behavior of PSAP. Finally, we evaluate the performance of our updated PSAPs in true n-of-one scenarios through simulation by “spiking-in” HGMD or ClinVar variants into the genomes of control individuals.

Using our analytic approach, PSAP p-values have improved resolution as compared to the original, implicating better power to identify variants of significance. Ethnicity-specific and gender-specific PSAPs both provide better calibration and performance, whereas we observe little difference when using ethnicity-match frequency parameters from gnomAD as compared to ExAC. We also found that using CADD as the scoring parameter in our PSAP calculations provides the most comprehensive annotation of the genome, and the best null-model calibration and performance. Using our updated PSAPs, we are able to identify 52% of disease genotypes as the most damaging unit in over 5 million simulated cases. PSAP is currently being used to identify potential pathogenic variants in studies of sporadic cases of disease, including ciliopathies and male infertility.

## PgmNr 1662: Representation learning for variant interpretation.

**Authors:**

R. Chung <sup>1</sup>; N. Ioannidis <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Center for Computational Biology, University of California, Berkeley, Berkeley, CA.; 2) Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA

---

Despite efforts to sequence and catalogue genetic variation within human populations, predicting the clinical and functional effect of a given variant is still difficult, particularly in noncoding regions of the genome. To understand this variation, computational tools have been developed to predict whether missense single nucleotide variants (SNV's) disrupt protein function and whether noncoding SNV's are associated with differences in gene expression profiles. Many such tools rely on classifiers built using manually curated genetic and epigenetic predictive features. Although many SNV's affect the expression of distant genes, most features used for variant effect prediction are local to the SNV.

In this work, we use deep learning based language models to generate features that capture the larger genomic context surrounding a variant of interest. To build these features, we train language models on the entire human reference sequence, using k-mers of varying lengths as tokens. We explore methods to optimize these features by fine-tuning them on several different genetic and epigenetic annotations. We evaluate the use of these learned representations as predictive features, either alone or in combination with canonical genomic annotations, for variant effect prediction tasks, including expression quantitative trait loci (eQTL) prediction. Our results demonstrate the importance of the surrounding genetic context for variant interpretation and the use of feature representations from language models to capture this information.

## PgmNr 1663: A toolkit for accelerating genomic analysis using NGS index formats.

**Authors:**

J. Didion; S. Osazuwa; A. Zalcman; N. Thangaraj

[View Session](#) [Add to Schedule](#)

**Affiliation:** DNAnexus, Mountain View, CA.

---

As genomic technologies move from the research laboratory to the clinic, the scale of genomic data is growing at an increasingly rapid pace. It is critically important for researchers, clinicians, and patients that data analysis pipelines do not become a bottleneck to diagnosis and treatment. Acceleration of next-generation sequencing (NGS) pipelines requires creative approaches to optimization that are informed by simultaneous consideration of biology, informatics, and systems architecture, especially in cloud-based computing environments. Common index formats, such as BAM Index (BAI) and Tabix (TBI), contain coarse-grained information on the density of NGS reads along the genome that may be leveraged for rapid approximation of read depth-based metrics. We present IndexTools, an open-source toolkit for extremely fast NGS analysis based on index files. We demonstrate that IndexTools 1) substantially accelerates parallel processing of BAM and VCF files by optimizing file splitting based on estimated read density; 2) provides reasonably accurate estimates of genome coverage (similar to indexcov, Pedersen et al., 2017); 3) is able to infer relative sex chromosome and mitochondrial genome number relative to the autosome; 4) is able to accurately call large deletions and copy-number expansions in minutes (compared to hours for traditional CNV pipelines). We further show that small variant and CNV calling pipelines implemented using IndexTools on the DNAnexus platform save considerable time and cost compared to equivalent non-accelerated pipelines. We expect that IndexTools will substantially reduce the turn-around time to deriving insight from NGS data.

## PgmNr 1664: Simultaneous multi-sample approaches significantly improve the accuracy of RNA-seq data analysis.

### Authors:

L. Florea<sup>1,2</sup>; G. Yang<sup>2</sup>; L. Song<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

- 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD;
- 2) Department of Computer Science, Johns Hopkins University, Baltimore, MD

---

Transcript assembly from RNA-seq reads and differential splicing (DS) analysis are two key objectives of transcriptomics applications. However, despite significant advances in designing accurate analysis tools, both transcript reconstruction and DS detection are not meeting their full potential. The current framework for RNA-seq analysis, where each sample is processed independently and results are merged across all samples, improves coverage but inherits the errors of individual data sets, limiting both accuracy and the ability to identify splice variants, in particular rare and low expression events. Parting with the traditional one-at-a-time approach, we developed two tools that simultaneously analyze all samples in an RNA-seq collection to produce significantly improved results. The first tool, PsiCLASS, is a novel transcript assembler that combines features from multiple samples into a unified splice graph data structure, which increases accuracy and consistency. Its algorithmic underpinnings include mixture statistical models for cross-sample subexon selection combined with splice graph based dynamic programming algorithms and a weighted voting scheme for transcript selection. PsiCLASS achieves significantly better sensitivity-precision tradeoff and precision up to 2-3 fold higher than the StringTie system and Scallop plus TACO, the two best current approaches. It is also highly scalable, assembling 667 GEUVADIS samples in 9 hours, and has robust accuracy with large numbers of samples. The second tool, MntJULiP, extracts a reliable set of introns from a set of RNA-seq samples and detects introns differentially present between groups of samples. In particular, it uses a Bayesian approach to model read counts of introns under a negative binomial model, combined with hypothesis testing to determine introns that are differentially *expressed* between conditions, and a Dirichlet multinomial model to detect differential *usage* (splicing) of introns within a gene. MntJULiP compares favorably in accuracy to existing DS tools, including LeafCutter, MAJIQ, rMATS and Junction-seq, and scales up to data sets of hundreds and thousands of samples. Collectively, PsiCLASS and MntJULiP represent a new generation of tools that leverage the latent information in a collection of related RNA-seq samples to enable more accurate, highly scalable and efficient transcriptomic analyses. Availability: <https://github.com/splicebox>. Funding: NSF DBI-1356078 and NIH R01-GM124531.

## PgmNr 1665: Predicting missense variant pathogenicity with C-MIP.

**Authors:**

P. Fradkin <sup>1</sup>; D. Merico <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Deep Genomics, Toronto, ON, Canada; 2) The Hospital for Sick Children, Toronto, ON, Canada

---

More than a third of rare and pathogenic mendelian variants are missense, resulting in an amino acid substitution. However, not all missense mutations are deleterious and accurately deciphering variant impact on protein function remains a challenge. Advancement of machine learning paradigms and increased availability of relevant genomic data have lead to the development of several novel methods to address this. Specifically, three core ideas have driven the development of these tools: multiple sequence alignments to evaluate conservation, prediction ensembles, and learning complex motif patterns using neural networks. We evaluated eight recent predictors that applied one or more of these core ideas (EVMutation, PrimateAI, Envision, MutationAssessor, MCAP, Polyphen2 HDIV, Polyphen2 HVAR, SIFT). We also introduce a new method, ClinVar Missense Impact Predictor (C-MIP), a gradient boosting ensemble predictor. We validated the performance of these models using pathogenicity classification from ClinVar and several deep mutagenesis experiments. We found that EVMutation, a method that utilized multiple sequence alignments and utilizes deep learning, outperformed other predictors. However, this method was limited to 18% of the proteome. Where EVMutation is not defined, C-MIP outperformed other methods. We compiled a comprehensive evaluation framework of these methods across four independent datasets and highlight the performance of EVMutation and C-MIP for predicting variant impact on protein function.



## PgmNr 1666: PEMapper2 and PECaller2 improve trio-based genotype calling results and allow for direct comparisons of cell-line and blood derived WGS samples.

### Authors:

H.R. Johnston <sup>1</sup>; M.R. Bishop <sup>1</sup>; P. Chopra <sup>1</sup>; S. Gabriel <sup>2</sup>; E.J. Leslie <sup>1</sup>; S.L. Sherman <sup>1</sup>; D.J. Cutler <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory Univ School of Medicine, Atlanta, Georgia.; 2) The Broad Institute, Cambridge, Massachusetts.

---

The original versions of PEMapper and PECaller allowed for fast and accurate analysis of whole-genome sequence data in an unbiased fashion. Here we present PEMapper2 and PECaller2. Each program has been improved, allowing for faster, more accurate analysis. PEMapper2 is improved primarily by shrinking output file sizes, enabling large savings of disk space in big studies. Building on the naïve calling models introduced in PECaller, PECaller2 includes dramatically improved priors for genotype models, more precise handling of model parameters, and improved detection of model parameter misconvergence. As a result, PECaller2 genotype calling for both SNPs and INDELs is significantly improved, and demonstrably better than GATK best practices.

For SNP calling, PECaller2 and GATK give remarkably similar results, but to the extent that they differ (which is rare), there is strong evidence that PECaller2 makes better calls. In a dataset of 395 European trios with over 33 million segregating sites, we observed 99.95% concordance in genotype calls; 43.9 billion genotypes were identical and 21.2 million were called differently. Of those differing calls, there were 19.2 million Mendelian inconsistencies in the GATK results, and only 0.74 million in the PECaller2 data, suggesting that when PECaller2 and GATK calls differ, the error occurs with GATK ~95% of the time.

The differences between GATK and PECaller2 are particularly noticeable for INDEL calling. Only 58% of sites called as INDELs by GATK are confirmed by PECaller2. At sites called an INDEL by both GATK and PECaller2, genotype calling agrees ~99% of the time. However, INDELs called by GATK produce 27.3 million Mendelian errors, or ~1 error per 11.5 INDEL calls. PECaller2 is more conservative in its calls and results in 63x fewer errors than GATK.

PECaller2 can also be used with a variety of sources of DNA. In our experiment, sequence was obtained at 30x depth from DNA derived from blood and from low-passage cell lines taken from 10 unrelated individuals. Data were mapped and called in two separate batches, standard QC applied, and final datasets compared. 99% of all sites called a SNP in one experiment were also called a SNP in the other. Conditional on a site being called a SNP in both experiments, the genotype calls agreed 99.98% of the time. For INDEL calling, only 90% of INDELs were called in both experiments, but conditional on the INDEL being called in both, the genotypes agreed of 99.5% of the time.

## PgmNr 1667: DeepClean: Correcting sequencing errors using unique molecular barcode based deduplication method.

### Authors:

S. Katiyar; K. Hyuk Jung; L. Sunghoon

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bioinformatics, Eone Diagnostics Genome Center, Incheon, Incheon, Korea, Republic of

---

A key objective of high-throughput sequencing experiments is the exact quantification of the total number of DNA molecules. However, PCR amplification produces multiple copies of the same molecule and it becomes challenging to distinguish among identical copies of the distinct DNA molecules. Unique Molecular Identifier (UMIs) are short oligonucleotide sequences, made by the random combination of nucleotides. One UMI is attached to one original DNA molecule and is carried over to all of the PCR amplified molecules. Because UMIs are randomly generated, identical copies from distinct original DNA molecules will have different UMIs and can be distinguished accurately. The introduction of UMIs allows very simple and accurate deduplication of PCR sequences. However, this simple process is challenged due to the erroneous nature of sequencing that introduces: i) sequencing errors in UMIs and ii) sequencing errors in the replicates of the same DNA molecule. In the context of the variant calling, errors in UMIs lead to the wrong quantification of DNA molecules (inaccurate variant allele-frequency) and errors in the replicates introduce false positive variants. We have developed a PCR sequence deduplication algorithm that overcomes both the challenges described above. The algorithm is implemented via software package DeepClean, written in python language. First, DeepClean accommodates the errors in UMIs by using a tree-based network model. DNA replicates that belong to the same UMI tree are clustered together. In the second step of error correction, DeepClean eradicates the errors among replicates of a cluster, based on the error-frequency and reference-base to alternate-base ratio. Improvement in quantification accuracy is demonstrated using both the simulated data and Tru-Q7 (1.3% Tier) reference standard cell line data. We also compared DeepClean with already published similar tools. DeepClean significantly reduced the general error rate post deduplication, as compared to PCR amplified data. DeepClean produced data showed the allele-frequency of Tru-Q7 expected variants with high accuracy along with the comparatively lower number of false positive variants. DeepClean will be made available to the public as an open source software package.

## PgmNr 1668: Exploring the feasibility of data augmentation while using smaller biobank data sets.

### Authors:

C.J. Lee <sup>1</sup>; A.R. Hsieh <sup>2</sup>; C.S.J. Fann <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Biomedical Sciences, Academia Sinica, Taiwan, R.O.C.; 2) Department of Public Health, China Medical University, Taiwan R.O.C.

Empowered by new computing technology and low genotyping cost, large biobank projects like UK Biobank (UKB) have had fruitful results in the advancement of biomedical sciences. However, there are several smaller biobanks sampling from different ethnic groups and the statistical power to detect any association from these datasets is lower. Data augmentation by synthesizing unobserved samples show promising results in the application of machine learning algorithms. Here, we hypothesized that augmentation of small biobank data can increase statistical power and detect reliable association signals.

A two-step strategy was adopted. First, control samples were filtered using Partition Around Medoids Algorithm, using the entire phenome to divide controls into clusters according to comorbidity. To reduce the heterogeneity, only samples not in the same cluster for the phenotype of interest were used as controls. Second, cases and controls were stratified by age and gender. By applying Synthetic Minority Oversampling Technique on each stratum, artificial cases and controls were generated. In this study, we chose to use asthma as the phenotype. Dataset from Caucasians in UKB (UKB-C,  $N_{\text{C-total}}=204,893$ ,  $N_{\text{C-case}}=31,303$ ) and a random sample were selected (UKB-CS,  $N_{\text{CS-total}}=24,000$ ,  $N_{\text{CS-case}}=3,612$ ). Fourteen linkage disequilibrium peaks ( $p \leq 10^{-8}$ ) from UKB-C GWAS were used as targets for comparison. Only HLA region was replicated using UKB-CS. Our strategy was then applied to UKB-CS. The real-to-artificial sample ratio (RAR) ranged from 4 (4 real and one artificial sample) to 1. Compared to targets from UKB-C data, 4 peaks were replicated when  $\text{RAR}=4$ , 5 when  $\text{RAR}=3$ , 6 when  $\text{RAR}=2$  and 11 when  $\text{RAR}=1$ . HLA region was prominent for every RAR. When  $\text{RAR}=2$ , false positive peaks seemed modest; almost half of the signals could be replicated when roughly 1/9 of the UKB-C samples were used.

The above procedure was applied to data from Taiwan Biobank (TWB,  $N_{\text{T-total}}=23,942$ ,  $N_{\text{T-case}}=2069$ ). Without augmentation, only HLA region was significant. When  $\text{RAR}=2$  for TWB and UKB-CS, GWAS results showed a similar trend. In addition to HLA region, only two other regions were replicated for TWB. Population heterogeneity may contribute to this discrepancy. Our results showed that data augmentation is promising, however caution needs to be taken with respect to input data quality and possible stratification, etc. More testing of augmentation algorithms should be done to further evaluate for performance.

## PgmNr 1669: New GeneHancer data in the UCSC Genome Browser displaying new interact track type.

**Authors:**

B. Lee; K. Rosenbloom; B. Raney; M. Haeussler; J. Casper; A. Hinriches; R. Kuhn; C. Lee; M. Speir; J. Navarro; L. Nassar; C. Powell; D. Schmelter; H. Clawson; A. Zweig; J. Kent

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genomics Institute, Univ. of Calif. Santa Cruz, Santa Cruz, California.

---

The UCSC Genome Browser has a new interact track format, used in native Browser tracks and available to users for their own custom tracks or hub tracks. Interact is used to display pairwise interactions as arcs or half-rectangles connecting two genomic regions on the same chromosome, and can represent cross-chromosomal interactions. The format can be used to display physical low-density chromatin interactions, such as Chromatin Interaction Analysis by Paired-End Tag Sequencing (ChIA-PET), as well as virtual (regulatory) relationships such as enhancers or SNP-gene interactions. The first native Browser track to use this new format displays data from the GeneHancer database linking human regulatory elements (enhancers and promoters) as arcs to their inferred target genes. Over 1 million regulatory elements obtained from seven genome-wide databases by GeneHancer are visualizable on the human hg19 and hg38 assemblies as color-coded curves ending on their respective target (<https://genome.ucsc.edu/s/PublicSessions/GeneHancer>). Recent new enhancements to the new track type include an option to invert the view from hills to valleys, an improved pack and squish mode and a cluster mode, which combines directional interactions with the same named target or source to display as a single bundled element.

## PgmNr 1670: Using genome graphs for native alignment and variant calling of non-reference sequences in African Americans in the Million Veteran Program.

### Authors:

S. Pyarajan<sup>1,2</sup>; X. Gong<sup>1</sup>; Y. Jeon<sup>3</sup>; S.-G. Ji<sup>1,3</sup>; Y. Shi<sup>1</sup>; M. Xu<sup>1,4</sup>; B. Gorman<sup>1,4</sup>; S. Tetikol<sup>3</sup>; C. Pan<sup>5,6</sup>; P. Tsao<sup>5,6</sup>; on behalf of Million Veteran Program

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) VABHS, Boston, MA.; 2) Harvard Medical School, Boston, MA; 3) Seven Bridges Inc., Charlestown, MA; 4) Booz Allen Hamilton Inc., McLean, VA; 5) VA Palo Alto Health Care System, Palo Alto, CA; 6) Department of Medicine, Stanford University School of Medicine, Stanford, CA

Despite the current usefulness of the human genome reference, its limitation on representing human genetic diversity has been consistently described (Yang *et al*, *Genome Biol.* 2019). Whole genome sequencing (WGS) studies of multiple human populations have reported novel non-reference contigs that are not present in hg38. A graph genome pipeline can efficiently represent these non-reference sequences as alternate paths through the graph and can be used to directly align reads and identify variants on the graph. The overall performance of this graph applied genome-wide has recently been published (Rakocevic *et al*, *Nat Genet*, 2019), and has subsequently been applied to non-repetitive, non-reference (NRNR) sequences from Kehr *et al*, *Nat Genet*, 2017. Applying the NRNR graph on 49 EUR individuals from the 1000 Genomes Project showed that the frequency of the NRNRs obtained from the graph pipeline had an  $r^2$  of 0.96 with the frequency reported by the Icelandic study. Furthermore, the genotypes of the NRNRs correlated well with the genotypes of SNPs in linkage-disequilibrium with the NRNRs.

Here, we applied this graph-based pipeline to 20 African American and 20 European American individuals of the Million Veteran Program (MVP). We compared alignment and variant calls from the BWA-GATK best practices pipeline against a global graph pipeline that incorporates all alternate contigs of hg38, as well as a curated set of high confidence variants. Although the graph pipeline had on average 40% lower compute cost, the overall results of the two pipelines were highly consistent. We observed 40% fewer reads with mates that map to different chromosomes and identified 4% more indels when the graph pipeline was used. To further improve performance, we developed an MVP graph by incorporating additional contigs recently identified from 3 trios of the 1000 Genomes Project (Chaisson *et al*, *Nat Commun* 2019) and the African pan-genome study (Sherman *et al*, *Nat Genet* 2019). We then used the gnomAD-SV statistics to filter for SVs that were present with at least 5% frequency in relevant populations. This led to 2,789 unique contigs, which amounted to 3.6 Mbp of human sequences not present in hg38. Finally, we independently reassembled the contigs using the Simons Genome Diversity panel and re-mapped them as alternate paths for optimal aligner performance. We demonstrate that the custom MVP graph has better alignments and variant call metrics than the global graph.

## PgmNr 1671: Secondary analysis of 1000 genomes in a week using 4 servers.

**Authors:**

M. Samadi; A. Sethia; S. Mahlke; T. Harkins

[View Session](#) [Add to Schedule](#)

**Affiliation:** Parabricks, Ann Arbor, Michigan.

---

As more and more organizations start using Whole Genome Sequencing (WGS) data for deeper analysis and insights, the demand for WGS analysis is rising rapidly. It is common to process hundreds of genomes in a week and there are predictions that analysis of 1000 genomes per week will become standard. But, there are large computational, infrastructural, and cost-related challenges in analyzing such a large number of genomes while maintaining the accuracy and fidelity of the analyses. To enable researchers to efficiently analyze trends in genomic data from entire populations, hospitals/clinics must be able to process large numbers of genomes every week at a reasonable time and cost, these challenges have to be overcome.

On the computing front, traditional general-purpose microprocessors (CPUs) no longer scale performance according to Moore's Law, thus they will not be able to meet the computing demands of WGS. Many compute intensive domains have adopted compute accelerators in the form of commodity Graphics Processing Units (GPUs) to scale performance at reasonable costs. GPUs are well suited for the parallel computation in genomic analysis and are already gaining a foothold in genomic analysis. We have developed a full suite of WGS secondary analysis software (GATK best practices, deepvariant, cnvkit, etc.) that is optimized for GPUs. Results demonstrate 30 to 50 times faster processing with GPUs and users can process a 30x whole genome in under an hour using 8 GPUs. Another major advantage of GPUs is that they are general purpose so they can be used for other computing tasks and readily available in the majority of the computing centers.

Using our software suite, 1000 Genomes can be analyzed every week by a center with commodity hardware using 4 GPU servers. Over a period of 5 years, this represents an analysis cost of one dollar per genome. This represents a marked departure from traditional clusters where thousands of nodes and tens of thousands of cores must be deployed to accomplish the same analysis. This large reduction in infrastructure, cost, and management overheads can provide significant advantages to organizations looking to scale their analysis of WGS data.

## **PgmNr 1672: Here, there, and everywhere: NCBI resources for genomic analysis and visualization in local, web, and cloud environments.**

### **Authors:**

V.A. Schneider; A. Kuznetsov; S. Rangwala; N. O'Leary; T. Madden

[View Session](#) [Add to Schedule](#)

**Affiliation:** IEB, NIH/NCBI, Bethesda, Maryland.

---

Genome-scale analyses play a central role in basic and clinical research. Decisions surrounding tool selection for these types of studies must consider many factors, including data size and available compute power, privacy concerns and the bioinformatics skill level of the analyst. Specifically, these parameters impact the user interface and the environment in which a tool must run. NCBI's collection of genome analysis and visualization tools supports a diverse range of user needs. Here, we highlight several of these, focusing on recent updates and new features. Genome Workbench is a desktop sequence analysis package that integrates analysis tools, such as BLAST, with graphical displays and whose local processing ensures data privacy. The functionality of this powerful tool has now been extended to support preparation of genome submissions to GenBank. Additionally, NCBI offers a suite of command-line alignment tools, such as BLAST+, IgBLAST and Magic-BLAST that can be run locally. Recent updates to BLAST+ will be discussed, including enhanced support for search limits and corresponding changes in BLAST databases. For users who prefer graphical interfaces, NCBI offers a diverse suite of web-based alignment tools, and we will present features of the redesigned BLAST results page. Genome Data Viewer (GDV) is NCBI's web-based genome browser. We will show recent performance enhancements and new features, such as track overlays, improved alignment displays, and expansions to our support for track hubs. Users looking for analysis inputs at the NCBI website, such as genome assemblies and annotation data, can now take advantage of an improved search experience that supports plain language queries, and presents results in a simple interface that supports data retrieval. We will demonstrate the features of this search and results display. NCBI also offers BLAST in the cloud. We will present recent developments in this area, including a BLAST+ Docker image that can be used to run analyses in the Google Cloud Platform and ongoing work to offer a web-based interface for BLAST searches that run remotely in the cloud. This collection of local, web, and cloud-based resources provides users with tools for genomic analyses that best fit their computational needs. This work was carried out by staff of the National Library of Medicine (NLM), National Institutes of Health, with support from NLM.



## PgmNr 1673: Accelerating research with the NCBI Sequence Read Archive on the commercial cloud.

### Authors:

S. Sherry; A. Boshkin; J. Brister; R. Connor; D. Deriy; A. Dillman; K. Durbrow; A. Efremov; L. Fleischmann; G. Godynskiy; D. Ignatovich; A. Iskhakov; A. Johnson; M. Kimelman; A. Klymenko; A. Kochergin; C. Leubsdorf; K. McDaniel; A. Mnev; C. O'Sullivan; S. Ponomarov; W. Raetz; K. Rodarmer, Jr.; K. Rodarmer, Sr.; O. Shutov; D. Soren; A. Stine; M. Vartanian; E. Yaschenko; V. Zalunin

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland.

---

The NCBI Sequence Read Archive (SRA) is NCBI's comprehensive collection of next generation sequence data from human and non-human organisms. At over 10 petabytes, SRA is NCBI's largest research archive and has exceeded the capability of most sites to replicate locally. Recognizing this, NCBI and NIH have recently replicated SRA content to both Amazon and Google commercial cloud platforms as part of the NIH STRIDES program. NCBI is providing community education, and training opportunities to build relationships with SRA users through webinars, online tutorials and a series of NCBI-hosted hackathons that bring together users to work with topical and problem-relevant subsets of SRA data.

Taken as a broad collection of sequences sampled across the tree of life, SRA data are mined for new discoveries about genomic sequence, natural variation, antimicrobial resistant genes, gene expression, methylation states, and previously undescribed genes and species, strains, or viral isolates. By size, SRA is equal parts public and controlled access data — public data includes non-human sequences and human RNAseq and genomic data where individual consent has been provided for the open and unrestricted use of their data. Controlled access data is sequence information from human research study participants supported by NIH and access is restricted and controlled through dbGaP approval protocols.

Moving SRA data to cloud platforms will benefit users by enabling meaningful and timely access to the exponentially growing corpus of SRA data. This migration of data is accompanied by improvements to the SRA Run Selector, a new data model that supports access to both submitted and normalized data formats, improvements to the SRA toolkit, and conformance to NIH's new standards for Identity and Authorization Management. Researchers are invited to join the NCBI cloud user community, participate in Codeathon events, and explore our developing knowledge base of self-directed education resources to design powerful and affordable cloud-based analysis workflows.

This presentation will describe new SRA resource content, procedures for data access and use, and new cloud-related educational resources.



## PgmNr 1674: Human Cell Atlas data distribution: Curation of harmonized open access single cell datasets.

**Authors:**

J. Zamanian; J.A. Hilton; J.S. Strattan; J.M. Cherry; Human Cell Atlas Data Coordination Platform Team

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetics, Stanford University, Stanford, CA, USA

---

The Human Cell Atlas (HCA) is an international scientific collaboration with the aim of generating a comprehensive map of human cell types. The completed atlas will connect molecular profiling with cell location and morphology information across healthy human tissues. This open access resource will enable researchers to accelerate scientific discovery and support the development of new diagnostic tools and medical treatments across diverse disciplines and diseases.

Data are collected and processed for the HCA through the HCA Data Coordination Platform (DCP) being developed as a collaboration between the Broad Institute, Chan Zuckerberg Initiative, the European Bioinformatics Institute, University of California Santa Cruz, and Stanford University. The DCP enables search and delivery of single-cell genomics data in raw and standardized formats with associated metadata files containing experimental details. All standardized processing pipelines used by the DCP are approved by the HCA community and are available to the public. An additional matrix service delivers gene by cell matrices for custom subsets of primary data.

In this poster we discuss the curation and analysis of harmonized single-cell transcriptomic data for open access by the scientific community. Data suitable for integrative analysis and biological interpretation are released in collections for consumption by researchers. We welcome scientific community engagement in developing primary data and analysis outputs useful for supporting research goals. Come visit the HCA website at <https://data.humancellatlas.org/> to learn more and to express an interest in contributing data.

## PgmNr 1675: An approach to develop multi-ethnic polygenic scores using GWAS meta-analysis and mega training cohort.

**Authors:**

J. Zhan; S.B. Laskey; K. Lin; J. Shi; P. Fontanillas; Y. Jiang; B.L. Koelsch; 23andMe Research Team

[View Session](#) [Add to Schedule](#)

**Affiliation:** 23andMe, Mountain View, California.

---

Polygenic scores (PGSs) have been used extensively to predict individuals' genetic liability for a range of phenotypes, from morphological traits to disease predispositions. PGSs are most often developed in European cohorts because individuals of European descent comprise the overwhelming majority of genetics research participants. However, Europeans represent a minority of global genetic diversity, and these PGSs typically perform worse — often dramatically so — in non-European cohorts. While efforts are being made to engage more non-European individuals in genetics research, it is also possible to leverage existing, abundant European genetic data to draw more generalizable scientific conclusions.

Adopting this approach, here we present a method to build PGSs that reduces performance disparities between European and non-European cohorts. This method consists of two steps: 1. Feature selection from a meta-analysis of GWASs integrating multiple populations, Europeans included, to identify tagging variants that are shared across ethnicities; 2. Model training on individual-level phenotype and genotype data from a mega-cohort of European and non-European individuals that leverages the power of the larger European cohort to boost predictive performance in non-European cohorts.

We tested this method on a broad set of phenotypes from the 23andMe database, using data from consented research participants of European, Hispanic/Latino, African/African American, East Asian, and South Asian descent. Comparing the performance of PGSs generated by this method with the European-only PGSs, we observed that our proposed method yields improved performance across response phenotypes and target populations. Specifically, using area under the receiver operating characteristic curve (AUROC) as a standard metric of model performance, we found that PGSs built with this method have significantly improved predictive accuracy for several non-European populations, with some models performing equivalently well in non-European and European cohorts. We believe that this method holds great potential for the development and implementation of PGSs across domains, with implications from direct-to-consumer genetics to clinical practice.

## PgmNr 1676: Increasing the precision of structural variant calls using machine learning.

**Authors:**

A.C. English; S.N. Shekar; A. Mangubat

[View Session](#) [Add to Schedule](#)

**Affiliation:** Spiral Genetics, Seattle, WA.

---

Structural Variants (SVs), genetic changes of at least 50bp, have been associated with an increased risk of neurological and cardiological conditions as well as in rare disorders. For SV calling to be widely adopted in clinical practice, it may be important to ensure that there are few false discoveries that may increase the risk of incorrect clinical diagnoses and unwarranted treatment.

However, in calling structural variants using short read Illumina data, the events are often larger than the unit of observation, a read, making it difficult to disambiguate between events in genomic repeats. Further, structural variant callers may be susceptible to false positives from calling structural variants using alignment information of reads where only part of the read maps to the reference or when using the expected mate pair distance, resulting in reference bias (<https://doi.org/10.1038/s41588-018-0145-5>).

We address this by calling variants by assembling whole reads across the genome, keeping track of paths that diverge and return to the reference. Applying the caller on the read data of HG002 sequenced at 40x using 150bp reads on an Illumina HiSeq, we recover 50.1% of the structural variants in the Genome in a Bottle Consortium Tier1 SV call set with a precision of 58.5%. Further investigation indicates that the false positives appear to represent sequence variation against the reference but with a location that does not match a call in the Tier1 set.

We take the VCF file of calls from HG002 that includes for each SV, the read coverage and minimum nucleotide overlap across the assembled reads. Taking 75% of all calls made at random, we apply the random forest machine learning method. We take the derived model and then apply it back to the data, we observe a recall of 46.7% of true positive SVs from the GIAB Tier1 set and a precision of 83.9%, a 25.4% increase.

It is further possible that dividing the variants into groups based on where they are located in the genome before running a random forest model may further improve precision without a reduction in sensitivity.

Overall, the application of a machine learning derived filter may increase precision and allow for the more regular use of structural variant calls in the bioinformatics analysis of rare disorders in clinical practice.

## **PgmNr 1677: AnnoSV for annotation and prioritization of complex coding and noncoding structural variants in human diseases.**

### **Authors:**

L. Fang; Y.Y. Ding; K. Wang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Children's Hospital of Philadelphia, PHILADELPHIA, Pennsylvania.

---

Recently, long-read sequencing technologies, such as linked-read sequencing, optical mapping, Oxford Nanopore sequencing and PacBio SMRT sequencing, are becoming increasingly used in human genetic studies, offering an unprecedented resolution to detect structural variants (SVs) that are missed by microarrays or short-read sequencing. Generally, over 20,000 SVs can be detected from an individual genome, but it remains challenging to pinpoint a small subset of SVs that are related to human diseases or contribute to phenotypic variations. Existing variant annotation tools offer limited functionality to understand the functional consequences of SVs, especially small non-coding SVs, complex translocations, nested SVs and those SVs located in highly repetitive genomic regions. In this study, we present AnnoSV, a software tool to provide comprehensive functional annotation for SVs including insertions, deletions, inversions, duplications, translocations, and complex events. AnnoSV can utilize gene annotations from GENCODE or NCBI RefSeq database, and annotate the affected genes/regions of an SV. In addition, AnnoSV searches and annotates the repeat sequences involved within or surrounding SV events, facilitating the understanding of the mechanism of SV generation. We also built a region importance score (RIS) to prioritize functionally important SVs, including non-coding SVs. RIS was built from multiple resources including population-based CNV/SV call sets, known disease-associated SVs, gene-disease association resources (such as predictions from Phenolyzer), functional annotations of genes (such as Gene Ontology), tissue-specific annotation of functional elements (such as those from Epigenome Roadmap Project, ENCODE and PsychENCODE), tissue-specific inter-chromosomal interactions. To test the performance of AnnoSV, we used AnnoSV to annotate the SV calls from a whole-genome long-read sequencing data set on the JIH-5 cell line. Among all SV calls, AnnoSV correctly identified two intra-chromosomal translocations as prioritized variants, including EP300-ZNF384 and EP300-CHD4 gene fusions, with the former being a recurrent alteration in B cell acute lymphoblastic leukemia. AnnoSV can take input files in VCF format and is compatible with output files of major short-read and long-reads SV callers such as Delly, and Sniffles. In summary, AnnoSV will facilitate annotation and clinical interpretation of SVs, especially non-coding SVs and complex SVs.

## PgmNr 1678: Improving the prioritization and characterization of structural variants detected by long read sequencing.

### Authors:

J. Reese <sup>1</sup>; D. Danis <sup>2</sup>; J. Jaconbsen <sup>3</sup>; M. Haimel <sup>4</sup>; C. Mungall <sup>1</sup>; P. Robinson <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Lawrence Berkeley National Laboratory, Berkeley, CA, USA; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.; 3) Genomics England, Queen Mary University of London, UK.; 4) The CeMM Research Center for Molecular Medicine, Austria

---

We describe a method for improved prioritization and characterization of structural variants using long read sequencing. A prioritization score for each SV is derived using information about overlapping neutral and disease-associated SVs, effect of SVs on gene elements (deletion, interruption, rearrangement, or duplications of exons), haploinsufficiency, and triplosensitivity. The effect of SVs on gene regulation is assessed based on information about the location and tissue specificity of enhancers and other regulatory sequences. We have previously shown that SVs that disrupt Topologically Associated Domain (TAD) boundaries are not uncommon and that they can be computationally prioritized by relating the phenotypes and tissue specificity of enhancers and genes on either side of a TAD.

Our Exomiser tool couples the calculation of phenotype similarity between human diseases and genetically modified mouse models with the evaluation of variants according to allele frequency, pathogenicity, and mode of inheritance. Exomiser compares HPO-encoded phenotype profiles with human diseases and mouse models; in cases where genes have no reference phenotypes, uses a guilt-by-association protein-interaction network approach. Here, we present software that integrates the assessment of SVs with the Exomiser in a way that allows phenotype-driven prioritization of SVs in long-read diagnostic genome sequencing. We present results of our algorithm on a suite of case reports encoded as Phenopackets (<https://github.com/phenopackets/phenopacket-schema>) and demonstrate how to use the software.

## PgmNr 1679: A graph genome of the Arab population with base pair accurate structural variation for population level variant genotyping.

### Authors:

S.N. Shekar <sup>1</sup>; A.C. English <sup>1</sup>; L. Herta <sup>1</sup>; A. Mangubat <sup>1</sup>; J. Bruestle <sup>1</sup>; E. Aliyev <sup>2</sup>; Y. Mokrab <sup>2</sup>; K.A. Fakhro <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Spiral Genetics, Seattle, Washington.; 2) Sidra Medical Research Center, Qatar; 3) Genetic Medicine, Weill Cornell Medicine-Qatar,

Our recent exploration of ancestral populations' genomes has shown that the widely used Human reference genome (GRCh37/GRCh38) lacks the diversity needed to represent populations of specific ethnicities. By incorporating Single Nucleotide Variants (SNVs) and indels found by alignment-based variant calling into GRCh37, a population-representative reference can be created which increases the proportion of alignable reads from an individual of the same population. Incorporating structural variants (SVs) would further improve alignment.

To more thoroughly capture structural variation in a sequenced individual would in principle require aligning their reads to a continually updated reference graph of all previously called variation. To avoid introducing erroneous variation, such a graph must be made up of accurate breakpoints. Currently, the detection of SVs from short-read data using inference methods can lead to inaccurate breakpoints, unresolved insertion sequences and high false discovery rates. Using a rapidly assemblable read graph-genome representation, BioGraph Assembly recalls on average 48.9% true positive SVs in the GIAB v0.6 Tier1 set across 5 replicates of HG002, higher than any other caller using Illumina whole genome reads. Calling to VCF completes in an hour on a 32 core AWS instance with 32GB of RAM.

Further, by aligning paired reads to a graph reference of SVs called in any of 5 WGS replicates of HG002, the number of true positive variants recalled across ALL 5 individuals increased to 57.5%. Similarly, by aligning reads to a graph genome of the variants discovered using BioGraph Assembly across the parents, the recall of true positive GIAB SVs in HG002 increases by 11.3%. By aligning reads to an Arab Graph Genome of variation from 106 WGS Qatari individuals, we show that we are sensitive to a greater number of SVs in an N+1 genome than from discovery alone in that genome.

Within the 106 individuals, there were 48, 17 and 19 individuals of Bedouin, Sub-Saharan African and Persian ancestry respectively. There were variants unique to each ancestry including >1000bp regions of the reference that had an alternative nucleotide sequence that reads could not map to (N locations =68 Persian, =259 Bedouin, =143 Sub-Saharan African) and sequence resolved insertions >1000bp (N = 5 Persian, =57 Bedouin, =6 Sub-Saharan African).

This suggests that aligning to an Arab Graph Genome of nucleotide resolved calls could improve SV recall over calling against GRCh38.

## PgmNr 1680: ClipSV: Structural variation detection by read extension, spliced alignment, and local *de novo* assembly.

### Authors:

P. Xu; Z. Chong

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of Alabama at Birmingham, Birmingham, Alabama.

---

Structural variations (SVs) are important sources of genomic mutations and are potential causes of various human diseases. Due to their relatively large mutation size (>50 bp), SVs are difficult to accurately detect using short reads from the next-generation sequencing (NGS) platforms. In previous studies, several methods have been developed to detect SV events, such as split reads, discordant read-pairs, (local) *de novo* assembly, and/or a combination of these methods. However, these methods are usually limited by low sensitivity, and the detected SV events are usually biased towards deletions.

In this study, we introduce ClipSV, which primarily relies on the spliced alignment of long sequences reconstructed from clipped reads and local *de novo* assembly to comprehensively characterize SVs. ClipSV searches the signals of SVs by collecting short reads that are clipped during genome alignment. To alleviate the limitation of read length, ClipSV generates longer sequences by iteratively overlapping of the clipped reads. After extending the read length, ClipSV performs spliced alignment to detect deletions and small insertion events (50~200 bp). For large insertions (>200 bp), ClipSV employs both read-pair alignments and local assembly approaches to increasing the overall sensitivity.

To evaluate the performance of ClipSV, we constructed a nonrandom simulation dataset by spiking in 12,745 known SV events of HG002 from the Genome in a Bottle project. ClipSV was able to discover 75.1% of the total true SVs at 30X coverage with 250bp paired-end reads, which is higher than Manta (59.5%), Lumpy (30.6%) and svABA (10.0%) at a comparable precision. Then, we evaluated the performance of ClipSV on the 30X 2x250bp real dataset of HG002, which has a high-confidence callset from multiple platforms. ClipSV was able to detect 8,423 (66.1%) true SV events, while Manta, Lumpy and svABA can only detect 6,002 (47.1%), 4,124 (32.4%), and 1,367 (10.7%) events, respectively. The better sensitivity of ClipSV was mainly due to its higher power in detecting insertions. Therefore, ClipSV will serve as an important tool for SV detection based on the NGS platforms which are still the mainstream sequencing capacity for routine genomic applications.

The source code is freely available at <https://github.com/penguab/ClipSV>.

## PgmNr 1681: Curation and quality assessment of 16S rRNA databases through natural language processing.

### Authors:

S.J. Kern <sup>1</sup>; J. Portman <sup>1</sup>; C. Webster <sup>1</sup>; I. Cernyte <sup>1</sup>; Y. Kiridooshi <sup>2</sup>; W. Suda <sup>3</sup>; C. Mueller <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Enthought Inc., Austin, TX; 2) JSR-Keio University Medical and Chemical Innovation Center, Tokyo, Japan; 3) Laboratory for Microbiome Sciences, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

---

A fundamental step in genomic and metagenomic data analyses is taxonomic annotation, which maps sequences onto known taxonomies, using tools such as Blast, HMMER and others. The underlying computational algorithm has an effect on the taxonomic assignment, but precision is also greatly influenced by the quality of the reference database. In the context of 16S rRNA sequencing there is a wide collection of curated and uncurated databases, including RDP, GRD, Silva, RefSeq and Core but it isn't always clear how accurate the information contained in these is, nor whether the information is consistent across sources. In fact some authors report that roughly one in five of these predictions might be wrong and others report significant differences in the accuracy of taxonomic ranks depending on the database used.

We have developed tools to automate the creation of a reference database combining inputs from any number of separate sources while at the same time assessing the quality of the reference data. To build the database we use natural language processing techniques that parse the information available in each one and map the input sequence names to NCBI taxonomies. The resulting matches are scored and low quality matches are flagged either as rejected or needing additional reviewing. We are able to map entries to NCBI taxonomies with a high degree of certainty for over 99% of the input sequences from Core, RDP, GRD and RefSeq combined.

In order to quantify the quality of each entry in the newly created database we use two criteria. First we look for sequences that are similar but have been assigned to a different taxonomy at a given rank. By clustering the sequences, we assign a score to each entry based on similarity to other taxonomies in the same cluster. The second criteria analyzes entries with the same taxonomy at a given rank but significantly different underlying sequences. We group sequences by taxonomy and use multiple alignment to assign each entry a score, representing how much difference there is between sequences sharing the same taxonomy. Combining the results from these two criteria, we are able to quantify the reliability of taxonomic assignments made in our 16S rRNA data analysis pipeline and track the provenance of the assigned sequence to the database of origin.

Though we have mostly applied our tools to 16S databases, they are general enough to be applicable to a wide range of sequence databases.



## PgmNr 1682: Evaluating the general utility of common phenotype algorithm refinement rules that define many GWAS cohorts.

### Authors:

F.D. Mentch<sup>1</sup>; H. Qiu<sup>1</sup>; M.E. March<sup>1</sup>; H.S. Hain<sup>1</sup>; A.I. Nesbitt<sup>1</sup>; J.J. Connolly<sup>1</sup>; P.M.A. Sleiman<sup>1,2</sup>; H. Harkonarson<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Childrens' Hospital of Philadelphia, Philadelphia, Pennsylvania.; 2) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Published phenotype algorithms are proliferating and may serve as guidelines for construction of generalized computable methods for automated cohort selection. One frequent component of these algorithms is an initial screening rule applied to structured phenotype data available in Electronic Health Records (EHR). The purpose of these rules is to reduce false positives in the cohort. These filters take the form of requiring a qualifying patient's EHR record to contain a set minimum number of unique-day encounters associated with disease or condition relevant ICD diagnosis codes and can be combined with similar requirements for medications or lab test results. We use a simple version of this rule on approximately 1600 EHR-derived phenotypes using the ICD diagnosis code groupings based on PheWAS codes (Denny et al. 2013) and apply it to a biobank of over 65,000 patients with more than 1.5 million encounter-related ICD diagnosis codes. The rule applied here requires presence in an individual's health record of two or more unique-day in-person encounters associated with one or more of the ICD codes in a PheWAS grouping. We define an in-person encounter as a face-to-face encounter with a clinical practitioner. For example, office visits, inpatient stays and implemented medical procedures qualify, while phone calls, canceled or no show visits and letters do not qualify. Over 450 PheWAS phenotypes had 500 or more candidate samples from our biobank and these were used for further analysis. We found the percentages of candidates filtered by the visit type and visit count restrictions to be similar within ICD-CM Chapters. For these 450 phenotypes the visit type restriction to in-person visits had a minimal effect with 314 phenotypes having 2 percent or less of samples filtered out, and 434 phenotypes having 5 percent or less filtered out. However, the removal of candidates based on having only one relevant ICD code had a very significant impact and is a severe restriction. 14 of the 17 ICD Chapters represented by these 450 phenotypes had on average more candidates with only one relevant in-person ICD code encounter than candidates with more than one in-person encounter. Other computable rules for identifying likely false positives in a phenotype cohort are also considered.

## PgmNr 1683: Evaluating the informativeness of variant-level deep learning annotations for human disease.

### Authors:

K.K. Dey <sup>1</sup>; B. van de Geijn <sup>1</sup>; S.S. Kim <sup>1</sup>; F. Hormozdiari <sup>1</sup>; D.R. Kelley <sup>2</sup>; A.L. Price <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, Harvard School of Public Health, BOSTON, MA. 677 Huntington Ave, Boston, MA, 02115; 2) Calico Labs, San Francisco, 1170 Veterans Blvd, South San Francisco, CA 94080

---

Deep learning models have shown great promise in predicting genome-wide regulatory effects from DNA sequence (Eraslan et al. 2019 Nat Rev Genet), but their informativeness for genetic associations to disease is currently unclear. We evaluated variant-level annotations that predict function from DNA sequence (based on the reference allele), trained using three deep learning models: two models previously trained to predict chromatin marks (DeepSEA and Basenji) and a new model that we trained to predict other genomic annotations (BiClassCNN). We applied stratified LD score regression to 41 diseases and complex traits (average  $N=320K$ ) to evaluate each annotation's informativeness for disease heritability, conditional on a broad set of coding, conserved, regulatory and LD-related annotations (including Roadmap and ChromHMM annotations). We meta-analyzed two metrics across traits: enrichment = (% of heritability)/(% of SNPs) and  $\tau^*$  = proportionate change in per-SNP heritability per 1 s.d. change in annotation value, conditional on other annotations.

In an analysis of all 41 traits, DeepSEA and Basenji annotations that were aggregated across tissues produced no conditionally significant signals ( $\tau^*$  all non-significant), despite high accuracy in predicting chromatin marks (AUC up to 0.97); the BiClassCNN annotations for coding and TSS regions were conditionally significant, but these signals could be entirely explained by local GC-content.

We next considered tissue-specific deep learning annotations. In an analysis of 11 blood-related traits, blood-specific DeepSEA and Basenji annotations produced no conditionally significant signals ( $\tau^*$  all non-significant). On the other hand, in an analysis of 10 brain-related traits, two brain-specific deep learning annotations were jointly conditionally significant after correcting for hypotheses tested: DeepSEA-H3K4me3 (enrichment=3.6,  $P=5e-8$ ;  $\tau^*=0.35$ ,  $P=1e-4$ ) and Basenji-H3K27ac (enrichment=2.3,  $P=1e-08$ ;  $\tau^*=0.27$ ,  $P=5e-6$ ); a sequence motif analysis of the brain-specific DeepSEA-H3K4me3 annotation identified enrichments of proximal ( $\pm 1kb$ ) CpG-rich motifs, potentially implicating nucleosome occupancy.

We conclude that deep learning models can provide powerful tissue-specific signals for diseases. However, high accuracy in predicting chromatin marks does not imply conditional informativeness for disease, underscoring the importance of appropriate evaluation procedures that incorporate disease data.

## **PgmNr 1684: Detecting methylation quantitative trait loci using a generalized genetic random field method.**

### **Authors:**

C. Lyu <sup>1</sup>; M.Y. Huang <sup>1</sup>; N.J. Liu <sup>1</sup>; C.A. Hobbs <sup>2</sup>; M. Li <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) School of Public Health, Indiana University Bloomington, Bloomington, Indiana.; 2) Rady Children's Institute for Genomic Medicine, San Diego, CA

---

The patterns of DNA methylation may be regulated by the genetic variants within a genomic region, referred to as a methylation quantitative trait locus (mQTL). The changes of methylation levels can further lead to stable alteration of gene expression, and influence the risk of various complex human diseases. Detecting those mQTLs may provide insights into the underlying mechanism of how genotypic variations and gene-by-environment interactions may influence the disease risk. In this article, we propose a generalized genetic random field (GGRF) method to detect mQTLs by testing the association between the methylation level of a CpG site and a set of genetic variants within a genomic region. The proposed GGRF has two major advantages. First, it considers multiple genetic variants, both common and rare, within a genomic region for genetic-epigenetic association. Second, it uses a beta distribution to model the methylation levels at a CpG site. Through simulations, we demonstrate that the GGRF may have improved power over other existing methods, especially when the sample size is small, which is a common scenario for methylation studies with specific tissue types. We further apply our method to a study of congenital heart defects with 84 cardiac tissue samples.

## PgmNr 1685: Colocalization of GWAS and regulatory QTL with incomplete summary statistics.

**Authors:**

F. Dunbar; E.A. King; J.W. Davis; J.F. Degner

[View Session](#) [Add to Schedule](#)

**Affiliation:** AbbVie Inc., 1 North Waukegan Rd., North Chicago, IL 60064

---

GWAS have uncovered hundreds of thousands of loci associated with complex traits and for many of these traits, it has become clear that the underlying genetic architecture is much more complex than many imagined. From one perspective, the results of GWAS have less direct utility for risk prediction or medical decision making due to each effect being relatively small. From another perspective, if GWAS associations point us toward causal mechanisms of disease, they may still have great utility for guiding drug discovery and development. Large efforts to understand the effects of genetic variation on low level cellular phenotypes such as the Gene Tissue Expression (GTEx) Project have greatly aided in generating hypothesis that explain the mechanisms of GWAS associations. However, there are sufficiently large numbers of trait associated and expression associated loci that significance of association of a single variant with both is not sufficient to assume a shared causal mechanism. Colocalization is used to determine whether the same variant is causal for multiple phenotypes and is stronger evidence for understanding mechanism than shared significance. Current colocalization methods require full summary statistics for both traits, limiting their use with the majority of reported GWAS associations (e.g. GWAS Catalog). We propose a new approximation to the popular coloc method (Giambartolomei et al. 2014, PLOS Genetics) that can be applied when summary statistics for one trait are only available for the top associated SNP. Our method imputes missing summary statistics using LD structure, and performs colocalization between the imputed statistics and full summary statistics for a second trait. As test of whether we are able to approximate the posterior probability of colocalization, we apply our method to colocalization of UK Biobank phenotypes and GTEx eQTL. We show good correlation between posterior probabilities of colocalization computed from imputed and observed UK Biobank summary statistics ( $R^2 = 0.92$  compared to using full summary statistics). Finally, we apply this method to estimate colocalization of GWAS Catalog traits and eQTL and use the results to prioritize drug targets. We find that colocalized trait-gene pairs are enriched in tissues relevant to those complex traits. Further, we find colocated gene-trait pairs are more likely to predict Mendelian disease genes for the same trait.

## PgmNr 1686: Quantifying the regulatory effect size for conditional eQTL data.

### Authors:

N. Ehsan <sup>1</sup>; B. Kotis <sup>1</sup>; F. Aguet <sup>2</sup>; A. Brown <sup>3</sup>; S. Castel <sup>4,5</sup>; T. Lappalainen <sup>4,5</sup>; P. Mohammadi <sup>1,6</sup>; GTEx Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Integrative Structural and Computational Biology, The Scripps Research Institute; 2) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) University of Dundee; 4) Department of Systems Biology, Columbia University, New York, NY; 5) New York Genome Center, New York, NY; 6) Scripps Research Translational Institute, La Jolla, CA, USA

---

Analysis of expression Quantitative Trait Loci (eQTL), genetic variants that affect expression in specific genes, is a widely used approach for functional characterization of the genome and its relation to higher level phenotypes such as disease. In recent years, conditional eQTL analysis has extended this analysis in large cohorts to allow for identification multiple independent loci impacting the expression in each gene. Previously, we introduced allelic Fold Change (aFC), as a unifying notion of effect size for eQTLs and Allele-Specific Expression (ASE) data that is biologically interpretable.

Here, we introduce a new tool to generalize aFC estimation to conditional eQTL data. We apply our method to Genotype Tissue Expression (GTEx) ver. 8 data to derive effect size estimates for 471,726 *cis*-eQTLs spanning over 34,548 genes in 49 human tissues. The data includes up to 16 independent eQTLs per gene in a tissue with ~26% of the genes associated with more than one eQTLs in each tissue. We use ASE data from phased genotypes of 838 individuals with WGS in GTEx data to show that these effect size estimates are consistent with observed allelic imbalance in ASE data at the population-level. We quantify the fraction of genetic regulatory variation at the individual level that is mechanistically described by known eQTLs. Next, we show how our estimates of regulatory effect size for conditional eQTLs can be used in a phasing-aware model of *cis*-regulation to impute gene expression in genotyped individuals for the purpose of estimating local gene expression heritability and transcriptome-wide association study. We compare the performance of our gene expression imputation method to standard regression-based models.

## **PgmNr 1687: TigeRFISH: A computational pipeline to design genome-scale oligonucleotide hybridization probes to visualize novel tandem repeats.**

### **Authors:**

R. Aguilar; Z. Deng; E. Hershberg; B.J. Beliveau

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genome Sciences, University of Washington Seattle, Seattle, Washington.

---

More than half of the human genome is comprised of repetitive sequences such as tandem repeats. These regions have been implicated in 3D genome organization and the maintenance of genome stability. Although these regions have established roles in the regulation of long noncoding genes, studying local repetitive elements has remained a technical challenge. Tandem repeats are not accessible with many conventional assays and these regions are not mappable which poses substantial roadblocks for visualization and mapping applications. FISH applications enables researchers to visualize the distribution of RNA and DNA in fixed samples. However more recent optimizations in synthetic biology has since provided researchers with the opportunity to construct FISH probes from oligonucleotides (oligos). Oligo-based FISH is a powerful single cell technique that allows researchers to control aspects of probe design such as binding energy and specificity. Here, we present TigeRFISH - a computational pipeline that allows for the design of optimal oligo FISH probes for tandem repeats genome wide. We implement existing computational tools including Tandem Repeat Finder coupled with OligoMiner to conduct k-mer based enrichment analysis to assess chromosome specificity followed by thermodynamic simulations of each probe. With TigeRFISH's capabilities to identify target repetitive candidates, we proceed with high-throughput microscopy at the single cell level to gain novel insight of the role these repetitive regions play in genome organization. Furthermore, the application of this technology as a part of the OligoMiner workflow may be used more broadly to facilitate the design of hybridization probes for a variety of applications genome-wide.

## PgmNr 1688: CasCADE: A fully integrated and open source pipeline to design CRISPR mutagenesis experiments.

### Authors:

V. Nelakuditi <sup>1</sup>; H. Elrick <sup>2</sup>; G. Clark <sup>2</sup>; L.M.J. Nutter <sup>2</sup>; A.K. Ramani <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Computational Medicine, the Hospital for Sick Children, Toronto, ON, Canada; 2) The Centre for Phenogenomics, The Hospital for Sick Children, Toronto, ON, Canada

---

CRISPR-mediated genome editing has enabled researchers to perform mutagenesis experiments with relative ease. Though tools exist for guide RNA (gRNA) design and predicting their off-target sites, there exists a need for a comprehensive platform to design, view, evaluate, store, and catalogue gRNAs and their associated primers. CasCADE (Cas Computer-Aided Design) integrates existing open source tools such as JBrowse, Primer3, BLAST, bwa, and Silica to create a complete allele design and quality assurance pipeline. It is available both from a public github repository and as a Docker image, for ease of installation and portability. CasCADE stores data locally, eliminating the security and privacy concerns that make external web-based applications less desirable. For any organism of interest, CasCADE downloads the genome along with annotations for genes, transcripts, and regulatory regions from Ensembl and loads them as tracks into the genome browser, JBrowse. Users can then select specific regions of interest, specify PAM sequences and design gRNAs to target these regions. Importantly, CasCADE provides the user with specificity scores for gRNAs and predicted off-target sites to assist with gRNA choice and then saves the information for user-selected gRNAs along with custom notes to an internal Mongo database. Once the gRNAs have been added to the database, users can trigger Primer3 to design primers for both the wild-type and mutant alleles. Primers are checked for specificity in the genome using BLAST and an *in silico* PCR check is performed with Silica. Selected gRNAs and primers are displayed visually as JBrowse tracks and stored permanently in the database. This allows users to retrieve information related to an experiment and re-analyze the target, if necessary. Additionally, once a design has been finalized, CasCADE generates files containing the guides, primers, and exons for a gene, which can be edited by sequence annotation tools such as SnapGene or ApE (A plasmid Editor). If a design fails, unsuccessful guides can be marked and their scores, off-targets, and genomic context used to make an informed decision about redesign. Currently only deletion alleles are supported, but future versions will allow for the design of point mutations (variants), conditional alleles, and mutations in non-coding regions.

## PgmNr 1689: Aggregated genome quality metrics can predict variant calling reproducibility in individual samples.

### Authors:

C. Colombo <sup>1</sup>; M. Bekritsky <sup>2</sup>; N. Johnson <sup>2</sup>; Z. Kingsbury <sup>1</sup>; E. Tsogka <sup>1</sup>; J. Denecker <sup>1</sup>; N. Walker <sup>1</sup>; M. Vasipalli <sup>1</sup>; T. Newington <sup>1</sup>; S.J. Humphray <sup>1</sup>; D.R. Bentley <sup>1</sup>; M.A. Eberle <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Illumina Cambridge Ltd., 19 Granta Park, Great Abington, Cambridge, CB21 6DF, United Kingdom; 2) Illumina Inc., 5200 Illumina Way, San Diego, CA 92122, USA

The advent of next generation sequencing has enabled a growing number of large-scale whole-genome sequencing projects. A critical aspect of these studies is variant calling accuracy. Multiple steps contribute to the final variant calls and each represents a potential source of variability: library preparation, sequencing platform, instrument calibration, operating conditions and software pipeline. Resources such as Genome in a Bottle or Platinum Genomes provide validation data for assessing variant calling performance but are limited to a subset of “confident regions” in a few samples. Alternatively, reproducibility analysis measures genotype concordance between different sequencing experiments of the same sample and does not rely on high-confidence variant truth sets. It can thus be applied to any sample as a complementary way to assess laboratory data quality.

To demonstrate the value of reproducibility analyses we sequenced the 17 members of the Platinum Genomes pedigree under a variety of conditions and compared SNV and indel calls between experiments. TruSeq DNA PCR-Free bulk libraries of the samples were sequenced on 2 runs on 1 HiSeq X instrument and on 6 runs on 2 NovaSeq instruments, with each run also containing 2 distinct libraries of one of the samples. Combined, each sample was sequenced at least 8 times and we calculated genotype concordance between all experimental combinations.

We found variant concordance ranges from 98.2% - 98.7% for SNVs and 93.1% - 95.4 % for indels, with different experimental conditions having comparable variability. Additionally, we hypothesized that discordant variants are not distributed equally across the genome, but tend to cluster in challenging regions. To test this, we aggregated data from a large cohort of samples and used per-base sequence-quality metrics to detect anomalous regions, predicting that these will correlate with lower variant calling accuracy. Indeed, stratified results show that while variant calls in confident regions are highly consistent (99.5% - 99.8% concordance for SNPs and 97% - 98.8% for indels), anomalous regions are hotspots for variability (80.3% - 87.9% concordance for SNPs and 89.8% - 92.9% for indels). Using aggregated data from multiple samples proves to be a promising approach for predicting variant call reliability in individual samples based only on the underlying genome quality, with potential clinical applications as a measure of confidence for candidate variant calls.



## PgmNr 1690: High-specificity variant filter enables joint analysis of whole genome sequence data from multiple studies and sequencing centers.

### Authors:

S. Gagliano Taliun<sup>1</sup>; Y. Li<sup>1</sup>; D. Ray<sup>2</sup>; P. Yajnik<sup>1</sup>; S. Lee<sup>1</sup>; L.J. Scott<sup>1</sup>; S.A. McCarroll<sup>3</sup>; C.N. Pato<sup>4</sup>; G.R. Abecasis<sup>1</sup>; M. Boehnke<sup>1</sup>; H.M. Kang<sup>1</sup>; NIMH InPSYght Consortium and NHLBI TOPMed Program

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, USA; 2) Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA; 3) Stanley Center for Psychiatric Research and Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 4) State University of New York Downstate Medical Center College of Medicine, Brooklyn, NY, USA

---

Joint analysis of whole genome sequencing (WGS) datasets can boost statistical power to detect disease risk-altering alleles. Here we focus on increasing power of a WGS case-control study by incorporating genetic ancestry-matched external WGS samples as additional controls. A first step to ensure data harmonization is to use a functionally equivalent pipeline to map and process alignments. However, in our experience, false positive associations in the resulting genome-wide comparisons still occur due to between-study differences in sequencing protocols or depth. Our goal is to develop high-specificity variant filtering procedures to eliminate such variants, and enable comparisons of WGS studies processed in a compatible manner.

To illustrate the challenges, we execute a joint analysis of African American WGS samples from the NIMH InPSYght study (N=3K controls, N=5K schizophrenia or bipolar cases, average depth 27x) with control samples of the same ancestry from the NHLBI TOPMed Project (N=15K, average sequencing depth 37x). Samples were sequenced at five centers, but sequence data were jointly processed and genotypes called together. We start with a comparison of InPSYght controls versus TOPMed controls, a scenario where we expect no true positive association signals. We adjust for genetic relatedness, sex and four principal components of ancestry, but observe a modest number of false positive associations: 113 common or low-frequency variants reach genome-wide significance ( $p \leq 5 \times 10^{-8}$ ) and an additional 158 also deviate from the expected null distribution. These false positive variants consistently had lower depth and genotyping quality in carriers, which we hypothesized could be driving the spurious associations.

We evaluate strategies for variant filtering using information such as duplicate concordance, Mendelian inconsistencies, sequencing depth and genotype missingness. We show that a strict set of variant filters that remove ~3% of variants (1,153,945 of 41,732,031 variants with allele count  $\geq 10$ ) enable joint analysis. Specifically, we remove variants with either a  $\geq 0.5\%$  discordance between duplicate samples or where  $\geq 2\%$  of the pairs have a missing genotype. In the variant-filtered control versus control comparison no variants reach genome-wide significance nor deviate from the null. Our duplicate-based variant filtering strategies allow for the addition of external controls in WGS datasets to boost power to detect disease associations.

## PgmNr 1691: Toward a more perfect genome: Methods for evaluating whole genome sequencing libraries.

**Authors:**

H. Kang; V. Onuchic; G. Berry; C. Glidewell-Kenney; N. Baird; N. Kim; W. Eap; K. Magdirila; A. Gross; N. Johnson; M. Bekritsky; K. Kruglyak; A. Crawford

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical Genomics, Illumina, San Diego, CA.

---

The utility of whole genome sequencing (WGS) to characterize germline variation across the genome has been well-documented, particularly with application to rare disease diagnosis. Compared to whole exome sequencing (WES) alone, WGS delivers more uniform coverage of exonic regions; enables the identification of expanded variant classes including copy number variants (CNVs), structural variants (SVs), and repeat expansions—all of which have been associated with clinical phenotypes; and allows for the detection pathogenic variants outside of the targeted exome.

We anticipate widespread adoption of clinical WGS in the near term for germline variant detection; however, given the range of available library preparation kits and analysis pipelines, it is critical to establish robust quality metrics to define and evaluate the clinical genome. Here we demonstrate the relationship between various sequencing-based metrics and performance of variant calling across different variant classes. While some correlations are intuitive and well-established—such as locus depth with small variant calling—others are more nuanced and novel. We summarize our observations across multiple library preparation techniques and analysis pipelines and demonstrate that our metrics are robust for evaluating clinical genomes across a range of conditions. Finally, we present our proposed evaluation strategy and resulting quality metrics. This evaluation framework can be leveraged to evaluate different clinical WGS protocols, or to evaluate the quality of individual sample results during routine clinical processing. Ultimately, the implementation of uniform quality metrics to define the clinical genome will facilitate both broader adoption of WGS as well as downstream data aggregation efforts to support genotype-phenotype studies.

## PgmNr 1692: High-quality human genomes achieved through HiFi sequence data and FALCON-Unzip Assembly.

### Authors:

Z.N. Kronenberg <sup>1</sup>; I. Sovic <sup>1</sup>; C. Dunn <sup>1</sup>; S. Kingan <sup>1</sup>; G.T. Concepcion <sup>1</sup>; J. Drake <sup>1</sup>; J. Korlach <sup>1</sup>; P. Peluso <sup>1</sup>; A.M. Wenger <sup>1</sup>; J. Mountcastle <sup>2</sup>; B. Haase <sup>2</sup>; O. Fedrigo <sup>2</sup>; E.D. Jarvis <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pacific Biosciences, Menlo Park, CA, USA.; 2) Laboratory of Neurogenetics of Language, Box 54, The Rockefeller University, New York, NY 10065, USA, Howard Hughes Medical Institute, Chevy Chase, MD, USA.

---

*De novo* assemblies of human genomes from accurate (85-90%), continuous long reads (CLR) now approach the human reference genome in contiguity, but the assembly base pair accuracy is typically below QV40 (99.99%), an order-of-magnitude lower than the standard for finished references. The base pair errors complicate downstream interpretation, particularly false positive indels that lead to false gene loss through frameshifts. PacBio HiFi sequence data, which are both long (>10 kb) and very accurate (>99.9%) at the individual sequence read level, enable a new paradigm in human genome assembly. Haploid human assemblies using HiFi data achieve similar contiguity to those using CLR data and are highly accurate at the base level<sup>1</sup>. Furthermore, HiFi assemblies resolve more high-identity sequences such as segmental duplications<sup>2</sup>. To enable HiFi assembly in diploid human samples, we have extended the FALCON-Unzip assembler to work directly with HiFi reads. Here we present phased human diploid genome assemblies from HiFi sequencing of HG002, HG005, and the Vertebrate Genome Project (VGP) mHomSap1 trio on the PacBio Sequel II System. The HiFi assemblies all exceed the VGP's quality guidelines, approaching QV50 (99.999%) accuracy. For HG002, 60% of the genome was haplotype-resolved, with phase-block N50 of 143Kbp and phasing accuracy of 99.6%. The overall mean base accuracy of the assembly was QV49.7. In conclusion, HiFi data show great promise towards complete, contiguous, and accurate diploid human assemblies.

# PgmNr 1693: Mappa monstrix: The benchmarking of alignment tools to establish best practices in long-read whole human genomics.

## Authors:

J. LoTempio <sup>1,2</sup>; S. Bhattacharya <sup>1</sup>; H. Barseghyan <sup>1,3</sup>; E. Délot <sup>1,3</sup>; E. Vilain <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Center for Genetic Medicine Research, Children's Research Institute, Children's National Medical Center, Washington DC 20010, USA; 2) Institute for Biomedical Sciences, School of Medicine and Health Sciences, George Washington University, Washington DC 20052, USA; 3) Department of Genomics and Precision Medicine, School of Medicine and Health Sciences, George Washington University, Washington DC 20052, USA

---

Long-read sequencing (LRS) technologies, including nanopore-based and single-molecule real-time sequencing, are capable of sequencing single reads spanning repetitive genomic regions and, potentially, entire chromosomes. This potential for *de novo* human assembly has led to studies vetting basecalling, assembly, and polishing tools for this application. In comparison, use of LRS for alignment to references, which has proven useful for smaller genomes, amplicons, and metagenomics, has been considerably less vetted for high-coverage human whole genome sequences (WGS). The number of tools available for alignment is growing in response to need, but the practical differences between alignment pipelines are not yet known.

Upon inspection of LRS alignment tools for their suitability for human WGS, we found some monsters lurking at the fringes of the genome maps. To slay some, we benchmarked the LRS alignment tools with the datasets generated from a healthy white human female, NA12878, on Oxford Nanopore's MinION-platform by the Whole Genome Sequencing Consortium and on the Pacific Biosciences SMRT-sequence data by the National Institute of Standards and Technology. We compared computational performance (peak memory utilization, CPU time, file size/storage requirements), genome depth (X-coverage), and differing breakpoints of structural variants between maps.

We serially ran tools, including minimap2, NGMLR and GraphMap, on the George Washington University high performance compute cluster, Pegasus, to understand the how consistent alignment files generated from each tool under the same conditions were across runs. We then compared read alignments and leveraged structural variant-calling tools, including Sniffles and NanoSV, to assay the physical difference between the maps generated by each alignment tool. Our investigations established that different alignment tools produced different maps and, subsequently, different structural variant callsets. The differences were largely found from comparisons across alignment tools, but also within tools, across runs. Based on these experiments, we propose best practices for use of these tools in long-read human WGS alignment.

## PgmNr 1694: Exploring genetic variation in 132,345 deep whole genomes with BRAVO.

### Authors:

D. Taliun<sup>1</sup>; P. VandeHaar<sup>1</sup>; X. Zhao<sup>1</sup>; J. Pleiness<sup>1</sup>; J. LeFaive<sup>1</sup>; D.K. Arnett<sup>2</sup>; A. Ashley-Koch<sup>3</sup>; K.C. Barnes<sup>4</sup>; R. Beer<sup>5</sup>; E. Boerwinkle<sup>6</sup>; S. Gabriel<sup>7</sup>; W. Gan<sup>5</sup>; R. Gibbs<sup>8</sup>; C. Laurie<sup>9</sup>; J. Mikulla<sup>5</sup>; D. Nickerson<sup>9</sup>; P. Qasba<sup>5</sup>; K. Rice<sup>9</sup>; S.S. Rich<sup>10</sup>; E. Silverman<sup>11</sup>; P. Srinivas<sup>5</sup>; B. Weir<sup>9</sup>; J. Wilson<sup>12</sup>; T.W. Blackwell<sup>1</sup>; H.M. Kang<sup>1</sup>; M. Boehnke<sup>1</sup>; G. Abecasis<sup>1</sup>; A.V. Smith<sup>1</sup>; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Michigan, Ann Arbor, MI; 2) University of Kentucky, Lexington, KY; 3) Duke University Medical Center, Durham, NC; 4) University of Colorado Anschutz Medical Campus, Aurora, CO; 5) National Heart, Lung, and Blood Institute National Institutes of Health, Bethesda, MD; 6) University of Texas Health at Houston, Houston, TX; 7) Broad Institute, Cambridge, MA; 8) Baylor College of Medicine Human Genome Sequencing Center, Houston, TX; 9) University of Washington, Seattle, WA; 10) University of Virginia, Charlottesville, VA; 11) Brigham & Women's Hospital, Boston, MA; 12) University of Mississippi, Jackson, MS

---

The NHLBI Trans-Omics for Precision Medicine (TOPMed) program aims to improve our understanding of heart, lung, blood and sleep disorders by adding deep whole-genome sequencing data to existing high-priority studies of these conditions. To date, TOPMed sequenced >130,000 samples. Harmonized data for the majority of participants are available through dbGaP. In addition to enabling phenotype-genotype association analyses, single nucleotide and short indel variants identified in the TOPMed data, when coupled with annotation, inform human genetic analyses. The TOPMed online variant browser (BRAVO - BRowse All Variants Online) enables researchers to access, explore, and analyze genome variation and supporting data. With BRAVO, researchers have access to ~705M genetic variants in a subset of 132,345 deeply sequenced (>38x) genomes from the latest TOPMed data set (Freeze 8), of which 215M variants (30%) are novel and 323M variants (46%) are singletons (observed in one heterozygous individual). Among the 8M protein-coding sequence variants, 458,513 are predicted loss-of-function, including stop-gain, splice site, and frameshift variants. Short indels constitute 53M (8%) of all identified variants. For each variant, BRAVO provides summary level information including sequencing depth, quality metrics, allele frequency, numbers of heterozygous and homozygous individuals, allele deleteriousness score (CADD), and putative functional consequences. Supporting reads are shown for each variant from anonymized, randomly selected individuals. BRAVO code is publicly available and enables other studies to establish their own compatible variant browser. Future work will enable browsers to communicate, facilitating comparisons of variant list and sequencing depth and other characteristics across studies. TOPMed data is available in BRAVO at <https://bravo.sph.umich.edu>.

## PgmNr 1695: A robust benchmark for germline structural variant detection.

### Authors:

J. Zook <sup>1</sup>; N. Olson <sup>1</sup>; N. Hansen <sup>3</sup>; F. Sedlazeck <sup>4</sup>; A. Wenger <sup>5</sup>; A. English <sup>6</sup>; C. Xiao <sup>7</sup>; A. Carroll <sup>8</sup>; M. Salit <sup>2</sup>; Genome in a Bottle Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD; 2) Joint Initiative for Metrology in Biology, Palo Alto, CA; 3) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, Rockville, MD; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston TX; 5) Pacific Biosciences, Menlo Park, CA; 6) Spiral Genetics, Seattle, WA; 7) National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD; 8) Google, Mountain View, CA

---

New technologies and analysis methods are enabling genomic structural variants (SVs) to be detected with ever-increasing accuracy, resolution, and comprehensiveness. Translating these methods to routine research and clinical practice requires robust benchmark sets. We developed the first benchmark set for identification of both false negative and false positive germline SVs, which complements recent efforts emphasizing increasingly comprehensive characterization of SVs. To create this benchmark for a broadly consented son in a Personal Genome Project trio with broadly available cells and DNA, the Genome in a Bottle (GIAB) Consortium integrated 19 sequence-resolved variant calling methods, both alignment- and de novo assembly-based, from short-, linked-, and long-read sequencing, as well as optical and electronic mapping. The final benchmark set contains 12745 isolated, sequence-resolved insertion and deletion calls  $\geq 50$  base pairs (bp) discovered by at least 2 technologies or 5 callsets, genotyped as heterozygous or homozygous variants by long reads. The Tier 1 benchmark regions, for which any extra calls are putative false positives, cover 2.66 Gbp and 9641 SVs supported by at least one diploid assembly. Support for SVs was assessed using svviz with short-, linked-, and long-read sequence data. In general, there was strong support from multiple technologies for the benchmark SVs, with 90% of the Tier 1 SVs having support in reads from more than one technology. The Mendelian genotype error rate was 0.3%, and genotype concordance with manual curation was  $>98.7\%$ . We demonstrate the utility of the benchmark set by showing it reliably identifies both false negatives and false positives in high-quality SV callsets from short-, linked-, and long-read sequencing and optical mapping. GIAB is working towards a new version of the benchmark set that will use new technologies and methods such as PacBio Circular Consensus Sequencing and ultralong Oxford Nanopore sequencing to expand to more challenging genome regions and include more challenging SVs such as inversions. We are also developing a robust integration process to make calls on GRCh37 and GRCh38 for all seven GIAB samples.

## PgmNr 1696: Scalable prediction of polygenic risk from summary statistics in multiple traits.

### Authors:

L. Gai <sup>1</sup>; S. Sankararaman <sup>1</sup>; E. Eskin <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computer Science, UCLA, Los Angeles, California.; 2) Human Genetics, UCLA, Los Angeles, California

---

Polygenic risk scores (PRS) aggregate the estimated effects from many variants to predict disease risk or other phenotypes, where the effects are traditionally estimated from a genome-wide association study (GWAS) on the trait of interest. However, one can improve the predictive power of the risk score by incorporating information from multiple related traits. Existing multi-trait approaches require individual level data for related traits, which may be difficult to obtain, or use GWAS summary statistics with strong assumptions about the genetic correlation across traits. Typically, variant effects in each trait are assumed to be drawn from a polygenic model where the correlation across traits is identical across all regions of the genome. This ignores patterns of local genetic correlation and can lead to inaccurate effect size estimates in regions where the global genetic correlation pattern is a poor match for the true pattern of correlation. To address these issues, we extend existing multi-trait approaches by modeling effects as drawn from a mixture of distributions with different correlation patterns across traits. Our method is flexible yet computationally efficient, and only requires summary statistics to obtain effect size estimates, which are then used for improved risk score prediction. We demonstrate the utility of this multi-trait approach in simulated data and the UK Biobank data in a variety of traits.



## PgmNr 1697: A scalable framework for identifying genetic variant set associated with polygenic-traits in UK Biobank.

### Authors:

Y.-C. Hwang <sup>1</sup>; P. Nguyen <sup>2</sup>; B.T. Hannigan <sup>1</sup>; J. Chin <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Science, DNAnexus, Mountain View, CA.; 2) Deep Learning, DNAnexus, Mountain View, CA.

---

Genome-wide association studies (GWAS) have been instrumental for discovering disease and trait-associated genetic variants, typically single-nucleotide polymorphisms (SNPs). While GWAS can identify SNPs that are marginally associated with traits using univariate tests, most traits in humans are polygenic – where a trait is influenced by more than one gene. Recently, lasso (least absolute shrinkage and selection operator) has been used as a multivariate prediction model for selecting a set of relevant SNPs that are useful for predicting a phenotype of interest. This regression method simultaneously performs variable selection and estimation.

UK Biobank, a large prospective population-based cohort study, collects extensive genotypic and phenotypic data of 500,000 individuals, aged 40-69 years, from the UK. The study includes genome-wide genotyping data (805,426 measured variants per individual), and more than 2000 health-related phenotypes. This ultra-high dimension, large-scale cohort makes finding subsets of genes associated with a polygenic trait statistically possible in human populations. However, the data size leads to a computational challenge in fitting the entire cohort into limited storage and memory. Recently, Qian et al. proposed a batch screening iterative lasso (BASIL) algorithm that reduces the problem to a manageable size by implementing lasso in an iterative fashion and parallelizing the screening problem in each iteration. Though it works with subsets of predictors, it does not compensate by providing an approximate solution. They implemented BASIL as a highly optimized R package, `snpNet`, and provided examples of finding SNP sets that can predict two quantitative traits (height and BMI) and two qualitative traits (asthma and high cholesterol). Using BASIL, they were able to calculate results 20% faster than other alternatives, and with better I/O efficiency.

In this study, we packaged `snpNet` into a portable docker image and deployed it into a cloud environment. After reproducing the discovery of SNP subsets associated with the original four phenotypes, we further extended the analysis to another 1000 phenotypes. Leveraging the flexible capacity of cloud computing, we are able to discover all SNP sets associations efficiently without worrying about overloading a local server. The `snpNet` docker image and cloud app enables easy adoption of the tool and leverages the massive cloud resource for analyzing high-dimensional datasets like UK Biobank.



## **PgmNr 1698: Joint modeling of winner's curse and study-specific confounding explains replication variability in genome-wide association studies.**

### **Authors:**

J. Zou <sup>1</sup>; J. Zhou <sup>1</sup>; S. Faller <sup>2</sup>; R. Brown <sup>1</sup>; E. Eskin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) UCLA, Los Angeles, California.; 2) Duke University, Durham, North Carolina

---

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with complex human traits, but only a fraction of variants identified in discovery studies achieve significance in replication studies. The “replication crisis” in GWAS studies has been well-studied in the context of winner’s curse, which is the inflation of effect size estimates in underpowered studies. Multiple methods have been proposed to correct for the effects of winner’s curse. However, winner’s curse is often not sufficient to explain lack of replication. Another reason why studies fail to replicate is that there are fundamental differences between the discovery and replication studies. A confounding factor can create the appearance of a significant finding while actually being an artifact that will not replicate in future studies. We propose a statistical framework that utilizes GWAS replication studies to estimate the effect of confounders in GWAS summary statistic data. We jointly model the effect of winner’s curse and study-specific confounders and correct for these effects. We apply this framework to 100 human GWAS studies and 22 CFW mouse GWAS studies and show that modeling both winner’s curse and study-specific confounding better accounts for replication between studies compared to an approach that only accounts for winner’s curse.

## PgmNr 1699: Matchmaking patterns at the Undiagnosed Diseases Program site of the Undiagnosed Diseases Network.

### Authors:

D. Adams <sup>1,2</sup>; J. Dizon <sup>2</sup>; M. Malicdan <sup>1,2</sup>; D. Mossbrook <sup>2</sup>; C. Lau <sup>2</sup>; C. Toro <sup>2</sup>; C. Tifft <sup>1,2</sup>; W.A. Gahl <sup>1,2</sup>; Undiagnosed Diseases Network, Common Fund, NIH

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NHGRI, NIH, Bethesda, Maryland.; 2) Undiagnosed Diseases Program, Bethesda, Maryland

---

The Undiagnosed Diseases Program (UDP), established in 2008, was created to evaluate and research study participants with severe diseases that remained undiagnosed despite extensive clinical workup. The UDP joined with other medical and research centers across the country to start the Undiagnosed Diseases Network in 2013. Clinical genomics has been an important tool for evaluating undiagnosed cases. Diagnostically unrevealing genome and exome studies are commonly encountered and are a research focus at many UDN sites. When non-diagnostic, genomic data is frequently used to generate research hypotheses around genes not currently associated with human disease. Matchmaking, through the Matchmaker Exchange (MME), PhenomeCentral, GeneMatcher and other mechanisms is used to in an attempt to identify interested researchers and build patient cohorts. In an effort to assess the effectiveness of our approach to matchmaking, we reviewed submitted and successfully matched cases from the UDP. For example, PhenomeCentral (our default MME node) contained 542 matches to 65 unique cases. Of those cases, evidence sufficient for a clinical diagnosis was eventually reached in 69% of cases (45)—although not all directly from matchmaking. Some form of matchmaking contributed to clinically actionable or publication suitable results in 38 instances overall. The mechanisms for successful matchmaking were highly diverse, including direct interactions, use of third-party databases and submission to multiple MME nodes. For example, 15 cases were solved via submission to GeneMatcher. Although use of potentially useful phenotypic matching could be demonstrated in PhenomeCentral, essentially all historically successful matches were accomplished via gene names. This observation does not likely account for person to person communication about cases of interest.

Matchmaking is a useful tool for building cohorts and establishing ample evidence to assert a gene-phenotype relationship. However, best practices have not been established and our own activity encompasses heterogeneous approaches. We present data from our own experience with matchmaking, including updated data, plans for refinement of approaches to case matching, and phenotype-based case matching.

## PgmNr 1700: Go Get Data: Simple, reproducible access to genomics datasets and annotations.

### Authors:

M. Cormier<sup>1,2</sup>; B. Pedersen<sup>1,2</sup>; J. Belyeu<sup>1,2</sup>; J. Köster<sup>3</sup>; A. Quinlan<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Utah, Salt Lake City, UT, USA; 2) USTAR Center for Genetic Discovery, University of Utah, Salt Lake City, UT, USA; 3) Algorithms for reproducible bioinformatics, Institute of Human Genetics, University of Duisburg-Essen, Essen, NRW, Germany; 4) Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA

---

A common and frustrating problem in genomics is managing the vast datasets and annotations germane to one's experiments. Common problems include the many disparate data and annotation repositories, differences in genome builds, inconsistencies with chromosome labeling, and different coordinate system standard among the various genomic data file formats. Researchers consequently waste valuable time identifying, collecting, and processing relevant datasets for integrative analysis with their experiments. Ultimately, this inhibits reproducibility and constrains research creativity.

Similar problems have existed with software access and usability, motivating the development of software package managers. Inspired by these modern software package management systems, we have developed a data management system, Go Get Data (GGD). GGD utilizes the Conda package management system, along with the infrastructure of Bioconda, to mitigate the many complexities with procuring and curating genomic data.

GGD provides a standardized system delivering simple, reproducible access to a growing number of data "recipes" for automatic data retrieval and processing. Each data recipe contains the set of instructions used to obtain, transform, and standardize genomic datasets, providing fast, easy access to processed genomic data.

Put simply, the GGD tools provide a simple, new approach for reproducible data recipe searching, retrieval, processing, and installation on one's computer. GGD utilizes Amazon cloud storage for data package caching, allowing for rapid data retrieval and processing. GGD allows and encourages community contribution, with multiple tools provided to help scientists easily develop and contribute new data recipes to the GGD ecosystem. This process is further simplified by the application of a continuous integration system within the GGD infrastructure for automatic data recipe testing, uploading data recipe to the cloud, and other system-wide tests. Moreover, GGD is structured to manage diverse -omics data, allowing for the extensive use of multiple data structures and types.

We will present the current GGD framework, existing recipes, and describe future plans for new functionality, along with the development, collection, community contribution, and distribution of data recipes. We anticipate that GGD will become a standard, community-driven ecosystem for -omics data access and data reproducibility.

## PgmNr 1701: WGS sequencing at scale: Experiences from a national genome sequencing project.

### Authors:

P. Fox; A. Hodgkins; D. Bunford-Jones; M. Nelson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Congenica Ltd, Biodata Innovation Centre, Wellcome Genome Campus, Hinxton, CB10 1DR, UK

---

### Introduction:

The UK 100,000 Genomes Project (100K) was launched to develop a national healthcare service based on whole genome sequencing (WGS) in order to halt the diagnostic odyssey experienced by patients and enable system wide transformation.

Congenica's platform was chosen as a clinical decision support platform provider for the 100K in 2015 and, after rigorous competitive evaluation, was announced in 2018 as the exclusive provider for the UK National Genomic Medicine Service (GMS)

**Objectives:** To provide a national clinical decision support platform capable of processing WGS at scale for the new UK GMS

### Methods:

From experience gained in deploying the 100K system the following activities were carried out to allow the system to scale and be operated at the levels required

Database and application code performance engineering to ensure that data was handled efficiently and that users were presented with the necessary data for clinical decisions to be made effectively and quickly.

Moving the system from static hardware to an auto scaling cloud based securely deployed in a cloud environment. The migration from a fixed data centre to a cloud environment enabled a high availability system to be engineered but also hardware to be used on demand rather than incurring high fixed capital costs.

Examining user and administrative actions to simplify and automate where possible, allowing large number of samples to be processed and acted upon with minimal human interaction.

### Results:

Congenica's clinical decision support platform has been successfully deployed on AWS infrastructure to support the new UK GMS. The expected processing capacity of 200+ WGS/day has been met, with flexibility to increase to 400 as required. The system has been iteratively designed to provide the necessary secondary, tertiary analysis and decision support systems to enable physicians and scientists to efficiently interrogate phenotypic and genomic data for diagnosis. The GMS and Congenica's clinical decision support platform is now being rolled out nationally.

### Conclusions:

Through careful engineering and working with users to gain an in depth understanding of the basis of clinical decision making processes it has been possible to produce a secure scalable clinical decision support platform that provides health professionals with standardised auditable workflows, integrated tools and a plethora of contextually relevant data and information to enable accurate clinical decision-making.

## PgmNr 1702: Explore expression patterns of blood cells for patients of non-hematological diseases.

**Authors:**

W. Hsieh; S. Li

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Tsing Hua University, Hsinchu City, Taiwan, Taiwan

---

Gene expression profiles are often studied for tissues at illness. It is interesting to know how the expression patterns of blood cells are affected when people get tumor at certain tissue or are ill with systemic problems such as chronic fatigue. Our study aims at finding the natural grouping of gene expression patterns of blood cells and looking for the common characteristics across diseases. It is also interesting to know if there are any disease-specific markers present in the blood.

Blood expression profiles are generally not available from studies of non-hematological diseases. We collected four expression datasets of moderate sizes from the GEO database. All of them are from peripheral blood cells profiled on the same array platform, Affymetrix HG-U133A, and none of them are from hematological diseases. To better identify potential patterns, we only consider the studies with sample size large enough. Control samples are also provided from each study for contrast. It also facilitates our normalization step.

We propose an iterative clustering strategy based on Sparse K-means and BUS (Batch effects correction with Unknown Subtypes) in search of subgroups of general population and biomarkers. The simulation based on specific design approved the effectiveness of the methodology. The real data analysis explored the difference across patients of lung cancer, chronic fatigue, peripheral arterial disease and different types of breast cancer, and there seemed to be a strong signal from malignant breast cancer in the blood.

## PgmNr 1703: The complete linear assembly and methylation map of human chromosome 8.

### Authors:

G.A. Logsdon <sup>1</sup>; M.R. Vollger <sup>1</sup>; S. Koren <sup>2</sup>; A. Rhie <sup>2</sup>; D. Porubsky <sup>1</sup>; K.M. Munson <sup>1</sup>; C. Baker <sup>1</sup>; U. Surti <sup>3</sup>; K.H. Miga <sup>4</sup>; A. Phillippy <sup>2</sup>; E.E. Eicher <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; 2) Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; 3) Department of Pathology, University of Pittsburgh School of Medicine, and University of Pittsburgh Medical Center, Pittsburgh, PA 15213, USA; 4) UC Santa Cruz Genomics Institute, University of California, Santa Cruz, CA 95064, USA; 5) Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

---

Obtaining a complete high-quality, sequence-resolved human genome is essential to understand the genetic basis of human health and disease. Recent efforts to assemble human chromosomes from telomere to telomere have led to the successful assembly of the X chromosome and draft assemblies of many others (Miga, Koren et al., unpublished). However, the assembly of a human autosomal chromosome has not yet been completed. Here, we utilize a combination of long-read sequencing technologies, including Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio), as well as a novel sequence assembly algorithm we developed, to complete the linear assembly of human chromosome 8. Specifically, we generate 20-fold coverage of ultra-long-read ONT data (16.6X > 100 kbp) and over 100-fold coverage of PacBio long-read data from the human hydatidiform mole, CHM13. We barcode each ultra-long ONT read with a set of singly unique nucleotide k-mers (SUNKs), which occur only once in the human genome and distinguish regions of the genome from one another. Then, we apply a novel assembly algorithm to link the barcoded ONT reads together, as well as a segmental duplication assembly algorithm to resolve sequence collapses, resulting in the linear assembly of the chromosome 8 centromere and two major segmental duplication blocks on the p- and q-arms. Using 24X PacBio HiFi (high-fidelity) reads, we polish each subassembly and incorporate these into the CHM13 assembly of chromosome 8. We validate the assembly with several orthogonal technologies, including BioNano Genomics, Strand-seq, targeted BAC sequencing, and HiFi reads. Using the polished chromosome 8 assembly, we identify methylated DNA bases, creating a chromosome-wide methylation map. We find that methylated cytosines are specifically enriched at the pericentromeric transitions, consistent with the repressed gene expression reported in these regions. Together, our work reveals the complete linear sequence of chromosome 8 and provides a framework for understanding chromosome-level biology.

## **PgmNr 1704: Toward a joint standardized “default” transcript set: Matched annotation from the NCBI and EMBL-EBI (MANE).**

### **Authors:**

J. Morales <sup>1</sup>; J. Loveland <sup>1</sup>; S. Pujar <sup>2</sup>; A. Astashyn <sup>2</sup>; R. Bennett <sup>1</sup>; C. Davidson <sup>1</sup>; O. Ermolaeva <sup>2</sup>; C. Farrell <sup>2</sup>; L. Gil <sup>1</sup>; J.M. Gonzalez <sup>1</sup>; V. Joardar <sup>2</sup>; M. Kay <sup>1</sup>; V. Kodali <sup>2</sup>; K. McGarvey <sup>2</sup>; A. McMahon <sup>1</sup>; S. Rangwala <sup>2</sup>; G. Threadgold <sup>1</sup>; F. Cunningham <sup>1</sup>; A. Frankish <sup>1</sup>; T. T. Murphy <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) EMBL-EBI, Hinxton, Cambridge, United Kingdom; 2) National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD, U.S.A.

---

Understanding the impact of clinically-relevant variants is dependent on accurate and comprehensive annotation of genes and transcripts. However, the large number of alternatively spliced transcripts per locus and the lack of standardized “default” transcripts for clinical reporting and for display across resources present challenges in the clinical context. To address this, EMBL-EBI and NCBI, the two providers of comprehensive and complementary human transcript annotation, are working together on the Matched Annotation from the NCBI and EMBL-EBI (MANE) project. This new initiative aims to generate a genome-wide transcript set that: 1) matches GRCh38, 2) includes pairs of Ensembl/GENCODE (ENST) and RefSeq (NM) transcripts that are 100% identical at the CDS and both UTRs and 3) is high-confidence based on conservation, expression and overall support. The initial output, the MANE Select set, will include one transcript at each protein-coding locus across the genome that is representative of biology at that locus. To select transcripts, each Institute developed independent computational approaches, taking into account evidence of functional potential such as expression levels, evolutionary conservation, and clinical significance. Transcript ends are defined using CAGE data from the FANTOM consortium and polyA data from conventional and nextGen sequencing efforts. Curators from both groups provide a layer of manual review to ensure that automated selections are ideal or to make transcript decisions for complex loci. The MANE Select set will be fairly stable, only allowing updates if absolutely required. To encourage widespread adoption as “default” on browsers and in clinical workflows, we are working closely with clinical partners to prioritize genes deemed to have high gene-disease validity. Currently, we have released MANE Select transcripts for 53% of protein-coding loci and expect an increase to ~ 75-80% by the autumn, with the ultimate goal of achieving genome-wide coverage by 2020. In addition, we intend to release an expanded set of transcripts, the MANE Plus set, to include additional transcripts per locus that capture significant features not present in MANE Select transcripts. Released transcripts are available on all genome browsers and on the NCBI’s FTP site. This work is supported by: Wellcome Trust-WT200990/Z/16/Z; EMBL-Core-Funds; NIH-U41HG007234 and NIH-Intramural-Research-program.

## PgmNr 1705: 5utr: A framework for the interpretation of 5' UTR variants in rare diseases.

### Authors:

S. Pajusalu <sup>1,2,3</sup>; N.J. Lake <sup>1,4</sup>; M. Lek <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Yale School of Medicine, New Haven, CT; 2) Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia; 3) Department of Clinical Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia; 4) Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Australia

---

The importance of untranslated regions (UTRs) in the regulation of gene transcription and translation has been widely appreciated. However, the UTR variants associated with Mendelian disorders have been understudied and underreported in comparison with protein-coding variants despite being captured by exome sequencing. Across all Mendelian disorders, there are less than 300 pathogenic 5' UTR variants reported in ClinVar and medical literature. This is due to our inability to interpret the functional impact of UTR variants and a lack of clear interpretation guidelines. In addition, computational tools applied on 5' UTR variants rarely consider surrounding sequence context, which is critical to assess the functional impact on the upstream open reading frames (uORFs), secondary structures, the effect on translational efficiency, Kozak sequence, and transcription factor binding sites. Taken together, these known functional elements in the 5' UTR can form a framework to assist the annotation and interpretation of rare UTR variants. We have implemented this framework called 5utr and plan to release it as a Variant Effect Predictor plugin.

To validate this framework we annotated a set of published pathogenic 5' UTR variants compared to common benign variants in ClinVar. Approximately 50% of pathogenic variants generated a novel start codon compared to 1% of benign variants. Variants generating novel uORF overlapping with the main ORF have the largest functional effect among 5' UTR variants. Also, reported pathogenic variants had a lower mean predicted translational efficiency and a trend to be more likely in TF binding sites. In contrast, changes in minimum free energy of dsRNA and G4 quadruplex in RNA secondary structure did not differ significantly between variant sets. Interestingly, pathogenicity prediction tools separated pathogenic and benign variants, however the differences were not sufficient for classifying each individual variant.

We show the utility of the framework by classifying variants reported as disease-causing in the literature, as well as present results from large neuromuscular disease cohorts where we have applied this framework to improve the interpretation of UTR variants. A framework to predict the functional impact of UTR variants will become increasingly important as routine efforts to re-analyze existing exome data in undiagnosed Mendelian disorders will require additional novel approaches to improve the overall diagnostic yield.



## **PgmNr 1706: Development of a clinical summary tool that reduces clinical note review time and standardizes phenotype terminology.**

### **Authors:**

J. Salvo <sup>1</sup>; W. He <sup>1</sup>; F. Xia <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Clinical Genomics Interpretation, Baylor Genetics, Houston, TX; 2) Human and Molecular Biology, Baylor College of Medicine, Houston, TX

---

Clinical laboratory testing requires accurate evaluation of physician notes to interpret patient phenotypes. This has become especially important with whole exome sequencing (WES), as it can be used in the molecular diagnosis of a wide range of diseases. As WES testing continues to become more prevalent, a need arises for both efficient clinical note interpretation, and standardization of phenotype terminology. We have developed a computer application, the Clinical Summary Clerk (CSC), that improves the rate and accuracy of clinical note interpretation using autocomplete entry fields and phenotype synonym matching. By using the terminology provided by the Human Phenotype Ontology (HPO), the CSC combines synonyms and standardizes patient phenotypes. Phenotype language is important for patient database consistency and using HPO terms improves querying ability by reducing the number of missed patients due to unknown synonymous or misspelled phenotypes. Additionally, variant prioritization software, such as Codified and Emedgene, often require a list of HPO terms to describe patient phenotypes. The CSC helps provide these terms without any additional time input, as list files are automatically generated. With the reduced time for HPO term or synonym lookup, quick phenotype list editing, and automated file generation, the CSC can reduce time spent summarizing clinical notes by more than 30%. While human interaction is still required for clinical interpretation, it is important to have tools like the CSC to reduce human error and improve efficiency.

## PgmNr 1707: Comprehensive haplotype resolved MHC sequences from whole genome shotgun sequencing from single individual.

### Authors:

J. Chin <sup>1</sup>; A. Diltthey <sup>2</sup>; A. Fungtammasan <sup>1</sup>; S. Garg <sup>3</sup>; E. Garrison <sup>4</sup>; M. Rautiainen <sup>5</sup>; M. Tobias <sup>6</sup>; J. Wanger <sup>7</sup>; Q. Zeng <sup>8</sup>; J. Zook <sup>7</sup>; The MHC team for Pan-genomics in the Cloud hackathon 2019

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) DNAnexus, Inc., Mountain View, California.; 2) Institute of Medical Microbiology University Hospital of Dusseldorf; 3) Department of Genetics, Harvard Medical School, Boston, MD; 4) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK; 5) Center for Bioinformatics, Saarland University, Saarbrücken, Germany; 6) Max Planck Institute for Informatics, Saarbrücken, Germany; 7) Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD; 8) LabCorp, Inc., NC

---

During the three-day Hackathon organized by UCSC gathering researchers and developers to explore approaches for pan-genomic DNA sequencing analysis in March 2019, we formed a team targeting resolving the MHC regions of a single individual HG002. The recent advance in long-read DNA sequencing technologies has made routinely assembling new human genomes less-daunting tasks than before so we can get a comprehensive view of each human genome in the near future. Meanwhile, not all regions of a human genome will be resolved equally due to variable complexity of the genome sequences. Major Histocompatibility Complex region is one of the many such examples. Given the medical importance of this region, various targeted approaches have been developed before to get additional haplotype resolved sequences that currently are as ALT contigs in GRCh38. In this work, we explore the possibility to reconstruct the haplotype resolved contigs for MHC using just single individual whole genome shotgun sequencing data.

While there are already many de novo human genomes published recently, most of them do not have fully resolved contigs for the MHC region due to either the accuracy or read-length limitation. Our work examines all currently available data from multiple sequencing vendors (10x Genomics, Pacific Biosciences, Oxford Nanopore Technologies) for HG002 collected by the Genome In A Bottle project. With the multiple independent sequencing datasets to complement to each other, we derive a strategy to increase the accuracy on phasing the SNPs and reads for a haplotype-resolved assembly. The multiple technology approaches overcome the accuracy and read-length limitation of each technology used along. We build reproducible pipelines in Jupyter notebooks for assembling de novo contigs, one for each haplotype, spanning the whole MHC region. We construct variant graphs representing intricate large-scale differences to a reference. The SNP-phasing accuracy is validated with the phased variants using the maternal and the paternal genomes. A full spectrum of all variations is identified in the MHC region from the variant graphs and contigs. We hope such comprehensive variation catalog of the MHC region will lead to the insight into the associated biology. Our reproducible pipeline ensures the reproducibility for such complicated bioinformatics tasks and makes the approach re-usable for future work on building better MHC genomic sequences for future pan-human genome projects.

## PgmNr 1708: Accurate, scalable and integrative haplotype estimation.

**Authors:**

O. Delaneau

[View Session](#) [Add to Schedule](#)**Affiliation:** Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

---

Combinations of alleles carried on the same chromosome, called haplotypes, are important genetic features to consider in disease or population genetics studies. As they cannot be experimentally measured, computational methods for their statistical estimation are therefore routinely used. Nowadays, the number of human genomes being genotyped increases exponentially and haplotype estimation methods able to process this amount of data in reasonable times are therefore required.

In this work, we present a new method, SHAPEIT4, which substantially improves over other methods to process large genotype and sequencing data sets. It notably provides very accurate haplotypes and has the nice property of having running times that scale sub-linearly with sample size. We compared SHAPEIT4 with widely used software such as Beagle5, Eagle2 and SHAPEIT3 on a gold standard data set, the UK biobank containing half a million samples. At this scale, SHAPEIT4 is able to provide highly accurate haplotype reconstructions in relatively short running times that are often correct across entire chromosomes.

In addition, we also describe an additional extension of the SHAPEIT4 core model. We notably show how it can conveniently incorporate phasing information naturally present in paired-end, long or barcoded sequencing reads as those provided by Illumina Hi-seq, PacBio or 10x Genomics sequencing technologies, respectively. Using this data together with population level information results in haplotype reconstructions for whole genome sequenced individuals with error rates as low as 0.1%.

Overall, SHAPEIT4 is particularly well suited to fully exploit the potential of the next generation of large datasets being generated using either SNP arrays or new sequencing technologies.

## PgmNr 1709: PoolHapX: De novo haplotype reconstruction from pooled pathogen next-generation sequencing data.

### Authors:

J. He <sup>1</sup>; C. Cao <sup>1</sup>; L. Mak <sup>1</sup>; J. Wang <sup>2</sup>; K. Ye <sup>3</sup>; D. Jeffares <sup>4</sup>; Q. Long <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada; 2) Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, Illinois, America; 3) Department of Automation, Xi'an Jiao Tong University, Xi'an, Shan Xi, China; 4) Department of Biology, University of York, York, England; 5) Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada

---

Pooled sequencing (pool-seq) is a next-generation sequencing (NGS) strategy where the genomes of several individuals from a population are grouped together and bulk-sequenced. Pool-seq provides an efficient and cost effective alternative to genome sequencing of individuals or single cells, especially in contexts where pathogen genomes are inherently mixed. To determine the frequencies of individual-level polymorphisms and linkage disequilibrium (LD) from a population, the aggregated variation data must be de-convoluted in silico into linkage blocks, or haplotypes, that ideally span whole genomes. Global haplotype reconstruction for genomic sequences estimates the likelihood that variants called from sequencing data occur in the same individual, thereby parsing bulk sequencing data at the single-molecule level. The PoolHapX program approximates the genotypic resolution of single-cell sequencing using only pool-seq data by integrating population genetics models with genomics algorithms to reconstruct haplotypes. PoolHapX first determines locations of long-range linkage uncertainty using short NGS reads, and then divides the full genome into shorter regions. Regional haplotypes are solved for and joined together for a parsimonious global distribution of haplotypes. Complex mixes of whole-genome HIV sequences, each representing the 'pool' of viruses from a patient, were simulated to mimic the diversity found in a HIV outbreak. 100 variant positions were scattered across the genome. Applying a prototypic version of PoolHapX to NGS reads simulated from these patient HIV populations resulted in an average of 99% accurate reconstructions of regional haplotypes. (J. He, C. Cao, L. Mak contributed equally)

## PgmNr 1710: MatchY: The Yale Matchmaker Exchange Node.

### Authors:

A. Zhao <sup>1</sup>; S. Pajusalu <sup>1, 2</sup>; N.J. Lake <sup>1, 3</sup>; S. Huang <sup>1, 4, 5</sup>; M. Lek <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Yale School of Medicine, New Haven, CT, USA; 2) Department of Clinical Genetics, Institute of Clinical Medicine, University of Tartu and Tartu University Hospital, Tartu, Estonia; 3) Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Australia; 4) Affiliated Hospital of Nantong University, Nantong, 226001, China; 5) The First Affiliated Hospital, Nanjing Medical University, Nanjing, 211166, China

---

Over the last few decades, researchers have sought to establish causal links between gene variants and rare Mendelian diseases. Identifying more rare disease patients with variants in the same gene can help validate the causality of novel disease genes. To bolster automated data sharing among rare disease centers across the world, the MatchMaker Exchange (MME) federated network has allowed researchers and clinicians to both share and query cases based on genotypes and/or phenotypes. Moreover, an increase in the number of MME nodes and updates of current node features can further augment data sharing efforts. So far, the MME network contains over 65,000 cases among seven member organizations and eight additional participating institutions.

Here, we present a new node of the MME network: *MatchY*, an open-source, web-accessible, and user-friendly platform for data sharing. This node implements the MME Application Programming Interface (API) and currently contains a substantial cohort of neurological and neuromuscular diseases. The scoring algorithm contained within *MatchY* allows users to match on gene, genotype, and phenotype and facilitates free-text phenotype matching to Human Phenotype Ontology (HPO) terms. In addition, we have added the flexible feature to match on equivalent functional classes of variants instead of restricting matches to the exact variants. For example, matching a nonsense variant with loss of function variants (i.e. nonsense, frameshift and essential splice) within the respective gene. Lastly, the implementation using an elasticsearch data will empower more advanced internal matching as this node grows in size. In summary, we believe *MatchY* can galvanize rare disease research by connecting researchers to more rare disease cases through the addition of cases from the Yale Center for Mendelian Genomics.

## PgmNr 1711: Circulating cell-free DNA detection on the GenapSys sequencer for organ transplant applications.

**Authors:**

M. Fallahi; S. Paliwal; A. Nabi; M.R. Barmi; H. Nezamfar; S. Bagchi; E. LoPrete; B. Dong; X. Gomes; M. Jouzi; A. Thomas; N. Tangprasertchai; H. Esfandyarpour

[View Session](#) [Add to Schedule](#)

**Affiliation:** GenapSys Inc., Redwood City, CA

---

Circulating cell-free DNA (cfDNA) can be used as a biomarker for transplant rejection in organ transplant patients. A cfDNA-based assay offers several advantages over traditional invasive tissue biopsy approaches, including lower cost, reduced complications, and reduced interpretation bias. Next generation sequencing (NGS) technology has enabled the accurate determination of the fraction of donor derived cfDNA in a transplant recipient's blood sample, which can be correlated with the potential for organ rejection.

GenapSys has developed a novel, portable, and cheap NGS technology that uses electronic detection to sequence DNA. Implementation of a cfDNA assay for organ transplant rejection on this platform could lead the way to enabling low cost and accessible point-of-care testing for transplant patients in the future.

We evaluated a comprehensive NGS-based cfDNA assay for organ transplants on the GenapSys sequencing platform and developed bioinformatic algorithms for determination of donor-derived cfDNA fraction. Amplicon libraries that target >200 single-nucleotide polymorphisms were generated from samples which contained donor-derived DNA at a range of different fraction levels, including 0% (control), <0.5% and >1%. Libraries were sequenced on the GenapSys sequencing system and sequencing data was analyzed using a custom bioinformatics approach. Briefly, adapter sequences were removed using cutadapt v1.16 and trimmed reads were aligned to the hg38 reference genome using BWA-MEM. The allele frequencies of heterozygous and homozygous SNPs were calculated using BCFtools mpileup and calibrated with reference to control samples.

We demonstrate successful implementation of the cfDNA assay on the GenapSys sequencing platform and the ability to accurately detect donor-derived cfDNA fractions at expected values, including for fractions lower than 0.5%. Additionally, sequencing replicates for a given library correlated well, indicating robust and reproducible performance.

## PgmNr 1712: Finding complex pathogenic variants using nanopore long reads.

### Authors:

S. Mitsuhashi <sup>1</sup>; S. Ohori <sup>1</sup>; L. Ming <sup>1</sup>; M.C. Frith <sup>2,3,4</sup>; N. Matsumoto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Yokohama City University, Yokohama, Japan; 2) 1.Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan; 3) Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan; 4) Computational Bio Big-Data Open Innovation Laboratory (CBBD-OIL) , Tokyo, Japan

---

A variety of complex genomic changes (e.g. tandem-repeat expansions/contractions, chromothripsis, transposable element insertions) are known to cause rare human genetic diseases. These changes are hard to find using conventional methods (e.g. short read sequencing). We have been developing new methods to find such complex genomic changes using long read sequencing data from patients with rare genetic diseases. We will present our two new pipelines; 1) *tandem-genotypes* which can predict tandem repeat copy number changes and prioritize possible pathogenic changes by comparing to healthy controls, 2) *rearranged-sequence-clumps* which can find patient-only rearrangements by subtracting rearrangements present in controls, and reconstruct them. Our methods are easy to use, applicable to multiple long read sequencing platforms and in fact can efficiently find pathogenic variants. We hope wider application of long read technologies will find pathogenic complex genomic changes that have been overlooked by short read sequencing.

## PgmNr 1713: AdaFDR: A fast, powerful, and covariate-adaptive approach to multiple hypothesis testing.

### Authors:

M. Zhang <sup>1</sup>; F. Xia <sup>1</sup>; J. Zou <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Electrical Engineering, Stanford University, Stanford, CA.; 2) Department of Biomedical Data Science, Stanford University, Stanford, CA.; 3) Chan-Zuckerberg Biohub, San Francisco, CA

---

Multiple hypothesis testing is an essential component of modern data science. Its goal is to maximize the number of discoveries while controlling the fraction of false discoveries. In many settings, in addition to the p-value, additional information/covariates for each hypothesis are available. For example, in eQTL studies, each hypothesis tests the correlation between a variant and the expression of a gene. We also have additional covariates such as the location, conservation and chromatin status of the variant, which could inform how likely the association is to be due to noise. However, popular multiple hypothesis testing approaches, such as Benjamini-Hochberg procedure (BH) and independent hypothesis weighting (IHW), either ignore these covariates or assume the covariate to be univariate. We introduce AdaFDR, a fast and flexible method that adaptively learns the optimal p-value threshold from covariates to significantly improve detection power. On eQTL analysis of the GTEx data, AdaFDR discovers 32% and 27% more associations than BH and IHW, respectively, at the same false discovery rate. We prove that AdaFDR controls false discovery proportion, and show that it makes substantially more discoveries while controlling false discovery rate (FDR) in extensive experiments. AdaFDR is computationally efficient and can process more than 100 million hypotheses within an hour and allows multi-dimensional covariates with both numeric and categorical values. It also provides exploratory plots for the user to interpret how each covariate affects the significance of hypotheses, making it broadly useful across many applications.



## PgmNr 1714: pCNN: A phylogeny-regularized convolutional neural network for predictive modeling of microbial data.

### Authors:

L. Chen <sup>1,4</sup>; Y. Wang <sup>1</sup>; Y. Jiang <sup>2</sup>; J. Chen <sup>3</sup>; X. Qin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science and Software Engineering, Samuel Ginn College of Engineering, Auburn University, Auburn, AL, 36849; 2) Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC 27599; 3) Division of Biomedical Statistics and Informatics and Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 4) Department of Health Outcomes Research and Policy, Auburn University Harrison School of Pharmacy, Auburn, 36849

---

With the development and decreasing cost of next-generation sequencing technologies, the study of the human microbiome has become an important research field accounting for huge potentials in clinical applications such as drug response predictions, patient stratification, and disease diagnosis. Thus, it is essential and desirable to build a microbiome-based prediction model for clinical outcomes based on microbiome profiles consisting of taxon abundance and a phylogenetic tree. One important characteristic is all microbial species are not uniformly distributed on the phylogenetic tree but tend to be clustered at different phylogenetic depths. Thus, the phylogenetic tree represents a unique correlation structure of microbiome, which might be an important prior information for the prediction. However, prediction methods that consider the phylogenetic tree in an efficient and rigorous way are under-developed.

We develop a deep learning prediction method pCNN (Phylogeny-Regularized Convolutional Neural Network) to predict phenotype/outcome (either quantitative or qualitative) by using both the taxon abundance and the phylogenetic tree. pCNN designs different convolutional layers to capture different taxonomic ranks (e.g. species, genus, family, etc.) with different convolutional kernels on each convolutional layer to capture the signals of evolutionary close microbiome species in a local receptive field. Together, the convolutional layers with its built-in convolutional kernels capture microbiome signals at different taxonomic levels while encouraging local smoothing induced by the phylogenetic tree. Comprehensive simulation studies demonstrate that pCNN outperforms other competing methods in the scenarios of large signal cluster or the dense signal clusters, while pCNN is still on par with other methods in other scenarios when the tree is informative. Moreover, pCNN has comparable prediction performance with other methods when the tree is non-informative. We further apply pCNN to two real datasets and find pCNN could predict age and gender more accurately compared to other methods.

## PgmNr 1715: Phylogeny-enhanced prediction tools for microbiome data.

**Authors:**

T. Wang; S. Wang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY

---

Associations between microbial community and human health have been well-studied and established. Sophisticated statistical methods have been developed to examine associations between microbiome data and health outcomes utilizing the distinctive features of microbiome data, i.e., large number of compositional and sparse taxa and the phylogenetic tree structure for taxa. Researches have also been done to predict health outcomes using microbiome data. However, these studies applied existing prediction tools directly, such as random forest (RF) or k-nearest-neighbors (KNN), when these methods ignore the unique feature in microbiome data, namely the phylogenetic tree structure. Here we propose a general framework that 1) enhances the existing prediction tools by accounting the phylogenetic tree structure through re-weighting the contribution of samples using multiple microbial distance metrics that capture various aspects of association between microbiome data and health outcomes, and 2) optimally selects one microbial distance metric that most accurately predicts outcomes of interest. We demonstrated the performance of the proposed general framework in enhancing in RF and KNN and developed phylogeny-enhanced RF (PhyEn-RF) and phylogeny-enhanced KNN (PhyEn-KNN). Simulation studies suggest the superior performance of PhyEn-RF and PhyEn-KNN over the original RF and KNN when different associations between microbiome and health outcomes were considered. We applied PhyEn-RF and PhyEn-KNN to gut microbiome data of malnourished Bangladeshi children (ENA: PRJEB5482) and found that the proposed methods consistently outperform the original ones in predicting several microbiome-related health outcomes using microbiome data.

## PgmNr 1716: Using mutual information to detect correlations between nucleotide positions in noncoding DNA.

### Authors:

J. Dannemiller

[View Session](#) [Add to Schedule](#)

**Affiliation:** Psychological Sciences, Rice University, Houston, Texas.

---

**Background:** The base at a given position in a coding DNA sequence is correlated with the bases at downstream positions because of biases in codon usage. This leads to a 3 bp periodic signal in base occurrence in coding DNA. Noncoding sequences lack this periodicity, but one might still expect to find some degree of predictability from upstream to downstream locations because of various conserved features (e.g., transcription factor binding site motifs in promoters) or because of sequence-dependent 3D conformational constraints. The goal of this research was to examine noncoding DNA sequences for evidence of significant short-range, distance-dependent correlations between base locations. **Methods:** Bias-corrected Mutual Information (bcMI) was used to quantify the dependency between any two locations in sequences of a) promoters, b) introns, and c) exons (at least 20,000 sequences for each type). Exon sequences were included as a check on the use of bcMI as the measure of dependency. bcMI was estimated at distances of 1 to 21 bp from various initial locations. Randomly shuffled sequences were used as negative controls; they showed bcMI values near 0 at all distances. **Results:** For all sequence types, bcMI generally fell (the downstream base became less predictable) with increasing distance between the two locations. There was, as expected, a clear 3 bp periodicity in the bcMI vs. distance function in exon sequences. This 3 bp periodicity was absent from the promoter and intron sequences, but there was a trend-reversing rise in bcMI at a distance of 6 bp in both sequence types with weaker evidence of another signal at a distance of 12 bp. Remarkably, this rise in predictability at a distance of 6 bp survived random misalignments (+/- 40 bp range) of the promoter sequences with respect to each other suggesting that it is a delocalized property of promoter sequences. The 6 bp signal in bcMI was also artificially produced by randomly inserting hexamer double tandem repeats in the shuffled promoter sequences. **Conclusions:** This rise in predictability at a distance of 6 bp in noncoding DNA can be seen in previous reports (e.g., Grosse et al., 2000), but its significance as a general property of noncoding DNA has heretofore not been articulated. It could reflect the presence of many short tandem repeats in noncoding DNA. Another possibility is that it reflects constraints on nucleotide sequence imposed by energetic or structural properties of the helical molecule.

# PgmNr 1717: Validation of germline structural variation reporting from clinically oriented ultra-deep amplicon sequencing.

## Authors:

S. Adams <sup>1,2</sup>; J. Larusch <sup>2</sup>; M. Haupt <sup>2</sup>; E. Orlova <sup>2,3</sup>; D. Whitcomb <sup>3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Shenandoah University School of Pharmacy, Fairfax, VA.; 2) Ariel Precision Medicine, Pittsburgh, PA; 3) University of Pittsburgh, Pittsburgh, PA

---

## Introduction

Structural variants (SV) are genetic variation spanning >50bp. Pathogenic SVs are found in >0.24% of human genomes. Techniques that estimate SV from next-generation sequencing (NGS) data have emerged & exemplify the versatility of genome & exome sequencing. These methods rely on depth of coverage (DOC) variation, split reads, & paired-end mapping to find SV breakpoints in areas of relatively uniform coverage. Amplicon sequencing is a robust & affordable alternative to WGS & WES. The extreme DOC for target regions supports novel variant discovery & highly accurate variant calling. Determining SV from amplicon sequencing is challenging due to variation in DOC across a target region & non-random read start positions. We developed & validated a DOC based method (ArielCNV) for amplicon-resolution SV estimation from a proprietary amplicon panel based on the Illumina AmpliSeq chemistry (San Diego, CA).

## Development

ArielCNV is implemented in the Nim programming language & utilizes a stepwise approach for CNV estimation, loosely adapted from the CNVPanelizer method described by Oliveira, et al. The tool uses input from a sample BAM file (unknown), a bed file that defines the contigs & a reference BAM file. Aligned read counts for the sample & reference samples meeting a MAPQ cutoff (e.g. 30) are counted for each contig. Read counts are normalized using the trimmed mean of M method. Bootstrap analysis of contigs within a given gene are subsampled & used to calculate a DOC ratio. A mean ratio & 2.5% - 97.5% quantile range are reported for each gene. Individual contig ratios are compared to baseline to estimate contig level copy changes by comparing the z score relative to all samples, then calculating a two-sided p value with subsequent correction with the Benjamini-Hochberg false discovery rate.

## Preliminary Validation

Samples of known CNV were sequenced with the proprietary Ariel panel. Cell lines/DNA samples were obtained from the NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research. Initial validation tested *CYP2D6* and *CEL*, which have SV and fusions with their neighboring pseudogene. Within four known samples, correct CNV detection was reported in all cases.

## Conclusions

ArielCNV is a novel method for the estimation of SV from amplicon sequencing. Preliminary validation showed concordance in known samples for *CYP2D6* and *CEL*. Future work will seek validation in other clinically relevant genes and clinical implementation of SV.

## PgmNr 1718: Doc2Hpo: A web application for efficient and accurate HPO concept curation.

### Authors:

C. Liu <sup>1</sup>; F. Kury <sup>1</sup>; Z. Li <sup>1</sup>; C. Ta <sup>1</sup>; K. Wang <sup>2,3</sup>; C. Weng <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical Informatics, Columbia University Medical Center, New York, NY.; 2) Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA.; 3) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA.

---

Phenotype terms from electronic health record (EHR) narratives can be used to improve the interpretation of sequencing data and expedite clinical diagnosis of patients with suspected Mendelian diseases. However, the process for clinical researchers to manually curate the standardized phenotype terms such as Human Phenotype Ontology (HPO) concepts from the clinical text remains laborious and can be error-prone.

To address the critical need for efficient and accurate HPO concept curation, we developed Doc2Hpo, a novel web application based on the human-computer collaboration design principle for semi-automatically extracting phenotype terms from the clinical text with automated concept normalization using HPO. Doc2Hpo automatically highlights clinical entities with their corresponding HPO mappings to facilitate manual review and revision if needed. Users can edit the HPO concepts automatically extracted by Doc2Hpo in real time, and export the final list of HPO concepts to gene prioritization tools for further analysis. Our evaluation showed that Doc2Hpo significantly reduced manual effort while achieving high accuracy in HPO concept curation.

Doc2Hpo is freely available at <https://impact2.dbmi.columbia.edu/doc2hpo/>. The source code is available at <https://github.com/stormliucong/doc2hpo> for local installation for protected health data.

## PgmNr 1719: Biological discovery and consumer genetics activate latent privacy risk in omics data.

**Authors:**

S.E. Brenner; Z. Hu

[View Session](#) [Add to Schedule](#)**Affiliation:** University of California, Berkeley, California.

---

Privacy risks from individuals' genomic data have garnered increasing attention. Sharing genomes without personal identifiers is common practice. However, recent studies and forensics underscored the ability to re-identify a person, using genomic relatives and quasi-identifiers, such as sex, birthdate and zip code. The additional availability of omics data, such as transcriptomics and methylation data, has implications for privacy, as it may also be linked to the genome, potentially allowing privacy breach. For example, sex and ethnicity information may be inferred directly from a genome, and the study may provide zip code. This could be linked to RNA-seq data from a disease study with attached birthdates and income. These combined quasi-identifiers may uniquely identify the person, and the study reveals the person's disease status.

Sequence reads from any experiment contain genetic variants, and thus can be directly linked to the genome. To avoid this risk, researchers may release only processed data instead of raw reads. For example, studies generating RNA-seq data may release only gene expression, isoform expression and exon read count data. However, a recent study suggested that gene expression data can be linked the genome based on genotypes inferred from expression QTLs (eQTLs). Splicing data has been reported as safe, as splicing QTLs (sQTLs) are 200 times lessas 0.5% abundant than as eQTLs, but we demonstrate that splicing data can be reliably linked to a single genome, from a population of thousands. Such a linking for DNase hypersensitive sites and gene expression data now enables the identification of a genome uniquely from the world population. This new risk has arisen due to growing biological knowledge, including from increased data that allow the identification of more molecular QTLs. Additionally, consumer genomics databases have made rendered many more individuals potentially recognizable, including via their relatives. Our study implies further hidden privacy risks in existing data, which will only manifest over time. While it is accepted that the security of many cryptographic security methods degrade over time, the need to preserve individuals' genomic privacy for their lifetime and beyond (for descendants) poses unique challenges to the effective sharing of high-throughput molecular data.

# PgmNr 1720: A sparse deep learning model for phenotype prediction.

**Authors:**

J. Chen; X. Shi

[View Session](#) [Add to Schedule](#)**Affiliation:** Dept. of Bioinformatics and Genomics, UNC Charlotte, Charlotte, NC.

---

**Motivation:**

Phenotype prediction has been widely utilized in many areas to help understand disease risks and susceptibility, and improve the breeding cycles of plants and animals. However, there are several barriers that make it extremely challenging to develop robust and powerful methods for phenotype prediction, including high dimensionality and collinearity in genomics data. Most of phenotype prediction are based on regularized statistical approaches which only consider linear relationships among genetic variants. Several recent studies have reported that deep learning has strong potentials to address high dimensional problems for prediction tasks in various domains including image analysis and natural language processing. Hence, we will explore novel deep learning models for phenotype prediction on high dimensional genomic data.

**Methods:**

In this study, we propose a Sparse Convolutional Predictor with Denoising Autoencoders (SCP\_DAE) for phenotype prediction. Our novel SCP\_DAE model takes the advantages of sparse convolutional network (CNN) and denoising autoencoder (DAE) architectures to address two critical challenges in high dimensional genomic data analysis. First, we utilize a sparse CNN to capture local structure information or correlations in genomic data and introduce model sparsity by incorporating an L1 regularization on the weight matrix to avoid overfitting. Second, we use an efficient representation of genotype data derived from DAEs in an unsupervised way and pretrain model layers to improve model performance.

**Results:**

We comprehensively evaluated the performance of our proposed models on a yeast genomic dataset that contains the genotypes of 28,820 genetic variants and 20 traits assessed in 4,390 samples. We first show that adding an L1 regularization to a CNN based predictive model can prevent overfitting and improve the prediction performance. Next, by adding DAE pretrained weights, our final model of SCP\_DAE significantly outperforms other methods including SCP without DAE, Lasso, elastic net, and random forest.

**Conclusion:**

This study demonstrates that a novel deep learning model based on sparse CNNs coupled with DAE pre-trained weights can significantly improve phenotype prediction on high dimensional genotype data. This new model for phenotype prediction can provide insights on how to better understand disease susceptibility and risks, and allocate resources for improving plant and animal breeding.

## PgmNr 1721: Expansion of the 1000 genomes project resources in the International Genome Sample Resource (IGSR).

### Authors:

S. Fairley; E. Lowy; G. Aldam; E. Perry; P. Flicek

[View Session](#) [Add to Schedule](#)

**Affiliation:** European Molecular Biology Laboratory - European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, CB10 1SD. United Kingdom

---

The 1000 Genomes Project created the largest fully public catalogue of human genetic variation, including openly available reference resources, which continue to be widely used. IGSR was established to maintain and expand those resources.

IGSR aims to make resources available on GRCh38, include new data types and add to the collection of openly consented samples.

The 1000 Genomes Project was based on GRCh37. IGSR has realigned the low-coverage whole genome and exome data from 1000 Genomes to GRCh38. These alignments have been used as the basis for calling variants from the 1000 Genomes data directly on GRCh38. This used GATK, BCFtools and FreeBayes for site discovery, GATK for genotyping and BEAGLE and SHAPEIT2, to produce an imputed and phased call set, which was shared in 2018.

Cell lines created for the 1000 Genomes samples continue to be used in data generation. The New York Genome Center (NYGC), funded by NHGRI, have sequenced the genomes of the 2,504 samples in the 1000 Genomes phase three panel to ~30x coverage. In addition, NYGC have shared a GATK call set generated from this data. IGSR hosts this data and makes it available alongside other data generated from 1000 Genomes cell lines and other openly consented samples.

Data in IGSR currently includes a wide variety of data types generated by the Human Genome Structural Variation Consortium (HGSVC), including Oxford Nanopore, PacBio and 10X Genomics data. The resources also include data on samples which were not present in 1000 Genomes. These are from the Gambian Genome Variation Project (GGVP), the Human Genome Diversity Project (HGDP) and the Simons Genome Diversity Project (SGDP).

Through our freely available analysis pipelines, we continue to extend the range of available open genomic data on GRCh38.

To support users in locating and using the data, our website contains a data portal, enabling exploration of available data by data type, population, collection or sample. Our website contains further information on data access and we also provide an email helpdesk, providing bespoke support for those using the data.

IGSR makes openly consented genomic data accessible to the wider community through collaboration



with data generation groups. We continue this work and remain open to establishing new collaborations where we can provide support for data sharing.

## PgmNr 1722: Ancestry clustering and classification in large-scale data using neural networks.

### Authors:

S. Gilhool<sup>1</sup>; H. Hakonarson<sup>1,2</sup>; P. Sleiman<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

It is well known that allele frequencies vary between populations of differing ancestry. This so-called population structure has an impact across multiple domains, including the diagnosis and treatment of disease in the clinic, and the identification of disease-associated loci in the lab. However, there is no publicly available way to quickly and easily answer the question, “Genetically, what is this individual’s ancestry?” In order to answer this question, we propose the use of two kinds of neural networks – an autoencoder to perform dimensionality reduction and clustering, and a multi-layer perceptron to perform classification.

An autoencoder is a type of deep neural network that operates by reconstructing its input data after passing it through a low-dimensional space that serves as a bottleneck. The fact that the data can be reconstructed implies that its more compact representation at the bottleneck is a meaningful way to encode the input data. We used the autoencoder to determine 2-D representations of SNP array data for tens of thousands of patients in the Children's Hospital of Philadelphia biobank. Plotting the results reveals distinct clustering by ancestry. We further used the autoencoder representations as inputs to the multi-layer perceptron classifier.

A multi-layer perceptron is a neural network that consists of multiple layers of artificial neurons. These successive layers are capable of learning complex features of the input data. Our network is designed to output probabilities of an individual’s membership in various racial or ethnic groups. Our approach achieved 98.7% accuracy in early tests with the sample restricted to Causasians and African Americans. Work to expand the number of classification groups is underway. Ultimately, we aim to return a probability distribution for each individual’s ancestry, which would be particularly useful in the case of admixed individuals.

## PgmNr 1723: Considering natural isotopes in peptide de novo sequencing: An implementation in R.

### Authors:

S. El Jadid <sup>1</sup>; R. Touahni <sup>1</sup>; A. Moussa <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ibn Tofail University, Kenitra, Morocco; 2) National School of Applied Science, Tangier, Morocco

---

Proteomics, as an -omic science, has made a massive progress in the past decade, generating a tremendous amounts of data that show considerable challenges for their interpretation. With the revolution of high resolution Mass Spectrometry, the mining of this flood of data for analysis and interpretation becomes possible. However, to supply the suitable biologically interpretation, an accurate identification of all proteins present in the sample is needed. Tandem Mass Spectrometry-based de novo peptide sequencing is a valuable tool for protein identification. Peptide de novo sequencing technology attempts to derive a peptide's amino acid sequence from the input spectrum. Several software based on different algorithms and programming language for de novo sequencing have been developed. To date, no tool performing de novo sequencing exists in R, the user-friendly programming language for biological researcher. Furthermore, existing tools doesn't handle isotopic information. It is popularly assumed that considering natural isotopes will considerably improve protein identification accuracy.

In this following work, we propose a novel R tool for de novo peptide sequencing. The tool implements UniNovo in R with considering the natural isotopes.

Our experimental results demonstrate that our tool significantly improves the accuracy of protein identification taking into account the isotopic information. In addition, our tool will be a valuable alternative for the R researcher community.

## PgmNr 1724: Uncovering causal noncoding regulatory variants in human disease through allelic protein-RNA.

### Authors:

S. Bao<sup>1,2,3</sup>; J. Wong<sup>1,2,3</sup>; E.D. Flynn<sup>1,2,3</sup>; C. Zhang<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Systems Biology, Columbia University, New York NY 10032, USA; 2) Department of Biochemistry and Molecular Biophysics, Columbia University, New York NY 10032, USA; 3) Center for Motor Neuron Biology and Disease, Columbia University, New York NY 10032, USA

---

It has become evident that a majority of genetic variants associated with human diseases reside in non-coding genomic regions which may play a role in regulating multiple steps of gene expression. A comprehensive catalog of precisely mapped regulatory variants (RVs) in human populations, along with the ability to determine disease-causing RVs in individuals, holds great promise for improving disease diagnoses and ultimately patient care. However, causal noncoding RVs remain extremely difficult to identify in most cases due to the insufficient resolution of genetic mapping and our incomplete understanding of gene expression regulation. In this study, we develop **approaches to identity causal noncoding RVs modulating post-transcriptional gene expression regulation, such as RNA splicing and stability**. Since RNA-binding proteins (RBPs) and their interactions with target transcripts play central roles in these regulatory steps, **SNPs affecting protein-RNA interactions represent one major source of disease-causing RVs**. We performed a proof-of-principle study to analyze CLIP data of over 100 RBPs in human cell lines from ENCODE and precisely mapped SNPs affecting protein-RNA interactions. In particular, the heterozygous SNPs showing allelic imbalance between reference and alternative alleles provide direct experimental evidence of altered protein-RNA interactions with potential impact on downstream post-transcriptional gene expression regulation and individual phenotypes. Furthermore, the concordance between the allelic imbalance of CLIP tags and changes in motif scores provides unbiased validation that the motifs reliably reflect RBP binding specificity and the impact of SNPs on protein-RNA interactions is direct. With this approach, we determined the specificity of a list of *de novo* candidate motifs for many RBPs that were previously poorly characterized. Our analysis also identified a list of SNPs as allelic protein-RNA interaction sites with high confidence and thus candidate RVs. Our approach overcomes the low signal-to-noise ratio associated with bioinformatic predictions of RBP binding sites and provides a valuable resource to annotate noncoding variants.

## PgmNr 1725: Covariate adjusted rare-variant association (CARVA) testing for millions of samples.

**Authors:**

R. Bohlender; Y. Yu; J. Chen; C. Huff

[View Session](#) [Add to Schedule](#)

**Affiliation:** Epidemiology, University of Texas MD Anderson Cancer Center, Houston, Texas.

---

As biobank scale sequencing datasets become more common, statistically rigorous and computationally efficient tools for the analysis of these data have become increasingly important. Covariate-adjusted permutation offers a robust and generalizable solution for assessing the statistical significance of rare variant association tests but has been underutilized due to computational concerns. Thus, analysis of biobank scale sequencing datasets has generally been limited to statistical analysis methods with analytical approximations. Here, we introduce the software package Covariate Adjusted Rare-Variant Association (CARVA), which is designed to conduct fast and memory efficient multithreaded association analysis of large sample sets, with support for highly scalable covariate adjusted permutation. CARVA currently supports 10 statistical tests, including CMC, SKAT, SKAT-O, WSS, VT, and VAAST. The core innovation in CARVA is a highly-optimized, adaptive permutation engine that implements two novel approaches for covariate-adjusted permutation of large datasets. First, CARVA permutes only minor allele carriers, which are the only samples that contribute to the statistical significance of a gene. Second, CARVA supports approximate binning of individuals with similar odds of being a case, reducing the number of categories to permute. In addition to adaptive permutation, these steps greatly reduce the computational burden of permuting large datasets. To limit memory usage, CARVA leverages sparse matrices where possible. CARVA is implemented in C++ using efficient linear algebra libraries, which further enhances memory efficiency and computational scalability. As such, it is possible to analyze a data set with one million samples on a 2015 MacBook Pro with a 2.2 Ghz Intel Core i7 and 16 GB of ram. Total run time for the analysis of a set of 500 genes in 500,000 cases and 500,000 controls, running on 5 threads was approximately 5 hours on the above laptop, with maximum memory usage at roughly 1.6 GB. Speed and memory usage scales roughly linearly with the number of threads. CARVA provides the means to conduct robust rare variant association studies on very large data sets that will be available in the foreseeable future.

## PgmNr 1726: Identifying effects of rare variants on gene expression with likelihood ratio test.

### Authors:

J. Li <sup>1</sup>; S. Hwang <sup>2</sup>; B. Han <sup>3</sup>; J.H. Sul <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA; 2) Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, Los Angeles, CA 90095, USA; 3) Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

---

Genome-wide association studies have discovered numerous non-coding genetic variants associated with complex diseases and traits. It has been hypothesized that those non-coding variants may affect diseases or traits by regulating gene expression, and expression quantitative trait loci (eQTL) studies have identified many common variants with regulatory effects. Recently, sequencing studies have started to identify roles of rare variants in complex diseases and traits, but regulatory effects of rare variants have not been thoroughly investigated. A traditional statistical approach to detect effect of a genetic variant through a single marker test may be underpowered for rare variants as their low occurrence limits power. To address this challenge, group-wise test methods that aggregate effect of rare variants in genes have been proposed. As these methods are susceptible to power loss when non-causal variants are included in tests, they often utilize prior information on variants such as minor allele frequency, Polyphen-2 or SIFT to remove non-causal variants from the analysis. In eQTL studies where many of eQTLs are in non-coding variants, prior information such as CADD scores and distance to a transcription start site may be used to prioritize causal non-coding variants. Here, we present LRT-q, a likelihood ratio test method that incorporates genotype data and prior information of genetic variants in a nonlinear manner to infer causal rare variants. It then uses this finding to identify rare variants associated with gene expression, assuming that individuals with abnormal gene expression are likely to enrich for causal rare variants. We demonstrated through simulations that LRT-q has higher power than other methods, including widely used SKAT-O. We also showed that LRT-q detected more novel genes regulated by rare variants that were not detected by common variants analysis in six tissues in Genotype-Tissue Expression (GTEx) by incorporating prior information about variants, in comparison with previous methods.

## **PgmNr 1727: Actionable Genomic Interpretation System (AGIS): A comprehensive and accurate knowledgebase system for variant clinical interpretation using ACMG-AMP standard guideline.**

### **Authors:**

W. Ding; C. Chen; J. Li; Q. Zhang; Y. Wan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Admera Health, South Plainfield, New Jersey.

---

With the continued development and adoption of massively parallel next-generation sequencing (NGS) technologies, one major hurdle in making precision medicine a reality is to turn genomic data into clinical information that can be easily used in medical practice. A truly effective solution at minimal requires: 1) a fully automated data analysis and interpretation workflow from raw data to clinically actionable report; 2) a comprehensive and high-quality database for fast and accurate clinical interpretation of variants; 3) a scalable and reliable IT infrastructure; 4) a clinically validated process that meets regulatory requirements; and 5) being demonstrated in real medical practice with great satisfaction. We have developed a web-based Actionable Genomic Interpretation System (AGIS) that has all above features with automated interpretation step and a manual adjustment step. Strictly following the 28 criteria provided by ACMG-AMP standards and guidelines, we also implemented an automated process to systematically integrate comprehensive information from Human Gene Mutation Database (HGMD), ClinVar, 1000 Genome Project, dbSNP, and other population, disease-specific, and sequence related databases, as well as a number of in silico predictive algorithms. The AGIS system is able to provide reliable and consistent genetic variant interpretation with the five-tier classification system (Pathogenic/Likely pathogenic/Benign/Likely Benign/Uncertain significance) and help human interpreters understand the clinical significance of genetic variants intuitively. Our implementation allows any potential new variants to be included and provides the most updated information to the scientific community. Currently it contains three major modules that can be used separately or in combination: Pharmacogenomics Module for guiding personalized disease treatment; Disease Risk Module for disease risk assessment; Genetic Disease Module for diagnosing hereditary diseases. The AGIS system can be easily used by researchers and clinicians, and will significantly enhance our understanding of the functional consequences of genetic variants in human diseases. The AGIS system is available at: <https://agis.admerahealth.com/>

## **PgmNr 1728: Assessment of statistical batch effect correction methods for different RNA sequencing data types.**

### **Authors:**

G. Cai <sup>1</sup>; S. Ni <sup>1</sup>; F. Xiao <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, SOUTH CAROLINA.; 2) Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SOUTH CAROLINA.

---

RNA-sequencing has rapidly become the mainstream method to profile and characterize transcriptome. However, batch effect makes downstream data analyses such as differential expression detection challenging and may affect the final conclusion. A handful models and methods have been developed and shown power in removing the batch effect in RNA-seq data. However, there is no clear guideline for method selection and application. Typically, two RNA-seq data formats are widely used for analysis, one is discrete RNA-seq raw read count and the other is transformed continuous FPKM/RPKM or CPM value. Researchers frequently ask which of these two formats should be used for batch effect correction. To rigorously answer this question and guide the method selection, we perform a systematic assessment of 13 existing batch effect removal methods for read count (edgeR, DESeq, limma-voom, SVASEq, RUVSeq, PCA) or continuous value (limma, LS, ComBat, SVA, RUVg, RUVs, RUVr, PCA). We evaluate their performance by the correction of between-batch variance, influence on within-batch variance, the AUC and type I error of downstream differential expression analysis. 6-site generated RNAseq data from the third phase of the MicroArray Quality Control (MAQC-iii) project were used. We also simulated outliers into the dataset to evaluate the robustness of batch effect removing. We found no one single method was superior than others in all scenarios. Based on the evaluation, we discuss the advantages, disadvantages, application scopes and matching data formats of these batch effect correction tools. We expect to provide a valuable application guide for researchers to make wise choices RNA-seq data batch effect correction accommodating to their specific experimental designs and analysis strategies.



## PgmNr 1729: IRNAA: An integrated shiny-based interface for differential expression analysis of RNA sequencing data.

**Authors:**

C.C. Huang; R. Azondekon

[View Session](#) [Add to Schedule](#)

**Affiliation:** Zilber School of Public Health, University of Wisconsin, Milwaukee, WI

---

**Background:** Given the significant progress made in the analysis of next-generation sequencing (NGS) data, numerous tools/packages have been developed for facilitating the downstream analysis of NGS data. Up to now, several R packages have been developed for differential gene expression analysis. Yet, those tools remain difficult to use for many researchers as they require extensive programming skills. In addition, they are difficult to integrate and to install without basic knowledge of command line scripting. More importantly, they lack the capacity to streamline RNA sequencing (RNA-seq) alignment, read quantification, and downstream analysis such as differential expression (DE) and pathway analyses.

**Implementation:** We present IRNAA, an integrated shiny-based interface for DE analysis of RNA-seq data, available for Ubuntu and related Linux distribution, as well as Microsoft Windows operating systems. IRNAA is built as a dashboard with multiple analysis components organized in boxes. In its GNU/Linux compatible implementation, IRNAA integrates the Salmon pseudo-aligner, FastQC and MultiQC, two bioinformatics tools for quality check assessment of RNA-seq reads, downstream DE analysis by DESeq2, limma-voom, edgeR, and pathway analysis by gProfile and WebGestalt. The application source code is freely available at <https://www.github.com/rosericazondekon/irnaa>.

**Conclusion:** IRNAA is very easy to install and does not require any programming experience or knowledge to preprocess RNA-seq FASTQ files and/or conduct DE analysis. IRNAA has the potential to facilitate the application of NGS methods in transcriptomics research.

## PgmNr 1730: A comprehensive evaluation of preprocessing methods for single-cell RNA sequencing data.

**Authors:**

S. Chu<sup>1,2</sup>; Q. Liu<sup>1,2</sup>; Y. Shyr<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America; 2) Center for Quantitative Sciences, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America

---

**Background:** Normalization and batch correction are critical steps in the preprocessing single cell RNA sequencing (scRNA-seq) data, which remove unwanted technical effect and systematic biases that unmask biological signal of interest. Although numerous computational methods have already been developed, there is no guidance for choosing the appropriate procedures in different scenarios.

**Method:** We benchmarked 23 procedures combining normalization and batch correction methods on both synthetic and real datasets in multiple scenarios, which considered relative magnitude of batch effects compared to biological effects and imbalanced cell compositions. The performance was evaluated by the capabilities to reduce batch effects, as well as to recover biological effects of interest, and the execution time.

**Result:** Batch effects can be removed at least by 60% on most procedures when they are not confounded with biological effects and they are not the major contributor to the variations. When batch effects confound with biological effects or their variations are larger than biological variations, the performance depends on their underlying mathematical models of batch correction (i.e., using linear model, nonlinear transformation, or linear-nonlinear combined adjustment).

**Conclusion:** The performance assessment of popular scRNA-seq preprocessing procedures can serve as a guideline to help users select the best method in the different scenarios.

## PgmNr 1731: Visualization and analysis of single-cell RNA-seq data by alternative clustering.

### Authors:

S. Liang <sup>1,2</sup>; Q. Liang <sup>3</sup>; R. Chen <sup>4</sup>; K. Chen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas.; 2) Department of Computer Science, Rice University, Houston, Texas.; 3) Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, Texas.; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.

---

Single-cell sequencing technology provides higher resolution for both cell differences and gene expression patterns, enabling unbiased global transcriptomic analysis. As the assay no longer targets a specific purpose, cells may be classified from various perspectives. For example, cells from macula and peripheral regions of retinae from multiple donors may be classified by their types or regions. Current visualization (e.g., tSNE, UMAP) and clustering (e.g., hierarchical clustering) methods only give an arbitrary single view of the data, which dissipates the power of scRNA-seq. Current batch effect removing methods may suppress one aspect to emphasize the others. However, they are widely criticized for exaggerating confidence in downstream analyses, because the differences between regions or donors are not technical noise and directly manipulating the expression data is problematic.

In this work, we focus only on discovering clusterings from different perspectives. We introduce “Varactor”, Visualization and Analysis of single-cell RNA-seq data by Alternative CLUSTERing. Varactor takes the expression matrix and the alleged nonpreferred clustering, which are either labeled beforehand, or previously discovered from the data. It redefines the pairwise distance of the cells, where the effect of the unwanted clustering is controlled, and use tSNE (which can be substituted with any distance-based clustering methods) on the new distances. It simply provides a new perspective for inspecting the similarity of the cells, instead of manipulating the expression data. The differential expression analysis may then be performed on the original data with the batches considered either strata or covariates.

On our retina dataset, we observe that Varactor successfully give tighter clusters for cell types validated by known marker genes while keeping subtle heterogeneity among donors as the detailed structure of each cluster. In contrast, even after performing Multi-CCA, Seurat still needs gene selection to give clear clusters, and cells from all donors are exaggeratedly mixed together. On a human brain single-cell dataset, we also observe that Varactor outperforms Multi-CCA implemented in Seurat in finding meaningful clusters, and the visualization result is comparable with the ones on a carefully selected gene sets provided in the original report. The new distance also adapts many clustering and trajectory inference methods. We expect Varactor to evoke more *ab initio* findings.

## PgmNr 1732: Surface protein imputation from single cell transcriptomes by deep neural networks.

### Authors:

Z. Zhou <sup>1,3</sup>; C. Ye <sup>2</sup>; J. Wang <sup>3</sup>; N.R. Zhang <sup>3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Genomics and Computational Biology Graduate Program, University of Pennsylvania, Philadelphia, PA., USA.; 2) School of Medicine, Tsinghua University, Beijing, China; 3) Department of Statistics, the Wharton School, University of Pennsylvania, PA., USA.

---

Recent technological advances allow the simultaneous profiling, across many cells in parallel, of multiple omics features in the same cell. In particular, high throughput quantification of the transcriptome and a selected panel of cell surface proteins in the same cell is now feasible through the REAP-seq and CITE-seq protocols. Yet, due to technological barriers and cost considerations, most single cell studies, including Human Cell Atlas (HCA) project, quantify the transcriptome only and do not have cell-matched measurements of relevant surface proteins that can serve as integral markers of cellular function and targets for therapeutic intervention. This motivates our inquiry of whether protein abundances in individual cells can be accurately imputed by the cell's transcriptome. Here we propose single cell Transcriptome to Protein prediction with deep neural network (cTP-net), a transfer learning framework that harnesses a multi-branched deep neural network to accurately predict single-cell surface protein relative abundances from scRNA-seq data.

We benchmark cTP-net's prediction on a diverse testbed of immune cell populations, and show that it achieves correlation  $> 0.9$  for the 10 proteins examined in a traditional holdout validation scheme. We also found cTP-net to have good generalization power, retaining high accuracy in tissues, cell types and technologies that differ from, but related to, the training data. cTP-net improves upon Seurat3, achieving higher accuracy in both the holdout scheme and generalization to unseen cell types. We generate predictions of 12 cell surface proteins for 530,000 cord blood and bone marrow mononuclear cells (CBMCs and BMMCs) that were deposited in the HCA portal and illustrate that multimodal data analysis by cTP-net can achieve a more detailed characterization of cellular phenotypes than transcriptome measurement alone. In addition, through interpolation analysis, we reveal that cTP-net utilizes a combination of cell type markers and genes related to RNA processing, protein localization and biosynthetic process to achieve multiscale imputation accuracy. With the accumulation of CITE-seq and REAP-seq data, cTP-net can be retrained to improve in accuracy and diversify in predictable protein targets. The results here underscore the need for more diverse multi-omic cell atlases and demonstrate how such resources can be used to enhance future studies. The cTP-net package is available at <https://github.com/zhouzilu/cTPnet>

# PgmNr 1733: NanoVar: A deep neural network for accurate detection of SNPs and Indels from long-read sequencing data.

## Authors:

M. Ahsan <sup>1</sup>; Q. Liu <sup>1</sup>; K. Wang <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Raymond G. Perelman Center for Cellular and Molecular Therapeutics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

**Purpose:** Long-read sequencing technologies, such as Oxford Nanopore and PacBio SMRT, produce substantially longer reads than Illumina short-read sequencing. Their unique advantage in detecting structural variants from sequence alignments enables examination of complex and repetitive regions of the human genome. However, their high base call error rates (typically >10%) complicates the detection of SNPs and Indels. Currently available software, such as GATK, show high variant calling accuracy for short-read data, but perform poorly on long-read data. But, recent developments in deep learning have shown promising results in improving the variant calling accuracy for long-read data.

**Methods:** We developed NanoVar, a deep convolutional neural network (CNN) that exploits error-prone Nanopore sequence reads for accurate variant calling. Using reads from a BAM file, it generates pileup images for candidate variant sites, which contain base information for a 33bp region around the site. The images consist of 5 feature channels; the first 4 record presence of the 4 bases and the 5th channel records base quality with reference match information. Pileup images are fed into a CNN model consisting of 3 convolutional and 3 fully connected layers, to predict alleles and genotype at the candidate sites. In a second neural network nearby variants sites are aggregated together to utilize haplotype information and re-score true variants and alleles. The model is subject to continuous adjustments and improvements.

**Results:** NanoVar was trained on both HG001 and HX1 whole-genome sequencing data and tested independently on chromosome 1 of each dataset. Ground truth calls for HG001 was sourced from Platinum Genomes Project, and high confidence variant calls for HX1 were generated from high-coverage Illumina data. NanoVar shows fast I/O run-time by leveraging parallel processing and outperforms variant callers designed for short-read sequencing data such as GATK. It also performs competitively against other variant callers designed for long-read data, including Longshot and deep neural network variant caller Clairvoyante.

**Conclusion:** Our preliminary analysis showed that NanoVar enables SNPs/Indel calling from long-read sequencing data, and can be easily adapted for a range of sequencing technologies. We conclude that long-read sequencing can generate accurate SNPs/Indels calls as a competitive alternative to short-read sequencing for genome analysis, at the expense of higher coverage.

## PgmNr 1734: An Omics Analysis, Search and Information System (OASIS) for mining association summary statistics from biobanks and knowledge portals.

### Authors:

J.A. Perry; B.D. Mitchell; J.R. O'Connell

[View Session](#) [Add to Schedule](#)

**Affiliation:** Program for Personalized and Genomic Medicine and Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD, 21201, USA

---

**Introduction:** Billions of associations between omic (e.g. genomic, transcriptomic, methylomic) variation and thousands of phenotypes have been computed and are quickly becoming publicly available. Knowledge Portals and Biobanks are already sharing summary statistics from their association analyses on public websites and a trend toward “open access” for data generated by government sponsored research is spreading worldwide. But there is a void in the arena of “scientist-friendly” tools for mining association results and transforming them into biological discovery. And while many annotation, visualization, and functional mapping resources are available, they are not automated or integrated into a meaningful process for discovery. Here we describe a tool for automating the mining process and integrating association results with functional annotation, data visualization and a variety of fine-mapping techniques.

**Methods:** An Omics Analysis, Search and Information System (OASIS) was constructed for mining data from the UK Biobank and from the data repositories used to populate the Type 2 Diabetes Knowledge Portal. OASIS is a web-based tool with robust, multi-faceted search capabilities. It provides on-demand linkage disequilibrium (LD) calculations, visualizations (e.g. boxplots, histograms, LocusZoom plots), and a broad spectrum of functional annotation. OASIS’s unique display of query results seamlessly integrates with online resources (e.g. dbSNP, genomAD, GTEx) and user-provided “known loci lists” to facilitate novel omic discoveries. User-generated summary statistics (i.e. GWAS or meta-analysis from the user’s lab) can also be loaded into OASIS for comparison with Biobank and Knowledge Portal findings.

**Results:** Association results from multiple repositories and user-generated analyses can now be easily mined and compared. Results for all variants within a locus can be viewed as a group and compared by LD values and functional annotation. Any set of variants can be selected for a customized display in the UCSC Genome Browser to understand which variants, of those with high LD, lie in regulatory regions (e.g. Dnase hypersensitivity sites, promoter/enhancer regions).

**Conclusion:** Transforming massive volumes of association results into “biological discovery” has been made dramatically easier. OASIS allows both analyst and non-analyst easy access to summary statistics from large repositories and uniquely integrates with multiple bioinformatic resources.

## PgmNr 1735: HiC-ACT: Improved chromatin interaction calling from Hi-C data via aggregated Cauchy test.

### Authors:

T. Lagler<sup>1</sup>; J. Rosen<sup>1</sup>; Y. Yang<sup>1</sup>; A. Abnoui<sup>2</sup>; M. Hu<sup>2</sup>; Y. Li<sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, The University of North Carolina at Chapel Hill; 2) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation; 3) Department of Genetics, The University of North Carolina at Chapel Hill; 4) Department of Computer Science, The University of North Carolina at Chapel Hill

---

Genome-wide chromatin conformation capture (3C) technologies such as Hi-C are commonly employed to study chromatin spatial organization. In particular, to identify statistically significant long-range chromatin interactions from Hi-C data, most existing methods such as Fit-Hi-C and HiCCUPS assume that all chromatin interactions are statistically independent. Such an independence assumption is reasonable for low resolution data (e.g., 40Kb bin), but is invalid for high resolution data (e.g., 5 or 10Kb bins) since spatial dependency of neighboring chromatin interactions is non-negligible in high depth data. When the independence assumption is violated,  $p$ -values corresponding to chromatin interactions may not be accurate. There is an urgent need to develop approaches that can not only capture spatial dependence, but are also computationally efficient and scalable. To address these challenges, we develop HiC-ACT, an aggregated Cauchy test (ACT) based approach, to improve the identification of significant chromatin interactions by post-processing calling results from methods relying on the independence assumption. HiC-ACT can be applied to results from any existing Hi-C interaction calling methods, enabling  $p$ -value local smoothing with negligible computational cost. To evaluate the performance of HiC-ACT, we re-analyzed publicly available deeply sequenced Hi-C datasets. Our results demonstrate that HiC-ACT improves sensitivity for the existing Hi-C interaction calling methods.

## PgmNr 1736: The *ped suite*: A collection of R packages for pedigree analysis.

### Authors:

M. Vigeland <sup>1</sup>; T. Egeland <sup>2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, University of Oslo, Oslo, Norway; 2) Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Aas, Norway.

---

We introduce a suite of packages for pedigree analysis in R. These packages, referred to as the *ped suite*, cover a wide range of topics and applications, but are built on the same fundamental data structures and coding principles. The packages are designed to work together, facilitating analysis of complex problems. A novel feature of the *ped suite* is the ability to handle *inbred founders*, leading to several improvements over existing methods.

The main packages in the *ped suite* are as follows:

**pedtools**: A comprehensive tool set for creating, manipulating and visualizing pedigrees and marker data. This is the core of the *ped suite*, imported by all the other packages.

#### Packages for marker-based analysis

**pedprobr**: Probability computations in pedigrees. The main content of this package is an implementation of the Elston-Stewart peeling algorithm for exact genotype probabilities in a given pedigree.

**forrel**: Forensic pedigree analysis and relatedness inference. This package also offers simulation of marker data, possibly conditional on known genotypes.

**pedbuildr**: Reconstructing small/medium-sized pedigrees from genotype data. The purpose of this package is to find the most likely pedigree connecting a collection of genotyped individuals.

**pedmut**: A framework for modeling mutations in pedigree computations.

#### Packages for marker-free analysis

**ibdsim2**: Simulation of the gene flow through pedigrees. In particular, such simulations can be used for studying distributions of chromosomal segments shared *identical-by-descent* by pedigree members.

**ribd**: Computation of pedigree-based relatedness coefficients, including kinship coefficients and Jacquard's identity coefficients. Both autosomal and X-linked coefficients are computed, in pedigrees of any complexity (including selfing). More specialized coefficients include generalized kinship coefficients and two-locus coefficients.



## PgmNr 1737: SAIGEgds: An efficient statistical tool for large-scale PheWAS with mixed models.

**Authors:**

Zheng; J.W. Davis

[View Session](#) [Add to Schedule](#)**Affiliation:** Genomics Research Center, AbbVie Inc, North Chicago, IL

---

Phenome-wide association study (PheWAS) is known to be a powerful tool in discovery and replication of genetic association studies. The recent development in the UK Biobank resource with deep genomic and phenotyping data has provided unparalleled research opportunities. To reduce the computational complexity and cost of PheWAS, the SAIGE (scalable and accurate implementation of generalized mixed model) method was proposed recently. However, it is still computationally challenging to analyze the associations of thousands of phenotypes with whole-genome imputed and whole-exome variant data, especially for disease diagnoses using the ICD-10 codes.

Here we develop a new high-performance statistical package (SAIGEgds) for large-scale PheWAS using mixed models. In this package, we implement the SAIGE method with optimized C++ codes and fully take advantage of sparse structure of genotype dosages. SAIGEgds supports efficient genomic data structure (GDS) files including both integer genotypes and numeric dosages. Benchmarks using the UKBiobank White British genotype data (N=430,235) with coronary heart disease and simulated cases, show that SAIGEgds is 5 to 6 times faster than the SAIGE R package in the steps of fitting null models and p-value calculations. When used in conjunction with high-performance computing (HPC) clusters and/or cloud resources, SAIGEgds provides an efficient analysis pipeline for biobank-scale PheWAS.

## PgmNr 1738: Coexpression is not a network.

**Authors:**M. Farahbod<sup>1,2,4</sup>; P. Pavlidis<sup>1,2,3</sup>[View Session](#) [Add to Schedule](#)**Affiliations:**

1) The University of British Columbia, Vancouver, British Columbia, Canada; 2) Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia, Canada; 3) The UBC Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada; 4) Graduate program in Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada

---

Coexpression analysis is commonly used as an aid to interpreting transcriptomic studies of disease. A motivation behind such analyses is the idea that coexpression is due to coregulation and/or that it reflects direct relationships between gene products. These assumptions are reinforced by the common representation of coexpression as a network. Here we show through computational analysis that much of observed coexpression identified in bulk brain tissue studies is due to the effects of cellular composition variation between samples. Therefore, interpretations of coexpression as a network representing physical or regulatory interactions is unlikely to reflect underlying biology. We focused on analysis of human neocortex, which has been the topic of many studies using coexpression to probe numerous psychiatric and neurological conditions. We first show that for most genes, differences in expression between brain cell types explain a large fraction of the variance of observed expression in bulk tissue (median adjusted  $R^2 = 0.34$ ). We then show that the similarity of expression patterns across cell types predicts the tendency for genes to be coexpressed in bulk tissue. These observations explain functional enrichment analysis of gene clusters in bulk brain tissue as simply reflecting cell-type-specific gene expression patterns. We also show that robust coexpression patterns observed in single nucleus RNA-seq data are highly distorted in bulk tissue, due to the effect of cellular composition. The impact of these results is several-fold. First, it accounts for the most prominent reported characteristics of coexpression as arising from differences in gene expression between cell types, not regulation per se. Second, it implies that the ability to extract regulatory information from coexpression is confounded by cellular composition effects, and therefore dramatically limited, helping explain why many algorithms designed to do so have high error rates. Third, it shows that interpretations of coexpression as networks can be highly misleading, and in particular the concept of “hubs” in coexpression does not parallel features of any physical network. While our analysis is of brain tissue, the same phenomenon is likely to explain coexpression in other tissues. Overall our findings urge a rethink of the value and interpretation of coexpression as a routine part of transcriptome analysis workflows.

# PgmNr 1739: Ranking transcription factor binding motifs in ChIP-seq data without p-values.

## Authors:

L. Sun <sup>1,2</sup>; S. Mastromatteo <sup>3</sup>; C. Viner <sup>4,6</sup>; D. Denisko <sup>5,6</sup>; L. Zhang <sup>1</sup>; J. Negrea <sup>1</sup>; Y. Tang <sup>1</sup>; M.M. Hoffman <sup>4,5,6,7</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 2) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Canada; 3) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 4) Department of Computer Science, University of Toronto, Toronto, Canada; 5) Department of Medical Biophysics, University of Toronto, Toronto, Canada; 6) Princess Margaret Cancer Centre, Toronto, Canada; 7) Vector Institute, Toronto, Canada

---

**Introduction:** Transcription factor binding *in vivo* is mapped by techniques like chromatin immunoprecipitation-sequencing (ChIP-seq) and cleavage under targets and release using nuclease (CUT&RUN). Sequence motifs found more frequently near the center of these methods' peaks often represent direct DNA binding by the targeted transcription factor.

The most common approach to quantitatively rank motifs enriched in peak centers uses p-values. For each candidate motif, this approach examines  $n$  peaks and identifies the number of peaks  $X$  where the best motif match is in a central window. It then calculates a  $p$ -value using a binomial distribution for  $X$  with parameter  $\theta$  (the probability for a sequence being aligned to a center window), under the null hypothesis of no central enrichment. Finally, it ranks motifs by their (one-sided) binomial  $p$ -values. Unfortunately, the use of  $p$ -value here is problematic because the  $p$ -value magnitude is dominated by severe heterogeneity in the number of peaks between motifs.

**Methods:** We created a new method for ranking central enrichment of sequence motifs that measures effect size instead of  $p$ -value significance. Our method ranks motif central enrichment by the lower bound of the confidence interval for the parameter  $\theta$ , after adjusting for multiple testing.

We compared the rankings of our method with three different classes of alternative methods in both simulated and published ChIP-seq data. First, we compared with the binomial  $p$ -value approach above. Second, we compared with Bayesian credible regions using either conjugate prior or AshR (Stephens 2017), an empirical Bayes method. Third, we compared with Cliff's  $\delta$  (Cliff 1993), which quantifies central enrichment of one distribution over another.

**Results:** Extensive simulation studies showed that our method outperforms the binomial  $p$ -value approach when sample sizes vary independently of, or are inversely proportional to, effect sizes. All approaches performed similarly when sample sizes stay constant between tests or vary proportionally to effect sizes.

Using published ChIP-seq data for GATA3, our method ranked the known targeted motif highest, even

among similar motifs for other GATA factors. Our method also ranked the known motif highest for SRF. In both cases, the  $p$ -value approach failed to assign the top rank to the known motif. While other effect-size methods performed similarly, we recommend the confidence interval method for its simplicity and ease of implementation.

## PgmNr 1740: A method to discover combinatorial variants within transcription factors that are associated with gene expression variability.

**Authors:**

Y. Tang; A. Gottlieb

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Precision Health, University of Texas Health Science Center at Houston, 7000 Fannin st, Houston, TX, 77030.

---

**Background:** Gene expression plays a key role in determining cellular states. Estimating the genetic component underlying gene expression can thus help our understanding of disease etiology or predict disease risk.

Recent approaches to estimate the genetic component of gene expression involve polygenic models that include combination of variants with large cumulative effects called “transcriptome imputation”. However, transcriptome imputation have thus far manages to explain only a small portion of expression variability for most genes. As it has been limited to variants in the *cis* regions of the gene, this suggests that other genetic components may be involved.

**Results:** We hypothesize that variants within transcription factors (TFs) may also contribute to variability in gene expression. We developed models to identify combinations of these variants. We introduce three models, corresponding to possible mechanisms by which variants in TFs can affect gene expression: effect on TF expression, on TF binding affinity or on both. We applied our TF models to four tissues – skeletal muscle, subcutaneous adipose, skin and whole blood and show that our TF models are robust across tissue types. We identified combinations of variants in TFs that explain the expression of 80 genes better than *cis* models and in 27 of the genes, the TF models include variants with higher association with expression than *cis* variants. These variants were undetected individually as trans-eQTLs.

Among our discoveries, we highlight variants discovered in TF models applied to adipose and muscle tissues associated with Type 2 Diabetes, including the genes *IRS1* and *ABCC8* and their corresponding TFs, *TCF3* and *REST*. Similarly, we highlight variants in TFs associated with basal cell carcinoma discovered in skin tissue.

**Conclusions:** Our approach is useful for identifying regulation mediated by combinatorial effect of genetic variations within TFs and can improve our understanding of genetic component of gene expression and ultimately the connection between genotype and phenotype.

## PgmNr 1741: Machine learning-based prediction of gene expression modifiers prioritizes non-coding variants causal for rare disease.

### Authors:

Q. Wang <sup>1,2</sup>; D. Kelley <sup>3</sup>; R. Tewhey <sup>4</sup>; Y. Reshef <sup>5</sup>; K. Karczewski <sup>1</sup>; L. Francioli <sup>1</sup>; A. O'Donnell-Luria <sup>1</sup>; D. MacArthur <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Institute of MIT and Harvard, Cambridge, Massachusetts.; 2) Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, Massachusetts.; 3) Calico Labs, San Francisco, California.; 4) The Jackson Laboratory, Bar Harbor, Maine; 5) Department of Computer Science, Harvard University, Cambridge, Massachusetts.

---

Whole genome sequencing (WGS) has the potential to elucidate non-coding regulatory mutations responsible for severe phenotypes. However, identifying non-coding causal variants from WGS data is still challenging. The availability of large-scale functional genomics resources, genomic constraint metrics, and direct measures of the effect of genetic variation on RNA expression provides an opportunity to improve the prioritization of such variants.

Here we propose to integrate functional genomics data and sequence-based regulatory activity prediction into a single score by training a predictor of the effects of SNVs on the expression of proximal genes. Our predictor combines publicly available epigenetic and transcriptional activity features (ENCODE/ROADMAP/FANTOM5) with sequence-specific features from the neural network-based method Basenji to output a tissue-specific Expression Modifier Score (EMS). We trained our predictor on fine-mapped expression quantitative trait loci (eQTLs), and used known literature-derived non-coding causal variants as test data.

When using eQTL data from Geuvadis, our predictor distinguished between variants with posterior probability (p.p.) of having a causal effect on gene expression  $>0.9$  ( $n=198$ ) and variants with  $p.p.<0.2$  ( $n=991$ ), with 76.8% true positive rate at 20.0% false positive rate, higher than using Basenji features or epigenetic features alone (70.0%, 68.4%).

We then hypothesized that a large fraction of non-coding variants causal for rare disease act by modifying the expression of proximal genes, and tested our predictor's ability to identify a set of such variants curated from the literature ( $n=209$ ). We defined two sets of negative control variants based on allele frequency, and tested the predictor's ability to prioritize causal variants over the control sets. The area under curve (AUC) ranged from 0.68 to 0.72, slightly higher when the rare control variants are used, suggesting that EMS could be useful in rare variant prioritization, despite relying on a training data set of common eQTLs.

Finally, we will present our ongoing analysis of causal variants in GTEx dataset, quantification of constraint in non-coding regions in large population genomics database (gnomAD), and show our framework for integrating such large-scale data to enhance the power of causal variant discovery.

## PgmNr 1742: PopSTR2 enables population-scale and clinical genotyping of microsatellites.

### Authors:

S. Kristmundsdóttir <sup>1,2</sup>; H. Eggertsson <sup>1,3</sup>; G. Árnadóttir <sup>1</sup>; B. Halldórsson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Statistics, deCODE genetics, Reykjavik, Iceland; 2) School of Science and Engineering, Reykjavík University, Reykjavik, Iceland; 3) School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland

---

Microsatellites, also known as short tandem repeats (STRs), are tracts of repetitive DNA sequences containing motifs ranging from one to six bases. Microsatellites are one of the most abundant type of variation in the human genome, after single nucleotide polymorphisms (SNPs) and Indels. However, microsatellite variations are rarely considered in whole-genome sequencing studies and clinical sequencing, in large due to a lack of tools capable of analyzing them.

PopSTR2 is an update and extension of our previous work 'PopSTR a population based microsatellite genotyper'. Given a list of samples and microsatellite markers it genotypes each sample at each marker. Microsatellites are highly polymorphic and often show a high somatic polymorphism rate. PopSTR2 automatically adjusts parameters to account for variable sequencing accuracy and has user specified parameters related to the sequencing read length. We supply a kernel to quickly estimate a sample specific slippage rate, making the genotyping more sensitive to each sample's properties. For clinical sequencing purposes, a panel of known pathogenic repeat expansions is provided along with a script to quickly check all markers in the panel for reads supporting an expansion.

Like its predecessor, PopSTR2 allows for the joint genotyping of samples at a population scale. To increase the power of downstream association we also provide a binning method to make markers more amenable to analysis within standard association pipelines.

# PgmNr 1743: Quantifying the quality of small INDELs called in sequencing analysis.

## Authors:

H. Dai; Y. Guan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC.

Current quality score of INDEL call generated by various variant calling bioinformatics tools were not well calibrated. Here we present a method to quantify the variant calling quality for small INDELs using calibrated posterior probability.

Our method is based on Nubeam, which can quantify the genetic difference between collections of reads. Nubeam represents nucleotides by matrices, transforms a read into a product of matrices, and based on which assigns numbers to reads. A collection of reads becomes a collection of numbers that form an empirical distribution. The genetic difference between collections of reads is quantified by the distance between empirical distributions.

When an INDEL is called, we obtain three collections of sequencing reads: reads used to call the INDEL (denote  $R_1$ ), reads simulated from alternative allele (denote  $R_2$ ), and reads simulated from reference allele ( $R_3$ ). We then calculate the Nubeam distance between  $R_1$  and  $R_2$  (denote  $d_{12}$ ) and that between  $R_1$  and  $R_3$  (denote  $d_{13}$ ). We obtain a likelihood ratio as  $L(d_{12}, d_{13}) = \exp(-\beta d_{12}) / \exp(-\beta d_{13})$  and treat it as a naïve Bayes factor (nBF), where  $\beta$  is a tuning parameter. The population frequency of the called INDEL is taken as the prior odds. Multiplying the nBF and prior odds we obtain the posterior odds, from which we calculate the posterior probability (which is a function of  $\beta$ ). We then tune the parameter  $\beta$  to calibrate the posterior probability.

We first used simulations to demonstrate that the nBF can predict if a collection of reads originates from alternative allele or reference allele. We simulated INDELs at random positions across human genome; for each INDEL, we simulated two collections of  $R_1$  reads, one originates from heterozygous alleles and the other from homozygous reference alleles. When  $\beta$  takes one, we predicted that the  $R_1$  originates from heterozygous alleles if  $\text{nBF} > 1$  and from homozygous reference alleles if otherwise. The accuracy was 100% in 1,000 simulations even when INDEL size was one. We then randomly selected 3,000 heterozygous INDELs from Genome in a Bottle (GIAB) benchmarking variant calls. Using reads from Illumina 2×148bp 300× WGS dataset, an nBF cutoff of one classified 97% of them as true variants. The preliminary results using nBF show that our method is promising in quantifying the INDEL calling quality by calibrated posterior probability.



## PgmNr 1744: GraphTyper2 enables population-scale genotyping of structural variation using pangenome graphs.

### Authors:

H.P. Eggertsson <sup>1,2</sup>; S. Kristmundsdottir <sup>1,3</sup>; D. Beyter <sup>1</sup>; H. Jonsson <sup>1</sup>; A. Skuladottir <sup>1</sup>; M.T. Hardarson <sup>1</sup>; D.F. Gudbjartsson <sup>1,2</sup>; K. Stefansson <sup>1,4</sup>; B.V. Halldorsson <sup>1,3</sup>; P. Melsted <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) deCODE Genetics/AMGEN, Reykjavík, Iceland; 2) School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland; 3) School of Science and Engineering, Reykjavik University, Reykjavik, Iceland; 4) Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland

---

Analysis of sequence diversity in the human genome is fundamental for genetic studies. Structural variants (SVs) are frequently omitted in sequence analysis studies although each has relatively great impact on the genome. We present GraphTyper2, which uses pangenomes to genotype SVs along with small variants using short-reads. We use public datasets to show that our SV genotyping is sensitive and variant segregation in families demonstrates the accuracy of our approach. We generated long-read sequences from 41 Icelanders to assess the quality our short-read SV calls. Using the long-reads, we validated 67.7% of our 8,568 high-confidence SVs on average per genome. We show that GraphTyper2 can simultaneously genotype tens of thousands of whole-genomes by characterizing 60 million small variants and half a million SVs in 49,962 Icelanders, thereof 80 thousand SVs with high-confidence. GraphTyper2 is a valuable tool for characterizing sequence variation in large-scale sequencing studies.

## PgmNr 1745: Evaluating variant calling best practices in a non-European population.

### Authors:

E.K.. Howell <sup>1,2</sup>; T.N. Phung <sup>1,2</sup>; A.M. Taravella <sup>1,2</sup>; M.A. Wilson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Life Sciences, Arizona State University, Tempe, AZ; 2) Center for Evolution and Medicine, Arizona State University, Tempe AZ

---

Analyzing human DNA sequence data allows researchers to identify variants associated with disease, reconstruct demographic histories of human populations, and further understand the structure and function of the genome. Currently, different strategies exist to genotype variants, remove false positives, and identify false negatives within sequence data. However, many current methods have been optimized for studying European populations. With the number of sequences analyzed from non-European populations growing, it is increasingly important to re-evaluate the “best practices” that guide the processing of human DNA sequence data and to assess how different methods may impact downstream analyses. Here, we evaluate the concordance and accuracy of different variant calling strategies such as joint and single-sample genotyping as well as different filtering procedures such as Variant Quality Score Recalibration (VQSR) and hard-filters within whole genome and whole exome sequences from individuals from Northwest Kenya. Since human males typically have one X chromosome and one Y chromosome and since current methods are optimized for genotyping the autosomes, we also evaluate how to best genotype the sex chromosomes. In addition, we compare the ability of joint and single-sample genotyping to accurately identify rare variants which may be associated with disease but are difficult to distinguish from sequencing errors due to their low frequency. We generated variant call sets for the autosomes, mitochondrial chromosome, and sex chromosomes using joint and single-sample genotyping and filtered using both VQSR and hard-filters. Genotype concordance between call sets was assessed, and the accuracy of variant calls was measured by comparing call sets to highly confident SNP array genotype data. We also evaluated the concordance between singleton sites found on the array and sites found in the joint and single genotyped call sets. Since both whole genome and exome sequencing are commonly used to obtain measures such as nucleotide diversity and the site frequency spectrum, we also evaluated how different genotyping and filtering techniques impacted these summary statistics. Characterizing the concordance and accuracy between these variant calling and filtering methods and comparing the differences in haploid and diploid genotyping of the Y chromosome will establish a more effective protocol for variant discovery in non-European populations and within the sex chromosomes.

## PgmNr 1746: Performance evaluation of different approaches for insertion and deletion variation detection from next generation sequencing data.

**Authors:**

S. Khan; N. Wang; L.L. Elo

[View Session](#) [Add to Schedule](#)

**Affiliation:** Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland

---

Insertion and deletion variation have been shown to associate with many diseases including autoimmune diseases and cancers. Insertions and deletions (indels) are the most common type of structural variation in the human genome and they can be identified using high throughput next generation sequencing techniques. While many computational tools for indel detection from sequencing data are available, the detection accuracy is still considerably lower for indel variation than for single nucleotide variation. The reliable indel calling is desirable in any application field, especially in genetic diagnosis. In order to dissect potential underlying reasons for the shortcomings in detection of indel variants, we studied the impact of the methods as well as the technical and biological features of the sequencing data. We evaluated seven widely used open source indel detection tools (DELLY, Pindel, FermiKit, Platypus, GATK Haplotype caller, VarScan, and Strelka2) with the underlying methodology spanning from assembly-based approaches to machine learning approaches. The impact of the sequencing data on indel detection was evaluated by using real world benchmarking data with varying sequencing coverages spanning from 60X to 5X coverage. Finally, the sequence context of the false positive indel detections was carefully studied. Our results demonstrated that the choice of the algorithm has an impact on the precision and recall of the indel calling, but no systematic bias related to the underlying algorithms was found. The indel detection performance benefited from the higher sequencing coverage compared to the 5X coverage. However, there was only a slight improvement in indel calling between 30X and 60X. The impact of the sequence context had major impact on the indel calling results with the presence of a repeat region being the most problematic sequence context for reliable indel calling. The results provide insights into the key factors influencing indel calling reliability and may be useful in designing of improved solutions for indel detection.

## PgmNr 1747: Pore-C: A method for genome-wide, multi-contact chromosome conformation capture.

### Authors:

S. Juul <sup>1</sup>; M. Pendleton <sup>1</sup>; N. Ulahannan <sup>2,3</sup>; X. Dai <sup>1</sup>; D.J. Turner <sup>1</sup>; E.D. Harrington <sup>1</sup>; M. Imielinski <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Applications, Oxford Nanopore Technologies, New York, NY; 2) Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY; 3) New York Genome Center, New York, NY

---

The DNA within the nucleus of an interphase cell is organised into a complex hierarchy of folds and loops known as the 3D Genome. The development of various chromatin conformation capture methods has enabled the detection of the structures that define each level of this hierarchy e.g. chromosome territories, A/B compartments, topologically associated domains (TADs) and promoter-enhancer loops. This in turn has facilitated functional studies which have uncovered some of the mechanisms behind the formation and maintenance of these structures, as well as their effect on gene expression. However, most of these studies rely on methods that can only capture interactions between two points on the genome, and thus lack the ability to resolve higher-order interactions. We will share our progress on Pore-C, a method to generate genome-wide, multi-contact chromatin conformation maps. Using this technique we were able to resolve misassemblies and join contigs in a *de novo* assembly of the genome of cell line GM12878. This way, we generated an exceptionally contiguous assembly with an N50 of 36 Mb and a longest contig of 129 Mb, which is ~89% of the entire chromosome 8 including the centromere. Furthermore, we will show how we used Pore-C to detect rearrangements and copy number changes and resolve complex structural variants in the human breast cancer cell line HCC1954.

## **PgmNr 1748: lncRNA-peptidome analysis through proteogenomic-approach: Quantifying tissue, cell-line, and plasma lncRNA-peptidome to discover peptide-based cancer biomarkers.**

### **Authors:**

S. Chakraborty <sup>1</sup>; G. Andrieux <sup>2,3</sup>; M. Hasan <sup>1</sup>; M. Ahmed <sup>4</sup>; M. Boerries <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Biochemistry and Molecular Biology, University of Dhaka, Bangladesh, Dhaka, Bangladesh; 2) Institute of Molecular Medicine and Cell Research, Faculty of Medicine, University of Freiburg, Freiburg, Germany; 3) German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Freiburg, Freiburg, Germany; 4) Princess Margaret Cancer Centre/University Health Network, Toronto, Ontario, Canada

---

The debate on the translational capacity of lncRNAs to guide the synthesis of polypeptides was ignited with the identification of evolutionary conserved ribosome-binding property of lncRNAs. Proteome-centric studies, although have identified numerous lncRNA-encoded polypeptides, lack differential expression analysis of lncRNA-peptidome across primary tissues, cell lines and cancer states. We developed an integrated proteogenomic pipeline to identify and quantify the polypeptides as the translated products of the lncRNA transcripts in the human genome. In short, we first predicted the hypothetical polypeptide sequence from each lncRNA transcript annotated in the human genome. We then exhaustively searched for these hypothetical polypeptides in the LC-MS/MS data to identify polypeptides that are actually translated in human primary tissues and cell lines. With the aid of the computational-proteogenomic workflow we re-processed publicly available LC-MS/MS data, which facilitated the identification of tissue-specific and universally expressed (UExp) lncRNA-polypeptides across 14 primary human tissues and 11 cell lines. The utility of lncRNA-peptidome as cancer-biomarkers was investigated by re-processing LC-MS/MS data from 92 colon-adenocarcinoma (COAD) and 30 normal colon-epithelium tissues. Intriguingly, a significant upregulation of five lncRNA UExp-polypeptides in COAD tissues was observed. Furthermore, clustering of the UExp-polypeptides led to the classification of COAD patients that coincided with the clinical stratification, underlining the prognostic potential of the UExp-polypeptides. Lastly, we identified differential abundance of the UExp-polypeptides in the plasma of prostate-cancer patients highlighting their potential as plasma-biomarker. The analysis of lncRNA-peptidome may pave the way to identify effective tissue/plasma biomarkers for different cancer types.

## PgmNr 1749: Genomic landscape of conjunctival melanoma.

### Authors:

K. Cisarova <sup>1</sup>; I. El Zaoui <sup>1</sup>; R. Pescini-Gobert <sup>1</sup>; B. Royer-Bertrand <sup>1</sup>; L. Zografos <sup>2</sup>; A. Schalenbourg <sup>2</sup>; N. Riggi <sup>3</sup>; A. Moulin <sup>2</sup>; C. Rivolta <sup>1,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; 2) Jules-Gonin Eye Hospital, Department of Ophthalmology, Fondation Asile des Aveugles, UNIL, Lausanne, Switzerland; 3) Experimental Pathology, Institute of Pathology, Lausanne University Hospital, Lausanne, Switzerland; 4) Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom; 5) Institute of Molecular and Clinical Ophthalmology Basel (IOB), CH-4031 Basel, Switzerland; 6) University of Basel, 4001 Basel, Switzerland

Conjunctival melanoma (CJM) is extremely rare but potentially lethal and highly recurrent form of cancer of the eye that, similarly to cutaneous melanoma (CM), originates from melanocytes. Contrarily to CM, however, CJM is relatively poorly characterized from a genomic point of view.

A better understanding of genetics of CJM would be beneficial for the improvement of its classification and especially of available treatment options.

To fill this knowledge gap and to gain insight into the genomic nature of CJM, we performed whole-exome (WES) or whole-genome sequencing (WGS) and RNA sequencing (RNA-seq) of tumor-normal tissue pairs in 14 affected individuals.

Our results show that, similarly to CM, CJM is also characterized by a very high somatic mutation load, composed of approximately 500 mutations in exonic regions. This, together with the presence of a clear UV light-induced signature, are signs of the role of sunlight in CJM etiology.

In addition, our analysis showed that the genomic classification of cutaneous melanoma proposed by TCGA seems to be well-applicable to conjunctival melanoma. Specifically, we observe four subclasses, defined on the basis of the most frequently mutated genes: *BRAF*, *NF1*, *RAS* and triple wild-type.

In line with these results, transcriptomic analyses show the presence of immune and keratin signature, again similarly to what has been reported in CM.

Alltogether, our results provide the first unbiased and complete genomic and transcriptomic classification of CJM.

## PgmNr 1750: Mega-analysis of transcriptomic field effect in colorectal cancer reveals novel features.

### Authors:

C. Dampier <sup>1</sup>; M. Devall <sup>2</sup>; V. Moreno <sup>3</sup>; G. Casey <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Surgery, University of Virginia, Charlottesville, VA; 2) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 3) Catalan Institute of Oncology, Bellvitge Biomedical Research Institute, and CIBERESP, L'Hospitalet de Llobregat, Barcelona, Spain

---

**Background:** Colorectal cancer is a relatively common malignancy that is curable when detected early but often fatal when diagnosis is delayed. A field effect of cancer in the colon has been reported and could have implications for diagnosis and therapy, but descriptions to date have been limited by small sample sizes and inappropriate tissue comparisons. To address these limitations, we performed a joint analysis of pooled transcriptomic data from all publicly available RNA-seq datasets of healthy and tumor tissue as well as a previously unpublished cohort of our own.

**Methods:** RNA-seq datasets from studies including healthy (HLT), tumor-adjacent but pathologically normal (NAT), and tumor (CRC) samples published in the Genomic Data Commons and Sequence Read Archive as well as our own cohort were systematically screened for inclusion. Samples of bulk sequencing with a depth of at least 10 million quantifiable paired-end reads from flash frozen healthy or tumor colorectal tissue obtained from surgical specimens and biopsies likely to contain a substantial mucosal component were eligible. After converting all available data to FASTQ format, we quantified transcripts using Salmon in GC bias mode for all samples. We aggregated transcripts at the gene level with the tximport R package. We used the SVA package to identify latent batch effects and performed differential expression comparisons across tissue types with DESeq2 using five surrogate variables to control for latent factors.

**Results:** After screening for eligibility and filtering out duplicates, we retained for analysis 924 samples from 11 studies, including 462 HLT samples, 94 NAT samples, and 368 CRC samples. In comparisons across tissue types in which we tested for absolute fold-change of 2 or more at FDR 5%, we found 1,701 differentially expressed genes (DEG) between HLT and NAT, 2,929 DEG between NAT and CRC, and 5,974 DEG between HLT and CRC. Expression of a subset of 43 DEG was found to increase consistently from HLT to NAT and from NAT to CRC, while expression of 18 DEG displayed the opposite pattern. A subset of 128 DEG were overexpressed in NAT compared with both HLT and CRC.

**Conclusion:** In this mega-analysis of the colorectal transcriptome, we leveraged increased power for discovery despite careful tissue selection to identify novel features of the field effect in colorectal cancer. These features may be useful for diagnosis and treatment of colorectal cancer.

## **PgmNr 1751: The enhancer connectome of ovarian cancer refines GWAS loci and identifies target genes of high grade serous ovarian cancer.**

### **Authors:**

B. Davis <sup>1</sup>; C. Giambartolomei <sup>2</sup>; M. Freedman <sup>3</sup>; B. Pasaniuc <sup>2</sup>; S. Gayther <sup>1</sup>; J. Plummer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Bioinformatics and Functional Genomics, Cedars-Sinai Medical Center, LOS ANGELES, California.; 2) Center for Health Sciences, University of California Los Angeles, Los Angeles, CA; 3) Dana-Farber Cancer Institute, Boston, MA

---

Genome-wide association studies (GWAS) have led to the identification of >30 risk loci for ovarian cancer. GWAS SNPs are annotated to the nearest gene with most fine mapped SNPs lying in non-coding genomic regions suggestive of their distant role on target gene expression. These disease-susceptibility variants frequently lie in cell-type-specific enhancer elements. To identify, interpret, and prioritize such risk variants, and thus identify their target gene, we use H3K27ac HiChIP, an efficient novel protein-mediated chromatin-conformation assay to obtain high-resolution interaction maps. By conducting HiChIP in fallopian tube precursor cells and ovarian cancer cell lines, we can confidently assign enhancer-promoter contacts in a cell-type specific manner and reveal principles that govern these interactions from normal to a cancer state. By establishing this ovarian enhancer interactome, we identify key regulatory elements involved in cancer risk and progression. These connectome maps in normal precursor cells and ovarian cancer cells were combined with summary stats from the Oncoarray to identify enhancer-promoter loops that overlap at GWAS signals (95% credible set). We observe that GWAS variants from ovarian cancer are enriched within the HiChIP interactions consistent within a role of enhancer-promoter interactions in the regulation of gene expression. We further integrate these results with GTEX, eQTL data, and TWAS data to explore the credible SNP-contact genes. This study shows that HiChIP interaction data can be leveraged with GWAS and eQTL to prioritize target genes for ovarian cancer which are supported by physical interactions between regulatory elements (enhancers) and their target genes.



## PgmNr 1752: Gene expression profiling from cervix cancer patients living in Senegal.

### Authors:

G. Diop<sup>1,2</sup>; M. Ndiaye<sup>1,2</sup>; S. Leclerc<sup>4</sup>; C. Derbois<sup>3</sup>; C. Coulonges<sup>4</sup>; A. Dieye<sup>1</sup>; A. Dem<sup>5</sup>; J.F. Zagury<sup>4</sup>; J.F. Deleuze<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Dakar, UCAD, Dakar, Senegal. BP:55, Dakar Senegal; 2) Institute Pasteur Dakar, Senegal. BP:220, Dakar-Senegal; 3) Centre National de Recherche en Génétique Humaine, CNRGH, 2, Rue Gaston Cremieux 91057 Evry France.; 4) Conservatoire National des Arts et Metiers, Université des Metiers, 292 Rue Saint-Martin, 75003 Paris, France; 5) Hopital Aristide LEDANTEC Avenue Pasteur Dakar-Senegal

---

Cervical cancer (CC) is a multifactorial diseases and the main aetiologic agent is the Human Papillomavirus (HPV). Despite the implementation of cervical uterine Pap smear (HPV) screening and vaccination to treat precancerous lesions , cervical cancer is still a public health problem.

The emergence of novel cervical cancer biomarkers is promising enough to reduce the cost of prevention and improve the specific detection of high-grade cervical lesions and early-stage cervical cancer, implying an improvement in the efficacy of cervical cancer treatment management.

The objective of this study will be to identify gene expression profiles and to investigate potential biomarkers through the analysis of significantly altered signaling pathways from patients with cervical cancer, cervical intraepithelial neoplasia (CIN 1) and healthy control subjects (CTR) using Illumina technology.

The Illumina BeadChips were used for a complete genome-wide transcript profiling of whole blood from 31 CC patients, 27 CIN and 29 CTR. Differentially expressed genes (DEG) were identified by comparing the mean normalized expression for each gene in the different sample groups. A hierarchical clustering was performed on the 3,435 significant genes and DEG at significant level between patients and the clusters generated were analyzed using Gene Ontology (GO) in String protein database to identify biological processes.

Few differences were observed between CIN patients and CTR with only 129 upregulated and 130 downregulated genes in CIN patients. In contrast, 1,569 genes were overexpressed in CC patients including 1,404 and 1,303 genes respectively from the comparisons with CIN patients and CTR. We identified in CC patients compared to CTR a high expression of a spectrum of genes involved in the immunity of CC patients and a weak expression of genes related to metabolism. Microarray data were validated by RT-qPCR in a set of seven genes showing a high degree of correlation.

Conclusion: This enrichment of genes involved in immunity overexpressed in CC patients could be related to their particular immune state regarding cervical cancer and the observation of genes related to metabolism under-expressed in CC patients could reflect the rewiring of the metabolism processes in cancer. Our study highlighted several new genes that could contribute in the identification of innovative clinical biomarkers for diagnostic procedures and therapeutic interventions.

## PgmNr 1753: Examination of whole-genome sequencing using SeqPlus sequencing methodology in esophageal carcinoma formalin-fixed, paraffin-embedded samples stored for more than a decade.

### Authors:

A.M. Goldstein<sup>1</sup>; S.T. Bailey<sup>2</sup>; B. Hicks<sup>1,3</sup>; B. Zhu<sup>1,3</sup>; N. Hu<sup>1</sup>; M. Wang<sup>1,3</sup>; C. Graham<sup>1</sup>; P.R. Taylor<sup>1</sup>; K. Jones<sup>1,3</sup>; J.R. Gulcher<sup>2</sup>; H. Sun<sup>2</sup>; J. Lund<sup>2</sup>; K. Oliner<sup>2</sup>; R.T. Williams<sup>2</sup>; S.J. Chanock<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD, USA; 2) WuXi NextCODE, Cambridge, MA, USA; 3) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

---

Collection of fresh-frozen tumor samples is often challenging or infeasible. Thus, many studies rely on formalin-fixed, paraffin-embedded (FFPE) tumor tissue samples. Although whole-genome sequencing (WGS) analysis of FFPE samples could enable novel insights from these vast archival sample collections, robust WGS has remained elusive due to poor yields of fragmented DNA, low complexity sequencing libraries, uneven genomic coverage, and sequencing artifacts attributed to FFPE fixation. To examine WGS of stored FFPE samples, we analyzed 20 paired esophageal carcinoma (EC) samples (primary tumors and matched germline samples) that had been stored for 10-16 years using a proprietary extraction and library preparation methodology (SeqPlus). We assessed SeqPlus performance on these older FFPE tissues and measured variant call concordance between WGS and a targeted, high-depth sequencing panel (269 genes, sequenced at >400x mean coverage). At a planned WGS tumor sequencing depth of 70x, 93% of the genome was covered by  $\geq 20$  reads, 99% of bases had 10x coverage; the average duplicate reads were 31%. Importantly, we observed similar transition/transversion ratios and mutational spectra as those reported from The Cancer Genome Atlas (TCGA) fresh-frozen EC specimens. The overall concordance of tumor-specific SNV and indel variant calls derived from WGS and the targeted panel was high at 86%. All 76 targeted panel variants above the WGS lower limit of detection (mutant allele frequency (MAF) >10%) were also detected by WGS, 2 variants in 2 tumors were detected only by WGS, and 12 variants at MAF  $\leq 6\%$  from 9 tumors were exclusively detected with the targeted panel. SeqPlus produced consistent coverage versus the targeted panel that showed uneven coverage across the targeted genes/samples. WGS detected an overall mean of 10.4 putative cancer variants per tumor, like the 12 variants per patient reported from frozen TCGA EC specimens. Copy number variation analysis of the WGS data also identified genomic amplification of *CCND1*, *EGFR*, *TP63*, and *SOX2*, frequent deletion of the *CDKN2A/B* tumor suppressor locus, and additional previously unrecognized genomic aberrations. In conclusion, our study further supports the utility of high-quality, uniform WGS sequencing of archival FFPE cancer samples with SeqPlus and unlocks the potential for massive-scale retrospective genomic analysis of millions of pathology samples with their associated clinical, therapeutic, and outcomes data.

## PgmNr 1754: Two NGS library prep panels for analyzing tumor cells for actionable mutations using a flexible, automated microfluidics platform.

### Authors:

T. Goralski; M. Gonzales; J. Perez; J. Alipaz; J. Brockman; S. Chamnongpol; B. Fowler; J. Geis; C. Kubu; R. Kung; J. Qin; G. Sun; C. Park; D. King

[View Session](#) [Add to Schedule](#)

**Affiliation:** Research and Development, Fluidigm Corp, South San Francisco, California.

Genomic analysis of tumor cells is becoming increasingly common as more cancer investigators adopt next-generation sequencing (NGS) and as DNA sequencing costs continue to decrease. Currently, targeted sequencing is the most prevalent strategy for screening large numbers of samples. However, library prep can be expensive and tedious to perform. To provide a solution, we developed two amplicon-based library prep panels targeting actionable variants in cancer cells. The Advanta Solid Tumor NGS Library Prep Assay identifies single nucleotide variants (SNVs), insertions and deletions (indels), and copy number variations (CNVs) in 53 genes. It is comprised of 1,508 amplicons, covering >230 kb of genomic DNA with a bias for protein coding regions, and >18,000 pathogenic COSMIC variants. The complementary Advanta RNA Fusions NGS Library Prep Assay detects gene fusion events from tumor RNA. Its 1,100 amplicons target >1,000 breakpoints in 385 fusion gene pairs reported to be involved in solid tumors and hematologic cancers.

Both panels use a highly multiplexed, nanoliter-scale, PCR-based enrichment method on the Juno system. Libraries can be prepared from 48 samples from one or both panels in a single 9.5-hour run that requires less than 4 hours of hands-on time. Automated assembly of amplification reactions is performed on an integrated microfluidic circuit (IFC). Primer pairs are distributed across eight 270 nL reaction wells, enabling overlapping coverage that reduces amplification of off-target artifacts. As little as 12.5 ng of genomic DNA or cDNA from 10 ng of RNA detects SNVs and indels at a variant allele frequency of  $\geq 5\%$ , CNVs at  $\geq 3.5$  copies, and RNA fusion transcripts at  $\geq 250$  copies of input. Here we compare the performance of these panels using genomic reference standards and formalin-fixed, paraffin-embedded, fresh-frozen, and blood samples employing potential alternative workflows. These new library prep research panels, when integrated with the Juno system, provide valuable content in a flexible and economical platform with convenient workflows for investigators who need comprehensive genomic profiling of tumor samples.

## PgmNr 1755: Understanding genetic variation in cancer using targeted long-read sequencing.

### Authors:

S. Iyer <sup>1,2</sup>; M. Kramer <sup>1</sup>; S. Goodwin <sup>1</sup>; W.R. McCombie <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Center, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 2) Program in Genetics, Stony Brook University, Stony Brook, NY

---

Cancer cells acquire varied genetic aberrations through different degrees of genomic instability during tumorigenesis, evolution, and disease progression. Understanding the origin and evolution of tumor complexity by studying tumor heterogeneity in cancer could help reveal new strategies for improved diagnosis and precision therapy. Structural variations (SV), a hallmark of genomic instability in cancer, include insertions, deletions, duplications, inversions, or translocations that can either activate oncogenes or inactivate tumor suppressor genes. While short-read sequencing has aided cancer genomics, it has performed poorly in SV detection, with false positive and false negative rates of 50% or more. Long-read, single-molecule sequencing methods such as Pacific Biosciences single-molecule real-time (SMRT) sequencing and Oxford Nanopore Technologies (ONT) sequencing can address larger variations as they typically generate read lengths of tens of thousands of bases. Long reads can span SV with a single continuous read, giving a clearer idea of the variation, its position, and size. In the recent years, these sequencing strategies have helped identify thousands of genomic features pertinent to cancer that were previously missed by short-read sequencing. However, the throughput and coverage offered by these platforms, coupled with the high cost of whole genome long-read sequencing makes it infeasible to conduct large-scale genomic studies, limiting our understanding of the distribution and frequencies of SVs in the population. Targeted sequencing significantly improves accuracy and coverage by offering the depth necessary to detect rare alleles in a heterogeneous population of cells. However, a lack of efficient long-read compatible targeting techniques makes it difficult to study specific regions of interest on existing long-read platforms. Currently we are evaluating amplicon-based and CRISPR/Cas-based systems for targeted long-read sequencing to enrich for specific regions of the cancer genome (BRCA1, HLA, BRCA2) in two breast cell lines – MCF 10A and SK-BR-3. Our overarching goal is to develop a targeted long-read sequencing strategy to facilitate large population SV analysis, which will help define the landscape of such variants in the population and help identify regions of therapeutic or diagnostic interest.

## PgmNr 1756: Capturing tumor heterogeneity in pre- and post-chemotherapy colorectal cancer ascites using scRNA-Seq.

### Authors:

N. Jinawath <sup>1,2</sup>; T. Poonpanichakul <sup>3</sup>; M.-S. Shiao <sup>4</sup>; N. Jiravejchakul <sup>5</sup>; P. Matangkasombut <sup>3,5</sup>; E. Sirachainan <sup>6</sup>; V. Charoensawan <sup>2,3,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Translational Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 2) Integrative Computational BioScience Center (ICBS), Mahidol University, Nakhon Pathom, Thailand; 3) Systems Biology of Diseases Research Unit, Faculty of Science, Mahidol University, Bangkok, Thailand; 4) Research Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 5) Department of Microbiology, Faculty of Science, Mahidol University, Bangkok, Thailand; 6) Medical Oncology Unit, Department of Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 7) Department of Biochemistry, Faculty of Science, Mahidol University, Bangkok, Thailand

---

Malignant ascites is the abnormal accumulation of fluid within the peritoneal cavity, caused by metastasis of several cancer types; colorectal cancer included. Cancer cells in ascites reflect poor prognosis and can be a source of recurrence. They also represent a good source of specimen for the study of tumor heterogeneity. Single-cell RNA sequencing has recently emerged as a powerful tool to explore and characterized cellular heterogeneity, and has been applied to different cancers. However, molecular profiles of cancer cells in ascites and their responsiveness to chemotherapy is poorly understood. In this study, we performed, to the best of our knowledge, the first scRNA-Seq analysis of pre- and post-chemotherapy malignant ascites-derived cells from a colorectal cancer patient in order to profile and characterize the cellular heterogeneity and different expression hallmarks within the population. Unbiased clustering of 17,967 cells in total reveals 16 sub-clusters of four major cell types: epithelial cells, myeloid cells, fibroblast, and lymphocytes. We identified distinct epithelial cluster that is mainly seen in pre-treatment samples, as well as another distinct cluster in post-treatment specimens. Interestingly, the percentages of recovered cells from different cell types are not only affected by the chemotherapy treatment, but also appear to be influenced by the single cell preparation protocols, namely enzymatic and mechanical separations. Single cell western blotting was subsequently used to validate the differential amount of recovered cell types between the two single cell preparation methods. Overall, differential expression (DE) analysis between pre- and post-treatment malignant ascites-derived cells reveals several differentially expressed transcription factors and ribosomal genes. Gene Set Enrichment Analysis (GSEA) emphasizes different states of cancer epithelial cells within the patient. In summary, our study underscores cancer heterogeneity within malignant ascites and uncover gene expression profile of ascites-derived colorectal cancer cells in single-cell resolution.

## **PgmNr 1757: Tumor mutational burden analysis: An ultra-multiplex amplicon-based NGS target enrichment approach for accurate, rapid and comprehensive profiling.**

### **Authors:**

L. Lee; K. Pendleton; V. Kapoor; G. Liu; Y.L. Liu; J. Liu; Z. Liu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Paragon Genomics, Inc., Hayward, CA.

---

Assessment of tumor mutational burden (TMB) using NGS methods, with its correlation to patient response to checkpoint inhibitor chemotherapy in the immuno-oncology field, is gaining significant importance. A whole exome sequencing approach has most often been used, although they can be expensive and time consuming. Alternatively, a targeted enrichment strategy using multiplexed PCR can provide better coverage of regions of interest and reduce costs. But, as required for tumor profiling, increasing multiplexity of PCR can lead to problems with non-specificity, primer-primer interactions, decreased uniformity, and increased GC bias. In this study we present a high-quality ultra-multiplex PCR-based strategy for targeting hundreds of genes. We designed a 500 gene TMB panel which uses 27,296 pairs of primers to target them. The lengths of the amplicons were designed to be 105-120bp to allow for target amplification from challenging samples like FFPE DNA and cfDNA. The 27,296 pairs of PCR primers are pooled into two multiplex PCR reactions and cover a total of 1635204bp (1192796 bp of coding sequence). Paragon Genomics' CleanPlex® technology enables a library preparation workflow which includes a multiplex PCR step with the targeted primers in two pools. After the multiplex PCR step, the two pools are combined. A subsequent cleanup step removes by-products, and a final PCR adds sequencer flow cell adapters and sample barcodes. With a read depth of ~2400 reads per amplicon, the TMB panel exhibits ~95% uniformity at 0.2X mean and little to no GC bias despite the highly multiplexed nature of this panel. High mapping and on-target rates of > 90% makes most of the data usable for downstream analysis. The CleanPlex® TMB Panel targets 500 genes using 27,296 amplicons generating ultra-multiplexed libraries with high coverage uniformity (~95%), low GC bias and high mapping and on-target rates. The advanced design of this panel along with the CleanPlex technology enables the rapid interrogation of these genes in a robust assay.

## **PgmNr 1758: Identifying cooperating disease-driving mutations in mutant p53-initiated invasive carcinomas and ductal carcinoma *in situ*.**

### **Authors:**

R. Morrissey <sup>1,2</sup>; J. McDaniel <sup>1</sup>; Y. Zhang <sup>3</sup>; S. Xiong <sup>1</sup>; G. Lozano <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Genetics, MD Anderson Cancer Center, Houston, TX.; 2) Genetics and Epigenetics Graduate Program, The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX; 3) Department of Pharmaceutical Sciences, Texas Southern University, Houston, TX

---

Ductal carcinoma *in situ* (DCIS) is an early breast tumor confined to the duct by myoepithelial cells. DCIS sometimes progresses to invasion, but does not always do so. Since it is difficult for clinicians to predict whether or not DCIS will become invasive, it is treated with surgical removal and radiation. To prevent overtreatment of patients, a strategy must be devised to stratify tumors that are likely to spread and tumors that will remain localized.

We have generated a somatic mouse model with a conditional wild-type to mutant (WM) R245W p53 mutation only in mammary tissue, conferring predisposition to breast tumorigenesis. Importantly, these mice retain a wild-type stroma and immune system. Pathological analysis indicated that the progression of breast cancer in our mouse model mirrors that in humans, and identified both invasive carcinomas (IC) and DCIS. These IC tumors have long latency (median survival of 1.5 years), thus, we hypothesize that additional mutations are required for DCIS to progress to IC.

DCIS and IC tumors will be identified via pathological examination and immunohistochemical staining. Lesions will be extracted through laser capture microdissection. Exome sequencing will be used to identify recurrent mutations cooperating with mutant p53 to drive breast cancer progression. RNA sequencing will be used to identify genes and pathways that are differentially expressed among DCIS and IC lesions. After identification of recurrent mutations and differentially expressed genes that are exclusive to IC, we will characterize the function of these mutations in driving breast cancer disease progression.



## PgmNr 1759: Molecular signatures of multifocal prostate cancer.

### Authors:

D. Noursome; R. Narayanan; J. Linehan; R. Gentry; K. Tran; S. Hsiang; D. Krasne; J. J alas; M.P. Salomon; D. Hoon; T. Wilson

[View Session](#) [Add to Schedule](#)

**Affiliation:** John Wayne Cancer Institute, Santa Monica, California.

**Background:** Prostate cancer (PCa) usually presents as histologically multifocal lesions (MF), with the potential for differences in aggressiveness among foci. In contrast, single focus (SF) patients are younger with lower Gleason score. Previous studies have attempted to characterize the somatic nature of MF tumors, but additional molecular studies are necessary to identify differences between foci. The goal of this study was to generate a molecular signature between foci in MF PCa, identify differences between MF and SF patients, and to characterize regional metastases to lymph nodes (LN) in patients who have undergone radical prostatectomy.

**Methods:** We performed miRNA sequencing using the HTG EdgeSeq miRNA whole transcriptome next generation sequencing assay for ~2,000 miRNAs on 27 patients (10 MF and 17 SF). The primary and secondary foci of MF patients were determined by MRI. There were 82 samples of LN metastases collected (mean=5.1 LN per patient). We tested differential miRNA expression between primary and secondary foci for each patient. In addition, we assessed differential expression between MF and SF primary foci. Using miRNA markers, we calculated Euclidean distances among foci and LN to assess the degree of relatedness within each patient. For a subset of patients, we performed RNA-seq (6 MF and 3 SF) to characterize cell types enrichment using xCell.

**Results:** For MF patient specimens, two miRNAs (*miR-221-5p* and *miR-378a-3p*) were differentially expressed between the primary and secondary foci. Between MF and SF patients, three miRNAs were differentially expressed (*miR-6515-3p*, *miR-670-5*, and *miR-5571-3p*). In MF patients, the primary lesion was more closely related to LN metastases compared to the secondary focus ( $P<0.001$ ). In MF patients, the primary foci had higher enrichment of T helper 2 cells ( $P=0.005$ ) and dendritic cells ( $P=0.07$ ) compared to the secondary foci. There was less enrichment of CD8+ cells in MF compared to SF patients ( $P=0.02$ ).

**Conclusions:** In this first miRNA expression profiling of MF PCa, we observed that multiple foci harbor distinct molecular signatures. miRNA and RNA signatures may be informative as biomarkers to discriminate MF from SF patients. These signatures indicate the different etiology between primary and secondary foci and also between MF and SF patients. Using molecular profiling to better understand intra-tumor heterogeneity may influence the development of novel surgical approaches in patient management.



## PgmNr 1760: Comprehensive detection of germline and somatic structural mutation in cancer genomes by Bionano genomics optical mapping.

### Authors:

A.W.C. Pang; J. Lee; K. Hong; T. Anantharaman; E.T. Lam; A. Hastie; M. Borodkin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bionano Genomics, San Diego, California.

---

The ability to identify structural variants (SVs) is crucial in cancer genetics. Karyotype and cytogenetics are manually intensive. Microarrays and sequencing cannot detect calls in segmental duplications and repeats, and miss balanced variants and low-frequency mutations.

We describe the Bionano Genomics's Saphyr platform to identify SVs in cancer genomes. DNA >100 kbp is extracted, labelled at specific motifs, and linearized through NanoChannel arrays. Molecule images are digitized and *de novo* assembled, creating chromosomal-arm scale genome maps. Cancer mutations >500 bp are detected by aligning the molecules or the genome maps to the public reference.

We ran Bionano's cancer workflow on multiple human cancer cell lines. While the number of SVs varies among samples, we typically observe > 3500 calls per genome. In the SK-BR-3 breast cancer genome, we detected a cluster of amplifications, and translocations on chr8, impacting the gene *MYC*. In the CML genome K562, the *BCR-ABL* translocation was detected, while we also detect novel rearrangements, such as insertion and inversion interrupting the gene *NAALADL2* in a prostate cancer cell line LNCaP.

In conclusion, with one platform, Saphyr can discover a broad range of traditionally refractory but relevant SVs, and improves our understanding of cancer.

## PgmNr 1761: Crowdsourced RNA-seq and gene ontology analysis of a human breast tumor in the context of an undergraduate course.

### Authors:

P. Soneral; K. Bishop

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bethel University, Department of Biological Sciences, 3900 Bethel Drive, St. Paul, Minnesota.

---

Breast cancer's lethality is due, in part, to inherent heterogeneity of breast cancer disease, posing challenges for clinicians to accurately diagnose and subtype tumors. Recently, RNA-seq technology and bioinformatic analysis of gene expression has made it possible to subtype tumors based on molecular classification. In this experiment, our aim was to capture tumor diversity while improving the redundancy and quality of library preparation for transcriptome analysis as a diagnostic tool for breast tumor subtyping. Using a crowdsourcing approach in the context of an undergraduate course, we prepared a whole transcriptome library from a single breast cancer patient, and used multiplex sequencing on twelve uniquely barcoded preparations, which were subsequently pooled and sequenced in a single run on a high-throughput instrument. Datasets obtained from the sequencing were subjected to quality control analysis using the Green Line of DNA Subway (Tuxedo Pipeline, FAST-X, FAST-QC, Tophat, and CuffDiff algorithms). Log-transformed gene expression ratios were subjected to the BiNGO gene ontology algorithm. Significant over-expression of genes for metabolic and cellular process pathways included cell death and cell division ( $p < 0.01$  Binomial). Significantly down-regulated pathways included those in development (e.g. organ development), regulation (stimulus response, inflammatory response, etc), and signaling (cellular communication, etc;  $p < 0.01$  Binomial). Furthermore, using previously established biomarker genes, we subtyped the cells as basal-like, showing little change in expression for Estrogen Receptor (ER), Her2/Neu (HER), or Ki-67 biomarkers, but a significant over-expression of the progesterone receptor ( $q = .0031$ ). Using the Mammprint suite of biomarkers we diagnosed a highly progressed tumor, wherein 60% and 73% of the Mammprint mid-hallmark genes and 38% late hallmarks were differentially expressed. Based on this dataset, we propose the biomarker MELK (maternal embryonic leucine zipper kinase) as a candidate for personalized chemotherapy. Taken together, our findings suggest that crowdsourcing RNA-seq experiments can be a valid approach for improving library preparation toward subtyping and diagnosis of human breast cancers. Furthermore, this can be achieved with undergraduate researchers in the context of coursework, contributing to their learning gains.

## PgmNr 1762: Human papillomavirus and the landscape of secondary genetic alterations in oral cancers.

### Authors:

D.E. Symer<sup>1</sup>; K. Akagi<sup>2</sup>; W. Xiao<sup>2</sup>; B. Jiang<sup>2</sup>; R.K.L. Pickard<sup>3</sup>; J. Li<sup>3</sup>; B.J. Swanson<sup>4</sup>; A.D. Agrawal<sup>5</sup>; M. Zucker<sup>6</sup>; B. Stache-Crain<sup>7</sup>; A.-K. Emde<sup>8</sup>; H.M. Geiger<sup>8</sup>; N. Robine<sup>8</sup>; K.R. Coombes<sup>6</sup>; M.L. Gillison<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Lymphoma & Myeloma, Univ. Texas MD Anderson Cancer Center, Houston, TX; 2) Thoracic/Head and Neck Medical Oncology, University of Texas MD Anderson Cancer Center, Houston, TX; 3) Division of Medical Oncology, Department of Internal Medicine, Ohio State University, Columbus, OH; 4) Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE; 5) Department of Otolaryngology – Head and Neck Surgery, Ohio State University Comprehensive Cancer Center, Columbus, OH; 6) Department of Biomedical Informatics, Ohio State University Comprehensive Cancer Center, Columbus, OH; 7) Complete Genomics, Mountain View, CA; 8) New York Genome Center, New York, NY

---

Human papillomavirus (HPV) is a necessary but insufficient cause of a subset of oral squamous cell carcinomas (OSCC) that is increasing markedly in frequency. To identify contributory, secondary genetic alterations in these cancers, we used comprehensive genomics methods to compare 149 HPV-positive and 335 HPV-negative OSCC tumor/normal pairs. Different behavioral risk factors underlying the two OSCC types were reflected in distinctive genomic mutational signatures. In HPV-positive OSCC, the signatures of APOBEC cytosine deaminase editing, associated with anti-viral immunity, were strongly linked to overall mutational burden. By contrast, in HPV-negative OSCC, T>C substitutions in the sequence context 5'-ATN-3' correlated with tobacco exposure. Universal expression of HPV E6\*1 and E7 oncogenes was a sine qua non of HPV-positive OSCC. Significant enrichment of somatic mutations was confirmed or newly identified in *PIK3CA*, *KMT2D*, *FGFR3*, *FBXW7*, *DDX3X*, *PTEN*, *TRAF3*, *RB1*, *CYLD*, *RIPK4*, *ZNF750*, *EP300*, *CASZ1*, *TAF5*, *RBL1*, *IFNGR1* and *NFKBIA*. Of these, many affect host pathways already targeted by HPV oncoproteins, including the p53 and pRB pathways, or disrupt host defenses against viral infections, including interferon and nuclear factor-kB signaling. Frequent copy number changes were associated with concordant changes in gene expression. Chr. 11q (including *CCND1*) and 14q (including *DICER1* and *AKT1*) were recurrently lost in HPV-positive OSCC, in contrast to their gains in HPV-negative OSCC. High-ranking variant allele fractions implicated *ZNF750*, *PIK3CA* and *EP300* mutations as candidate driver events in HPV-positive cancers. We conclude that virus-host interactions cooperatively shape the unique genetic features of these cancers, distinguishing them from their HPV-negative counterparts.

## **PgmNr 1763: Single NGS assay used to evaluate MHC loss of heterozygosity (LOH) and tumor mutational burden (TMB) and determine their impact on response to immunotherapy.**

### **Authors:**

J. Thompson; G. Cerqueira; L. Keefer; D. Nichol; J. Meyer; N. Dracopoli

[View Session](#) [Add to Schedule](#)

**Affiliation:** Research & Development, PGDx, Baltimore, Maryland.

Many tumor cells survive due to a suppressed immune response that prevents their destruction. This may be overcome during treatment via checkpoint inhibitors. These act via a monoclonal antibody that blocks the interaction of the checkpoint receptors with their cognate ligands. Tumor Mutational Burden (TMB) serves as a surrogate marker of immune response because it provides a readout on altered proteins that should but are not recognized by the immune system. TMB is measured through sampling regions of the cancer genome to estimate the number of mutations/Mb. A high TMB score is associated with better response to immunotherapy because the tumor carries more somatic mutations and has a higher chance of presenting an immunogenic neoepitope. However, not all high TMB tumors are impacted by therapy. One reason can be loss of antigen presentation caused by mutations and loss of heterozygosity (LOH) of the beta-2-microglobulin (B2M) and MHC Class I genes. Consequently, the outcome for high TMB tumors will be dependent on their ability to present antigens with their MHC Class I/B2M complex.

TMB should be considered together with the tumor's ability to present these putative neoantigens. To test this hypothesis, we used the PGDx elio™ tissue complete assay (in development; >500 genes covering 1.3 Mb) to measure TMB and antigen presentation in the same assay. We tested 190 cancer patients and showed that in FFPE tissue samples with ≥20% tumor content, we could detect LOH of the MHC Class I with 88% accuracy. This confirmed our hypothesis that it was possible to measure TMB and evaluate antigen presentation in the same NGS analysis.

We have replicated the combined analysis of TMB and LOH of MHC in a publicly available NSCLC cohort with patient outcome results to checkpoint inhibitor treatment. Patients that were either TMB-low (≤120 mutations per exome) or TMB-high but with predicted LOH of MHC had a poorer outcome and higher hazard ratio (7.9, CI 95% 1.3-49) than patients with high TMB and intact MHC. Thus, our algorithm can be used in the combined evaluation of TMB and the potential for antigen presentation to predict patient outcome. These methods are being used to further evaluate TMB and somatic alterations of the antigen presentation complex in different cancers and to further validate the hypothesis that adding neoantigen presentation capability to TMB scores will significantly improve prediction of response.

## PgmNr 1764: Structurally resolved immune profiling in ovarian carcinoma using whole-transcriptome spatial sequencing.

### Authors:

S.R. Williams <sup>1</sup>; L.D. Gibbs <sup>2</sup>; N.I. Weisenfeld <sup>1</sup>; D.W. Craig <sup>2</sup>; N.F. Delaney <sup>1</sup>; D. Da Silva <sup>3</sup>; Y. Yin <sup>1</sup>; S. Rotimi <sup>2</sup>; J. Chew <sup>1</sup>; R. Bassiouni <sup>2</sup>; M. Frey <sup>1</sup>; J. Qian <sup>2</sup>; M. Webb <sup>2</sup>; H. Miller <sup>3</sup>; E. Velazquez <sup>2</sup>; L. Murderspach <sup>3</sup>; T. McEachron <sup>2</sup>; L. Roman <sup>3</sup>; J.D. Carpten <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computational Biology, 10x Genomics, Pleasanton, California.; 2) Department of Translational Genomics, Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Department of Obstetrics and Gynecology, Keck School of Medicine, University of Southern California, Los Angeles, CA

---

In the United States approximately 22,000 women will receive a new diagnosis and ~14,000 women will die from ovarian cancer (OC) each year, making this disease the fifth most deadly cancer among women. Like many cancers, OC can be genetically heterogeneous making studies difficult to plan, execute, and interpret. Bulk methods such as whole-genome or whole-transcriptome sequencing are limited in their ability to resolve fine grain molecular signatures which hinder their utility in dissecting the underlying biology of individual tumors.

Here we present investigation of 11 subjects, previously diagnosed with OC using whole-transcriptome spatial sequencing on the 10x Genomics Spatial Transcriptomics platform. This system requires no tissue dissociation keeping fragile cell types intact. Using serial sections of solid tumors from each subject we were not only able to profile each section at 100um resolution but also spatially resolve gene expression signatures and cluster regions of tissue based on these signatures. Next, we aggregated the transcriptional profiles of serial sections from each case increasing our power to cluster similar regions and identify differentially expressed genes within these tissues.

Using this information, we went on to conduct a refined analysis of 18 key inflammatory genes that have been previously known to play an important role in the immune response of OC and developed a spatially resolved inflammation score for each section from each subject. Not only did we observe statistically significant regionalized inflammation scores within each subject, we also found intersubject differences which may give insight into response to treatment. Further, we observed clear transcriptional substructure in some tumors, where different tumor regions were defined by molecular pathways that are associated with tumorigenesis (i.e. UV Response, Epithelial to Mesenchymal Transition, and Estrogen Response pathways), regions defined by immune populations (B cells, Macrophages, and CD4+ T cells), and molecular pathways that are associated with immune infiltration (i.e. complement, IL2, and IL6 signaling pathways). These observations highlight the intra-subject heterogeneity of OC in our study. This approach demonstrates the power of using spatial whole-transcriptome sequencing in solid tumor studies to help unravel the complexity of heterogeneous cancers.

## PgmNr 1765: Somatic mutation and copy number alteration profile in chordoma.

### Authors:

R. Yang <sup>1</sup>; H. Koka <sup>1</sup>; W. Zhou <sup>1,2</sup>; W. Luo <sup>1,2</sup>; S. Yepes <sup>1</sup>; K. Jones <sup>1,2</sup>; A. Vogt <sup>1,2</sup>; M. Wang <sup>1,2</sup>; B. Zhu <sup>1,2</sup>; B. Hicks <sup>1,2</sup>; D. Parry <sup>1</sup>; A. Goldstein <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) div Cancer Epidemiology & Genetics, NCI, NIH, Bethesda, Maryland.; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Frederick National Laboratory for Cancer Research, Frederick, MD, USA

---

Chordoma is a rare bone tumor, which is believed to originate from notochordal remnants and occurs in the axial skeleton at cranial, spinal and sacral sites. Chordomas are considered slow-growing, however, local recurrences are common and treatment options are limited particularly for those with advanced disease, highlighting the need for improving our knowledge of the disease biology to discover novel druggable targets. In this study, we conducted targeted panel sequencing in paired tumor (Formalin-Fixed Paraffin-Embedded) and germline DNA of 57 chordoma patients residing in the United States or Canada. We also evaluated somatic copy number alterations (SCNAs) in a subset of the tumors (n=49) using an Illumina Infinium global screening array. The chordoma site distribution was 49% skull-base, 25% spinal, and 26% sacral, respectively. The most recurrent somatic mutations in this tumor set included *PIK3CA* (n=6), *LYST* (n=5), *SWI/SNF* genes (*PBRM1* [n=2], *SETD2* [n=3], and *SMARCB1* [n=1]), and *USP9X* (n=3). Most *LYST* mutations (4 out of 5) occurred in tumors from females or sacral chordomas, while all 3 mutations in *USP9X* occurred in spinal tumors. Other mutations included *TP53*, *PALB2*, *MAP3K4*, *ATM*, *TSC2*, each in a single tumor. SCNA arm-level profile was consistent with previous reports, including frequent gains of chromosomes 1q, 2, 5, and 7, and deletions of 1p, 3, 4, 9, 10, 13q, 14q, and 18. Gains of chromosomes 1q and 2 and deletions of 13q were more likely to occur in skull-based tumors, whereas deletions of chromosomes 5, 9, and 18 were more prevalent in sacral tumors or chordomas from older patients. GISTIC analysis identified several significant focal SCNA regions, including the deletion peak of the 9p21.3 region containing the *CDKN2A* gene. Homozygous 9p21.3 deletions were seen in 4 tumors. Focal amplification of the *T* gene was seen in 3 tumors (two skull-base, one sacral), which occurred at a much lower frequency compared to the frequency (>20%) reported in a published sacral chordoma sequencing study. In summary, results from our study provide further evidence that mutations in *SWI/SNF* complex genes, *PIK3CA*, and *LYST*, as well as *CDKN2A* homozygous deletions might be driver events in chordoma and some of these genomic events might vary by chordoma sites.

## **PgmNr 1766: Correlation between mutations found in FFPE tumor tissue and paired cfDNA samples.**

**Authors:**

B. Niccum; L. Saunders; A. Hur; A. Patel

[View Session](#) [Add to Schedule](#)

**Affiliation:** Beckman Coulter Life Sciences, Indianapolis, IN.

---

Cell free DNA (cfDNA) consists of small (150 – 500 bp) DNA fragments that circulate in the blood. cfDNA levels tend to be low in healthy, non-pregnant patients, and increase in patients with cancer, pregnancy, or extensive damage to tissue. cfDNA is believed to derive mostly from apoptotic cells, and biomarkers for a variety of diseases have been found in cfDNA. As cfDNA is extracted from blood, it is a non-invasive way to detect disease; however, there is some concern that cfDNA does not contain the same biomarkers as tumor tissue. This study measures the efficacy of cfDNA as a biomarker detection medium by comparing mutations found in both FFPE tumor samples and paired cfDNA samples. This study determines whether the same biomarkers are found in each sample type, and which of those can be used as biomarkers in both and which are preferred biomarkers for only a single sample type. Trends in mutation detection with the two different sample types are discussed.

## **PgmNr 1767: Comparison between mutation profiles of paired whole blood and cfDNA samples.**

### **Authors:**

A. Patel; L. Saunders; A. Hur

[View Session](#) [Add to Schedule](#)

**Affiliation:** Beckman Coulter Life Sciences, Indianapolis, Indiana.

---

Liquid biopsies are increasingly becoming a tool of choice for cancer detection and monitoring. NGS of cfDNA is coming into maturity as a non-invasive method to identify mutational profiles in many cancer types . Here we describe a simple method to isolate both gDNA and cfDNA from a patient blood sample and discuss the automation of both extractions. We show the efficacy of cfDNA as reliable biomarker analysis tool by comparing mutations in cfDNA vs whole blood. The study determines if the difference between tumor and germ-line mutations can be established and what are the limitations.



## PgmNr 1768: Functional validation of *RCCD1* as a candidate susceptibility gene for breast and ovarian cancer risk.

### Authors:

S. Chen; J. Plummer; S. Gayther; J. Kanska; F. Dezem

[View Session](#) [Add to Schedule](#)

**Affiliation:** Cedars-Sinai Medical Center, Los Angeles, California.

---

Breast and ovarian cancers share not only common etiologies and but similar genetics, such as mutations in *BRCA1/2* are responsible for most multi-case breast/ovarian cancer families. These cancers also share common genetic susceptibility variants suggestive of pleiotropic risk for both breast and ovarian cancer. Previous GWA and cis-eQTL analyses identified *RCCD1* as a likely candidate gene for both breast and ovarian cancer risk. To confirm *RCCD1* as the target gene, we utilize chromosome conformation capture assays to identify the physical interactions between this susceptibility gene and risk associated SNPs. We answer the question whether the same risk SNPs have shared interactions with *RCCD1* in their respective cancer or whether different SNPs at this locus act in a cell specific manner with respect to breast and ovarian cancer. Specifically, we use precursor and cancer cell lines for both cancer types to identify interactions between the *RCCD1* promoter and possible candidate causal variants at the 15q26 pleiotropic risk locus. The pleiotropic risk SNPs, rs763280402 and rs570569800 showed 4C interactions with the *RCCD1* promoter in both breast and ovarian cancer cells, hence at this locus the same SNPs appear to contribute to *RCCD1* regulation. Given regulatory elements are the likely functional targets of most risk-associated SNPs, we used regulatory feature annotation (Statehub) to show both SNPs lie in predicted enhancers in their respective cell type. Moreover, normal precursor cells exhibit more interactions (>5-10) compared to their respective cancer cells, demonstrating there are cell type specific interactions from normal to cancer states. Trans interactions were seen between the *RCCD1* promoter and 8q24/*MYC* another pleiotropic risk locus (q value=8.12 e<sup>-48</sup>). In fact, breast and ovarian cancer patients who have high *RCCD1* expression (P<0.004) show prolonged survival curves (KM Plotter). TCGA data shows an increased expression of *RCCD1* expression in ER positive tumors (P value <0.00001). We evaluated the functional role of *RCCD1* in these cancers by perturbing its expression in breast and ovarian cancer precursor cell lines. CRISPR knockdown of *RCCD1* show neoplastic phenotypes consistent with its role in oncogenesis. These findings show that using an integrated functional genomics approach to comprehensively evaluate interactions of a given breast and ovarian cancer risk locus can identify candidate susceptibility genes for both cancer types.

## PgmNr 1769: High throughput high molecular weight DNA extraction from human tissues for long-read sequencing.

### Authors:

K. Liu <sup>1</sup>; M. Kim <sup>1</sup>; J. Laugharn <sup>2</sup>; H. Khoja <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Circulomics Inc, Baltimore, MD; 2) Covaris Inc, Woburn, MA

---

Structural variants (SVs) such as inversions and translocations are increasingly recognized as drivers of genetic disease. SVs are difficult to detect with traditional short-read sequencing due to their size. However, recent studies using long-read sequencing and optical mapping methods from PacBio, Oxford Nanopore, and Bionano Genomics have found SVs to be 5-10X more prevalent than previously thought.

The ability of these methods to detect SVs is predicated on high purity, high molecular weight (HMW) DNA. Whereas germline SVs are readily detected using blood or saliva, somatic SVs can only be detected using tissues. Tissues present a particular challenge as they contain extracellular material that must be efficiently disrupted without damaging delicate HMW DNA.

The gold-standard for HMW DNA extraction is agarose plug lysis, a tedious two day method that can only process a small handful of samples. In contrast, spin column and magnetic particle methods are fast but fragment HMW DNA, compromising read lengths and SV detection sensitivity. High-throughput methods are needed to enable large scale studies of SVs and to underpin future clinical tests.

We present a high throughput method for HMW DNA extraction. First, Covaris cryoPREP automated dry pulverizer is used to cryofracture frozen tissue samples in tissueTUBEs. Compared to Dounce homogenizer or TissueRuptor, cryoPREP can rapidly process large numbers of samples with high reproducibility. Then, Circulomics Nanobind magnetic disks are used to extract HMW DNA from the cryofractured tissue powder. Each Nanobind disk is covered with micro- and nanostructured silica that protects DNA from shearing to enable higher DNA size, extraction yield, and purity than competing methods. The entire protocol can be completed in only 2 hours.

This method was tested on a comprehensive panel of human tissues including breast, kidney, colon, uterine fibroid, pancreas, and skeletal muscle. DNA size up to 300 kb was obtained across all tissue types with size surpassing 500 kb for some samples. 25 mg of tissue was used in each extraction and resulted in 3-40 µg of DNA. Extraction yields varied across tissue type but were consistent across replicates (11% CV). The DNA were then sequenced on Oxford Nanopore GridION. High read length N50 over 36 kb and throughputs up to 9 Gb were obtained. These high quality DNA samples are suitable for diverse methods including PacBio and 10X Genomics as well as Illumina sequencing.

## PgmNr 1770: Genomic landscape of subependymal giant cell astrocytomas: Unique gene expression profile with few somatic events.

### Authors:

K. Giannikou <sup>1,2</sup>; J. Kim <sup>2</sup>; Z. Zhu <sup>1</sup>; M.E. Tyburczy <sup>1</sup>; D. Marron <sup>3</sup>; J.S. Parker <sup>3</sup>; Z. Hebert <sup>4</sup>; K.D. Willen <sup>5</sup>; D.M. Meredith <sup>6</sup>; S. Alexandrescu <sup>7</sup>; D. Rosenberg <sup>2</sup>; A. Bongaarts <sup>8</sup>; M. Nellist <sup>9</sup>; K. Kotulska <sup>10</sup>; S. Jozwiak <sup>10</sup>; O. Devinsky <sup>11</sup>; M. Mandra <sup>12</sup>; H. Lidov <sup>13</sup>; A. Tracy <sup>2</sup>; G. Getz <sup>2</sup>; E.A. Thiele <sup>14</sup>; M. Sahin <sup>4</sup>; E. Aronica <sup>15,16</sup>; D.W. Ellison <sup>17</sup>; K.L. Leigon <sup>7</sup>; D.J. Kwiatkowski <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cancer Genetics Laboratory, Division of Pulmonary and Critical Care Medicine and of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, 02115, MA, USA; 2) Broad Institute of Harvard and MIT, Cancer Genome Program, Cambridge, MA, USA; 3) Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 4) Molecular Biology Core Facility, Dana Farber Cancer Institute, Boston, 02115, MA, USA; 5) Department of Neurology, Boston Children's Hospital, Harvard Medical School, 02115, Boston, MA, USA; 6) Department of Pathology, Boston Children's Hospital, Harvard Medical School, Boston, MA, 02115, USA; 7) Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, 02115, USA; 8) Department of (Neuro)Pathology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 9) Department of Clinical Genetics, Erasmus Medical Centre, Rotterdam, The Netherlands; 10) Department of Neurology and Epileptology, Children's Memorial Health Institute, Warsaw, Poland; 11) Department of Neurology, New York University School of Medicine, 223 E 34 Street, New York, New York 10016, USA; 12) Department of Pediatric Neurosurgery, Medical University of Silesia, Katowice, Poland; 13) Department of Pathology, Boston Children's Hospital, 300 Longwood Ave, Bader, Boston, MA, 02115, USA; 14) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 15) Stichting Epilepsie Instellingen Nederland (SEIN), The Netherlands; 16) Swammerdam Institute for Life Sciences, Center for Neuroscience, University of Amsterdam, Amsterdam, The Netherlands; 17) Department of Pathology, St. Jude Children's Research Hospital, Memphis, 38105, TN, USA

**Background:** Subependymal giant-cell astrocytomas (SEGAs) are rare, slow-growing glioneuronal brain tumors, seen in 10-15% of individuals with Tuberous Sclerosis Complex (TSC). Though histologically benign, SEGAs can cause serious neurological problems and be fatal. SEGAs show *TSC1/TSC2* biallelic loss and distinct gene expression patterns compared to normal brain.

**Aim:** To define other somatic genetic events beyond *TSC1/TSC2* that occur during SEGA development and examine transcriptional differences in comparison to other CNS brain tumors.

**Material and Methods:** Exome sequence analysis was performed in 21 paired resected SEGA tumors (Illumina platform). Whole transcriptome RNA-seq analysis was performed on 16 SEGAs and was compared to TCGA Low Grade Gliomas (LGG; n=530), glioblastoma (GBM; n=171) and gangliogliomas (n=10).

**Results:** Fifteen germline and 5 somatic variants were identified in *TSC2* and 5 germline variants in *TSC1* in 21 SEGAs. Copy neutral Loss of Heterozygosity (CN-LOH; size range: 2.9-46Mb) was seen in 76% (16/21) of SEGAs (35% chr9q and 65% chr16p). Thirty-six somatic non-synonymous variants in

genes other than *TSC1/TSC2* were identified in 12/21 (57%) SEGAs (range:0-6 variants, average of 2.6, median AF 16%), mainly missense changes of uncertain clinical significance. Differential gene expression analysis of SEGAs compared to TCGA LGG, GBM and gangliogliomas RNA-seq data revealed >1500 differentially expressed genes, including several homeobox transcription factors (TFs) with pivotal roles in brain development; *HMX3*, *HMX2*, *IRF6*, *VAX1*, *SIX3*, *TWIST2*, *ZBTB20*, and *EOMES* were all expressed at median >9-fold higher ( $p < 0.003$ ,  $FDR < 0.05$ ) in SEGAs. These TFs were validated by IHC in multiple sections from SEGAs in comparison to normal brain, cortical tuber, GBM, and LGG subtypes including diffuse astrocytoma, oligodendroglioma, ganglioglioma, pleomorphic xanthoastrocytoma and subependymoma. Unsupervised weighted gene co-expression network analysis (WGCNA) identified 24 modules of co-expressed genes, and we found that 6 of the 8 TFs above belonged to the module that showed the most consistent difference between SEGAs and LGG/GBM.

**Conclusions:** SEGAs have one of the lowest somatic mutation rates known for human cancers. *TSC1/TSC2* loss appears to be sufficient, while other somatic variants are likely background. The highly expressed TFs we identified are candidate key drivers of SEGA growth, and serve as novel diagnostic markers and potential therapeutic targets.

## PgmNr 1771: DNA “nanomapping” using CRISPR/Cas9 as a programmable nanoparticle.

### Authors:

S. Koebley <sup>1</sup>; A. Mikheykin <sup>1</sup>; A. Olsen <sup>1</sup>; K. Leslie <sup>1</sup>; D. Guest <sup>1</sup>; O. Payton <sup>2</sup>; L. Picco <sup>1</sup>; J. Reed <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Physics Dept, Virginia Commonwealth University, Richmond, VA; 2) Interface Analysis Centre, H. H. Willis Physics Laboratory, University of Bristol, Bristol, UK; 3) Massey Cancer Center, Virginia Commonwealth University, Richmond, VA

---

Next-generation sequencing (NGS) has revolutionized biomedical research, yet the effective transfer of NGS to the clinical sphere is limited by its per-patient cost, time to construct an assembly, and short-read limitations. Our lab has developed a powerful complement to NGS: a high-speed atomic force microscopy (HSAFM) platform for “nanomapping” the genome. Our HSAFM, the fastest atomic force microscope ever developed, “feels” DNA with a nanoscale tip to yield topography maps with sub-nanometer z-resolution and 15 bp lateral resolution. Enzymes like Cas9 can be easily distinguished from DNA and precisely located in these images. We thus use nuclease-inhibited Cas9 as a programmable biomarker for nanoscale mapping, in which an observed pattern of sgRNA-Cas9 labels is compared to an *in silico* reference to reveal the presence of variants. Our approach is applicable as a programmable assay tuned to any hotspot of genetic mutation and can accommodate DNA fragments sized from <100 bp to >1 Mb. We have demonstrated the method’s efficacy and sensitivity using multiple oncogene targets in blinded cell line and clinical sample experiments, yielding metrics that meet or exceed those of standard assays. Unlike many standard or proposed assays, sample preparation for our technique is cheap and simple, image processing is conducted in near real-time, and the most expensive component of our HSAFM—its optical unit—can be replaced by a commercial DVD optical pickup. Thus, at a fraction of the cost and time of NGS and other comparable techniques, HSAFM nanomapping has the potential to serve as a viable tool for the assessment of genetic disease in the clinical realm.

## **PgmNr 1772: Genome analysis through electric signal detection of nucleotide incorporation using a portable DNA sequencer.**

### **Authors:**

D. Harrison; J. Idol; C. Wong; C. Ngan; C. Wei

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genome Technologies, The Jackson Laboratory, Farmington, CT.

---

Advancing our ability to produce accurate, high throughput and cost effective sequencing data is the driving force for biomedical discovery and technology innovation. Here, we evaluate a new sequencing platform from Genapsys for its workflow and performance in sequencing quality, output and robustness. The Genapsys technology utilizes a novel sequencing-by-synthesis approach that employs electrical signal detection of nucleotide incorporations on a semiconductor chip, allowing for highly accurate sequence detection. The system adopts standard library preparation followed by DNA template amplification on beads and the sequencing reaction on a scalable and portable format. Using microbial genomes comprised of a wide range of sequence contexts as well as the human reference NA12878 genome, we assess the performance of the Genapsys technology on its accuracy, coverage, error profile, GC bias and alignment ratio in comparison with the Illumina SBS technology. With an early access system, we have demonstrated the high quality of the sequencing data and utility in mutation detections in targeted gene panels. Beyond sequencing of standard control samples for reference benchmarking, we have detected the mutations in genes frequently altered in glioblastoma tumor samples. The quality of data, combined with its highly portable nature and cloud-based data processing workflow, should make this as promising platform for clinical utility.

## **PgmNr 1773: Identification of germline variants for hereditary cancer using the rapid, CleanPlex® multiplexed NGS panel.**

**Authors:**

K. Pendleton; V. Kapoor; G. Liu; L. Lee; Y.L. Liu; J. Liu; Z. Liu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Paragon Genomics, Hayward, CA.

---

The demand for genetic testing for hereditary cancers has been increasing, especially to aid with testing and diagnostics of patients. High quality variant calling and ease of workflow are critical for these assays. We have developed the CleanPlex® Hereditary Cancer Panel V2 for rapid production of NGS libraries with high-quality coverage of relevant hereditary cancer targets. This panel utilizes multiplex PCR of 1445 primer pairs to cover the coding sequence of 37 genes, two SNPs in the BRCA1 3'UTR, and the MSH2 Boland inversion. Here, we present that with as little as 10ng of input and ~1000 reads per amplicon, this panel exhibits a uniformity of ~99% at 0.2X mean bias, little GC bias, a mapping rate of ~96%, and identification of over 95% of SNPs and indels. With the CleanPlex® Hereditary Cancer Panel V2 simple workflow, efficient and economical targeted NGS libraries to investigate inherited cancers can be produced within 3 hours, with high variant calling confidence and excellent coverage uniformity of genomic regions of interest.

## PgmNr 1774: Whole exome sequencing of clinical samples on the GenapSys NGS Platform and performance comparison to industry standard technology.

### Authors:

H. Esfandyarpour; M. Fallahi; N. Tangprasertchai; A. Barua; A. Thomas; B. Dong; S. Bagchi; H. Nezamfar; E. LoPrete; M. Jouzi; M.R. Barmi; S. Paliwal

[View Session](#) [Add to Schedule](#)

**Affiliation:** GenapSys Inc., Redwood City, California.

---

**Introduction** Advances in Next Generation Sequencing (NGS) technologies and NGS-based diagnostic applications have ushered in a new age of clinical testing. These clinical diagnostic applications include germline and somatic mutation testing, non-invasive prenatal testing etc. Clinical testing is typically performed at centralized laboratories, partly due to the large capital costs of NGS platforms and high run costs, which necessitates pooling of large numbers of samples. There is a strong need for decentralized NGS-based clinical testing, e.g. at hospitals, enabling cheaper and faster results for doctors and patients. The GenapSys NGS platform offers an accurate solution with low capital equipment and run costs, making it ideal for point-of-care testing. In this study, the GenapSys system performance was characterized by sequencing clinical samples from Novogene and compared to industry standard Illumina technology.

**Methods** GenapSys NGS libraries were generated from patient samples, and sequencing metrics, as well as germline and somatic mutation calling, were compared with Illumina sequencing. Genomic DNA was extracted by Novogene from patient FFPE and blood samples, and genomic libraries were generated following mechanical fragmentation, adapter ligation, size selection and PCR. Hybrid capture-based enrichment on GenapSys libraries was done using the IDT Exome Research panel, which targets a 39 Mb region (19,396 genes), to generate a Whole Exome Sequencing (WES) library. WES was performed for both FFPE and blood sample libraries on the GenapSys and Illumina systems, with typical coverage >100x. Sequencing reads were aligned to the hg38 reference genome using BWA-MEM. Variant calling analysis involved additional training of the Google DeepVariant model based on GenapSys sequencing data.

**Results** Comparison of GenapSys and Illumina sequencing demonstrated high concordance, with the F1 score of ~95% on SNV detection in high confidence regions. Average read length of GenapSys sequencing was > 125 bp, and the library on-target rate was > 85%.

**Conclusions** This study demonstrated that the GenapSys NGS platform generated comparable sequencing performance to the industry standard NGS technology. It identified germline and somatic mutations in clinically relevant WES libraries generated from patient FFPE and blood samples, thereby showing promise for use in decentralized clinical testing applications.



## PgmNr 1775: Compositional and functional analysis of the microbiome in tissue and saliva of oral squamous cell carcinoma.

### Authors:

Z. Zhang <sup>1,2</sup>; C. Wantao <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Oral Maxillofacial Head and Neck Oncology, Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, 200011, China; 2) Shanghai Key Laboratory of Stomatology & Shanghai Research Institute of Stomatology; National Clinical Research Center of Stomatology, Shanghai 200011, China

---

Oral squamous cell carcinoma (OSCC) is affected by the interaction between oral pathogen and holobionts, or the combination of the host and its microbial communities. Studies have indicated the structure and feature of the microbiome in OSCC tissue and saliva, the relationships between microbiota and OSCC sites, stages remain unclear. In the present study, OSCC tissue (T), saliva (S) and mouthwash (W) samples were collected from the same subjects, and carried out the microbiome study by 16S sequencing. The results showed the T group was significantly different from the S and W groups with the character of lower richness and diversity. Proteobacteria were most enriched in the T group at the phylum level, while Firmicutes were predominant in groups S and W. At the genus level, the predominant taxa of group T were *Acinetobacter* and *Fusobacterium*, and for group S and W, the predominant taxa were *Streptococcus* and *Prevotella*. The genera related to late stage tumours were *Acinetobacter* and *Fusobacterium*, suggesting microbiota may be implicated in OSCC developing. Both compositional and functional analyses indicated that microbes in tumour tissue were potential indicator for the initiation and development of OSCC.

## PgmNr 1776: Low allele fraction variant discovery via electrical detection of nucleotide incorporations.

### Authors:

J. Bell <sup>1</sup>; M. Kubit <sup>1</sup>; H.P. Ji <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanford Genome Technology Center, Palo Alto, CA.; 2) Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA

---

Among the problems in human genomics research is that of successfully discovering allelic variants that occur at low frequencies. Germline sequencing allows robust claims to be made for variation from a reference even at low sequencing depth, because the true proportions for an allelic variant are *a priori* known to be 50% or 100%. Tumor samples, however, owing to stromal contamination, heterogeneity, or polyploidy, may have allelic fractions ranging from nearly 0% to nearly 100%. The discovery of low-frequency variants is particularly important from the clinical perspective because if such variants occur in coding or regulatory DNA, they may be indications of subclonal elements of the tumor that can take over and cause recurrence, should the primary clone(s) be successfully treated. Thus, methods for confidently discovering and validating low-frequency variations are crucial for improving the quality of cancer genomics research, particularly because of the rising importance of liquid biopsies and the analysis of cell-free DNA.

We have begun testing a new DNA sequencer (Genapsys) that uses a sequencing-by-synthesis approach. This sequencing technology employs electrical detection of nucleotide incorporations. The solid state instrument detects a steady-state signal allowing for highly accurate sequence detection. Prior to loading on the sequencer, a library of DNA fragments is clonally amplified with DNA primers conjugated to beads. Amplified beads are loaded into a Complementary Metal-Oxide Semiconductor (CMOS) sequencing chip containing a scalable number of sensors. Nucleotides are injected one base at a time and incorporations are measured as electrical signals from each sensor that collect on the instrument's solid-state drive. The incorporation data files are transferred in real time to a secure cloud-hosted server and storage environment. At the end of the run, a FASTQ file becomes available.

Via a number of admixture experiments, we have tested the ability of this sequencer to produce data that allows detection of low-frequency variants in important genetic regions, ranging to a proportion of coverage of less than .1%. Even given very low coverage proportions, we have found that the Genius sequencer allows detection, giving results similar to a comparison ddPCR assay. As a low-footprint, highly sensitive sequencing technology, the use of solid state sequencing should expand the space for effective targeted testing of important variants.

## **PgmNr 1777: Simultaneous capture of short and long RNA subtypes from liquid biopsy substrates and FFPE tissue samples by a novel RNA-Seq library preparation method accelerates biomarker discovery.**

### **Authors:**

S. Okino <sup>1</sup>; A. Bibillo <sup>1</sup>; P. Patel <sup>1</sup>; Y. Wang <sup>1</sup>; L. Lingelbach <sup>2</sup>; Y. Jouvenot <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Life Science Group R&D, Bio-Rad Laboratories, Hercules, California.; 2) Life Science Group Marketing, Bio-Rad Laboratories, Hercules, California.

---

Technological advances in recent years have accelerated the study of biological markers as diagnostic, prognostic, or predictive indicators of a physiological state. Through the use of RNA sequencing (RNA-Seq) technology, researchers are querying the differential expression of messenger RNA (mRNA) transcripts and non-coding RNAs (lncRNA) as potential biomarkers of disease. One limitation that impedes discovery is the inability to query all RNA subtypes in a sample simultaneously. Most commercially available RNA-Seq library preparation methods are limited in their ability to capture only long or short RNAs. Here, we describe a new library preparation kit that utilizes a novel enzyme to capture all RNA subtypes at great sensitivity and independently of their size, enabling the simultaneous differential expression analysis of mRNA, lncRNA, miRNA, snoRNA, and other subtypes. Furthermore, we demonstrate the robustness of the enzyme and its ability to capture RNA from complex and highly degraded biological samples, such as liquid biopsies substrates and FFPE tissue samples.

## **PgmNr 1778: Anatomical and transcriptional characterization of breast tumor heterogeneity using spatial RNA sequencing.**

### **Authors:**

S. Ziraldo; C. Uytingco; J. Chew; M. Frey; Y. Yifeng; F. Meschi; E. Iyer; S. Giacomello; A. Jurek; J. Chell; S. Chatterjee; S. Williams; A. Hartnett; E. Borgstrom; N. Delaney; N. Weisenfeld; Z. Bent; J. Herschleb; T. Mikkelsen

[View Session](#) [Add to Schedule](#)

**Affiliation:** 10X Genomics, Pleasanton, CA

---

The tumor microenvironment is consisted of highly heterogeneous cellular components that interact and communicate with each other dynamically. Significant advancement in single-cell RNA sequencing allowed capturing thousands of cells and has revealed many subpopulations of cells in tumor tissues. However, dissociation of the tissue into single cells results in the loss of its important architectural information. The recently introduced spatial transcriptomics technology resolves localization of cells within a tissue section. Here we present an improved version of this spatial technology with increased tissue coverage, higher spatial resolution, and significantly improved sensitivity. We applied our improved technology to tumor tissue sections from human breast tumors and analyzed tissue-wide transcriptomics profiles to locate cancer related genes and infiltrating immune cells within spatial context and revealed intra-tumor heterogeneity within a tissue section. Elucidation of the spatial heterogeneity of tumor cells can shed light into understanding the disease states and progression, thus possibly aiding treatment decisions.

## **PgmNr 1779: An automated high-throughput custom single-cell targeted DNA sequencing platform to reveal rare clones and clonal evolution in cancer and characterize mutation profiles in CRISPR-edited cells.**

### **Authors:**

D. Mendoza; N. Li; A. Sciambi; M. Manivannan; J. Ho; B. Liu; K. Gokhale; K. Thompson; J. Marin; S. Chow; C. Curt; J. Wang; F. Carrerou; L. Nousheen; S. Sahu; S. Wang; S. Gulati; D. Kim; N. Vissa; N. Beard

[View Session](#) [Add to Schedule](#)

**Affiliation:** Mission Bio, Inc, South San Francisco, CA, USA

---

A great challenge in precision medicine is the understanding of cancer heterogeneity and clonal evolution, which has major implications in targeted therapy selection and disease monitoring. Single-cell DNA sequencing has the potential to provide unique insights on the cellular and genetic composition, drivers, and signatures of cancer at unparalleled sensitivity. Previously we have developed an automated high-throughput single-cell DNA analysis platform (Tapestri™) that leverages droplet microfluidics and a multiplex-PCR based targeted DNA sequencing approach. Here we present a new generation of Tapestri™ system which features a new instrument with simpler workflow and a cartridge with new microfluidic design, coupled with optimized and updated chemistry, resulting in drastically higher cell throughput, higher genomic coverage, better precision, and robust performance. The system now take input from 5,000 to 100,000 cells and generate genotypes of 1500 to 30000 cells in one run. The employment of AI (artificial intelligence) engine during panel design improves amplicon performance resulting in >90% panel uniformity and coverage of amplicons with high GC content (65-72%) across panels. Catalog panels such as AML, Myeloid, CLL and Tumor Hotspot are readily available for direct use or adaptation. We also have developed a web-based custom panel design portal for all human or mouse exome targets. Custom panels from 20-2000 amplicons can be easily built with a turn-around time of 4-6 weeks from design to validation and delivery. The new biochemistry allows maximum flexibility for panel optimization, enabling easy addition of new targets into existing panels for improved coverage or updated study design. The new analysis pipeline and software also employ AI-based algorithms resulting in better sensitivity and specificity for rare mutation detection and de novo discovery. We have validated the system with fresh, frozen and fixed cells, including cell lines, blood cancer samples, and solid tumor tissues, demonstrated detection of rare subclones of <0.1% prevalence, identified cancer mutation co-occurrence, and constructed clonal phylogenetic trees during disease progression and drug treatment. Additionally, the system enables unprecedented throughput and insights in mutational profiling of CRISPR-edited single cells, enables distinction of individual mutation zygosity and multi-mutational co-occurrence, further advancing research into cell therapies.

## PgmNr 1780: Spatial nucleus barcoding integrates cellular genomics and tissue architecture in normal and malignant breast tissues.

### Authors:

K. Wang <sup>1</sup>; M. Hu <sup>1</sup>; S. Bai <sup>1</sup>; M. Rabbani <sup>2</sup>; H. Chen <sup>3</sup>; B. Lim <sup>4</sup>; N. Navin <sup>1,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, UT MD Anderson Cancer Center, Houston, TX 77030, USA; 2) School of Health Professions, UT MD Anderson Cancer Center, Houston, TX 77030, USA; 3) Department of Pathology, UT MD Anderson Cancer Center, Houston, TX 77030, USA; 4) Department of Breast Medical Oncology, UT MD Anderson Cancer Center, Houston, TX 77030, USA; 5) Graduate School of Biological Sciences, UT MD Anderson Cancer Center, Houston, TX 77030, USA; 6) Department of Bioinformatics and Computational Biology, UT MD Anderson Cancer Center, Houston, TX 77030, USA

---

The expression programs of cell types in tissues and their spatial organization are critical for understanding homeostatic functions of normal tissues and the progression of diseases such as cancer. However, current single cell RNA-seq methods inherently lose all spatial information on cellular localization. To overcome this technical obstacle, we developed a spatial nucleus barcoding (SNUBAR) technology that delivers spatial barcodes into tissue regions, after which barcoded nuclei are used for high-throughput single nucleus RNA-seq. We validated this method by sample barcoding and intermixing experiments with cell lines, which showed a high efficiency of barcode delivery and doublets identification. We applied SNUBAR to 36 spatial regions of a normal breast tissue which identified 9 major cell types and 3 distinct spatial regions (fatty, fibroblast and epithelial). The fatty region was mainly composed of adipocytes that colocalized with a distinct lipo-fibroblast cell expression state, while the fibroblast region contained matrix-producing fibroblasts that colocalized with pro-angiogenic macrophages and vascular endothelial cells. The epithelial areas were composed of three distinct epithelial cell types that colocalized with lymphatic endothelial cells, and another fibroblast expression program. We further applied SNUBAR to 15 spatial regions of an invasive breast tumor, which identified 4 cell types in the microenvironment in addition to the tumor cells. While most immune and stromal cells were uniformly distributed across the 15 spatial regions, the tumor cells consisted of two distinct expression programs that were spatially segregated and harbored distinct copy number alterations. The two tumor clones exhibited differences in gene signatures associated with epithelial-to-mesenchymal transition, MYC targets, hypoxia and the expression of several cancer genes (eg. *VEGFA*, *AKT1*, *AKT2*). Additionally, we found two distinct macrophages expression states that colocalized with different tumor sub-clones. These data demonstrate the utility of applying SNUBAR to delineate the spatial organization of cell types and their expression programs in normal and malignant tissues to understand how spatial localization influences gene expression programs. This new technology is expected to have numerous applications in broad fields where the spatial organization of cell types is key to understand biological functions and the progression of diseased states.

## PgmNr 1781: A new sample extraction method for multiomics profiling and signaling pathway analysis.

### Authors:

S. Trusiak <sup>1</sup>; B. Patel <sup>2</sup>; S. Snovida <sup>2</sup>; E.I. Chen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Thermo Fisher Precision Medicine Science Center, Cambridge, MA.; 2) Thermo Fisher Scientific, Rockford IL

---

A multiomics approach to translational and clinical research aims to create more comprehensive molecular profiles to elucidate both molecular phenotype and mechanism of disease. One of the challenges in creating multiomic molecular profiles is the limited amount of relevant clinical material provided by FFPE and small surgical biopsy tissue. Often, one must choose between obtaining only genetic or proteomic information from the sample. Even when the sample source is plentiful, as with cells and large tissue, variation in cell populations or tissue subsections can create discordance between genetic and proteomic results when processed separately to extract nucleic acid and protein. We have developed a procedure to extract both NGS ready DNA and RNA and mass spec compatible peptides from the same FFPE section, cell pellet or tissue sample without the need to split the lysate. The entire workflow takes less than five hours. We applied our multiomic sample prep workflow to two applications; molecular profiling in FFPE cancer tissue and targeted signaling pathway analysis in human cells. DNA and RNA extracted from FFPE tissue were screened for clinically relevant DNA mutations and RNA fusions through targeted sequencing on the Ion Torrent GeneStudio S5 platform; these variants were then functionally verified by analyzing the extracted peptides on a Thermo Scientific Orbitrap Fusion Lumos mass spectrometer. Both mutation and fusion detection through the multiomic sample prep workflow was on-par with results from standard FFPE DNA/RNA extraction methods, but with the added benefit of verification of mutated protein presence. For signaling pathway analysis, we applied the multiomic extraction procedure to IGF treated human cancer cells to compare both RNA transcription and protein expression levels of key genes in the *AKT/mTOR* pathway. Using both targeted NGS and mass spec, we were able to determine the mechanism by which IGF treatment dysregulated the *AKT/mTOR* pathway within the cells. Our new sample extraction workflow enables genetic, transcriptomic and proteomic profiling from even limited sample sources while eliminating potential variabilities introduced by subsampling. This multiomics approach allows for DNA variant, RNA fusion and transcription detection, in addition to functional validation through proteomics.

## PgmNr 1782: Single-cell RNA sequencing of B16 melanoma model reveal the trajectory of distant metastasis of melanoma.

### Authors:

Y. Kim <sup>1,2,3</sup>; S. Shin <sup>1,2,3</sup>; M. Lee <sup>3</sup>; S. Jung <sup>2,4</sup>; S. Lee <sup>4,5,6</sup>; Y. Chung <sup>1,2,3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Microbiology, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 2) IRCGP, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 3) Precision Medicine Research Center, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 4) Cancer Evolution Research Center, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 5) Biomedicine & Health Sciences, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea; 6) Department of Pathology, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea

---

**Background:** Understanding the mechanism for distant metastasis is one of the fundamental issues in controlling metastatic melanoma. The purpose of our study was to elucidate the trajectory for distant metastasis of melanoma using single-cell RNA sequencing (scRNAseq) of B16 melanoma model.

**Methods:** We performed scRNAseq of murine melanoma cell line (B16F0) and its highly pulmonary metastatic variant (B16F10). The transcriptomic data of 2,884 and 1,827 cells were generated and analyzed from B16F0 and B16F10, respectively. The Seurat package was used to identify the subpopulation, differentially expressed genes between clusters, and cellular heterogeneity in genetic signatures. Copy number alternations between subpopulations were inferred by the InferCNV package. Pseudo-temporal ordering and RNA velocity analyses were performed by the Monocle package and the Velocyto package, respectively.

**Results:** Unbiased clustering of single-cell transcriptomes found two major populations according to the cell type, in which known cancer driver genes and additional putative genes were differentially expressed. Further finer clustering revealed five subpopulation, among which putative intermediate clusters with mingled B16F0 and B16F10 cells showed higher proliferative signature and lower invasive melanoma signature of melanoma, and putative highly metastatic cluster showed high expression of cancer driver genes. Inferred copy number alternation profiles suggested genetic aberrations between subpopulations. Pseudo-time trajectory and RNA velocity analyses supported the trajectory from putative intermediate subpopulation to putative highly metastatic subpopulation.

**Conclusions:** We identified transcriptional heterogeneity of melanoma cells during the metastasis progression using scRNAseq.



## PgmNr 1783: QIAseq FastSelect -rRNA and -Globin: One-step, rapid removal of rRNA and/or globin mRNA during whole transcriptome NGS library prep.

### Authors:

J. Shaffer <sup>1</sup>; D. Kim <sup>1</sup>; M. Hussong <sup>1</sup>; C. Howard <sup>1</sup>; N. Tolstrup <sup>2</sup>; S. Rulli <sup>1</sup>; E. Lader <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) QIAGEN Inc., Frederick, Maryland; 2) QIAGEN Nordic, Vedbæk, Denmark

---

Whole transcriptome next-generation sequencing (NGS) enables the characterization of both mRNA and long noncoding RNAs (lncRNAs) from biological samples. For optimal results, significantly overrepresented RNAs, such as ribosomal RNA (rRNA) and globin mRNA must be removed. Elimination of these contaminating RNAs simultaneously increases sensitivity and decreases the sequencing budget required per sample.

Various methodologies exist to deplete rRNA and globin, including hybridization/capture methods employed both as a sample pre-treatment or during post-library construction and methods which utilize enzymatic digestion in combination with target specific probes. However, these methods do not efficiently remove rRNA and globin, distort transcriptomic profiles, are ill-suited for damaged samples, are time consuming and potentially cause sample loss.

To remedy these shortcomings associated with rRNA and globin depletion, we have developed QIAseq FastSelect -rRNA HMR and -Globin. The FastSelect technology utilizes a novel, one-step rRNA depletion method that prevents cDNA synthesis of unwanted RNAs in cell, tissue, and whole blood RNA, whether the molecules are intact or degraded. Furthermore, FastSelect is compatible with virtually all stranded RNAseq library methods, and works with human, mouse and rat samples, as well as other model organisms. Custom kits are also available for any species or RNA.

Here we present examples of utilizing QIAseq FastSelect -rRNA HMR and -Globin with strand-specific RNAseq to analyze the whole transcriptome of matched normal and tumor lung cancer FFPE samples as well as whole blood samples. Such resulting differentially expressed RNA signatures are typically being utilized for pathway analysis, biomarker discovery, or for sample stratification. QIAseq FastSelect -rRNA HMR and -Globin rapidly eliminates rRNA and globin to enable the discovery of novel gene signatures. FastSelect is more efficient, tolerant of degraded samples, faster, and more cost effective than existing solutions.

## PgmNr 1784: Lotus DNA Library Prep Kit: A single flexible workflow with endless applications.

### Authors:

H. Huang; M. Hong; R. Lopez; A. Royall; K. Lai; L. Lewis; U. Das Chakravarty; Y. Zheng; S. Henck

[View Session](#) [Add to Schedule](#)

**Affiliation:** Integrated DNA Technologies, Redwood City, CA.

Recent advances in NGS technology provides unprecedented value to the scientific community in understanding how genomic diversities translate to complexities in biological processes. As NGS is increasingly adopted in clinical and diagnostic settings, high throughput laboratories would benefit from a streamlined workflow that seamlessly transitions from library preparation to target enrichment, and enables multiple applications using a single library prep kit. We present the Lotus DNA library preparation kit that employs a combination of enzymatic fragmentation and unique end repair mechanisms in a single-tube with efficient adapter ligation strategy to generate high-quality NGS libraries in approximately 2 hours. The flexibility in the chemistry allows the users to pair any TA-ligation compatible adapter (full-length or stubby) with a sample indexing strategy of their choice for PCR-free, PCR-amplified, and targeted sequencing applications on Illumina platforms. This library prep kit is compatible with multiple custom IDT NGS adapters, including TruSeq<sup>TM</sup>-Compatible Full-length Adapters and TruSeq<sup>TM</sup>-Compatible Stubby Adapter and Indexing Primers, as well as xGen suite of reagents for hybridization capture. This workflow can be used with bacterial or eukaryotic genomic DNA (gDNA) as well as FFPE DNA samples.

To assess the performance of the Lotus workflow, libraries were prepared using full-length adapters from a range of human gDNA and FFPE DNA inputs (1-250 ng). Whole genome sequencing (WGS) analysis showed highly uniform GC coverage and low bias from PCR-free and PCR-amplified samples. As exome sequencing of coding regions within the human genome reduces sequencing costs and increases sample throughput, we evaluated our workflow in targeted sequencing using the TruSeq<sup>TM</sup>-Compatible Stubby Adapter and Indexing Primers and xGen Lockdown Exome hybridization capture panel. We observed high on-target mapping and uniform sequence coverage across the entire human exome, including GC-rich exons. To evaluate sensitivity and specificity in low frequency somatic variant detection, libraries were generated using full-length xGen Dual Index UMI adapters from mixtures of genome in a bottle gDNA (NA12878 and NA24385) and captured using a 75 kb custom panel targeting single nucleotide and copy number variants. In summary, the Lotus workflow provides a fast and reliable strategy for preparing libraries to address a wide variety of applications.

## PgmNr 1785: Imputation of plasma metabolites using whole exome sequencing data.

### Authors:

N.A. Bihlmeyer<sup>1</sup>; L.C. Kwee<sup>1</sup>; W.E. Kraus<sup>2</sup>; C. Haynes<sup>1</sup>; S.H. Shah<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Duke Molecular Physiology Institute, Duke University, Durham, NORTH CAROLINA.; 2) Division of Cardiology, Department of Medicine, Duke University, Durham, NORTH CAROLINA.

---

**Introduction.** Metabolomic profiling is a powerful tool for identification of novel disease biomarkers and mechanisms. For example, we have described two discrete metabolic clusters that are associated with obesity, insulin resistance, type 2 diabetes and cardiovascular disease (CVD): branched chain amino acids and short-chain dicarboxylacylcarnitines. However, many large cohort studies do not have metabolomic profiling; or they may have metabolic data that do not include the metabolites of interest, but which otherwise would serve as important studies of these cardiometabolic diseases. Many of these cohorts, including the UK Biobank and Trans-Omics for Precision Medicine (TOPMed), have existing deep genomic data that could be leveraged to understand relationships between metabolites and disease. This study sought to evaluate using whole exome sequence (WES) data to impute plasma metabolite levels using metabolomic quantitative trait loci (mQTL).

**Results.** WES with 20X coverage and data on 66 targeted mass-spectrometry profiled amino acids and acylcarnitines were generated on individuals enrolled in the CATHGEN cardiovascular study. The cohort was split into a training set (N=3,441) and a validation set (N=860). Genetic variants were filtered based on minor allele count ( $\geq 10$ ) and significance of GWAS association with individual metabolite ( $p < 5e-5$ ); indels and non-autosomes were removed. Machine learning models controlling for age and sex fit in the training set included between 2 and 874 genetic variants; they were evaluated for efficiency in the validation set using two methods. (1)  $R^2$  between directly measured and imputed metabolites were measured and ranged from 0.01 to 0.38. (2) Associations between metabolites and Homeostatic Model Assessment of Insulin Resistance (HOMA-IR), were evaluated to see that the clinical value of these metabolites remained after imputation. For example, propionylcarnitine is strongly associated with HOMA-IR ( $p < 3e-30$ ) and had only a slight attenuation of effect size within the validation set after imputation (0.74 vs 0.24, measured vs imputed) indicating the imputed version of this metabolite has some value.

**Conclusions.** This method shows promise for extracting additional understanding of obesity and CVD by enabling the many large cohorts with genomic data on their participants to engage in metabolomic profiling studies. Further work will be needed to refine our machine learning models from these preliminary results.

## PgmNr 1786: A genome-scale Crispr screen identifies modulators of doxorubicin-induced cardiotoxicity.

**Authors:**

C. McDermott-Roe; W. Lv; K. Musunuru

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of Pennsylvania, Philadelphia, PA.

---

Doxorubicin is a potent chemotherapeutic used to treat various cancers. Despite efficacy, doxorubicin can cause severe and irreversible cardiotoxicity. We carried out a genome-scale CRISPR-Cas9 screen in HL1 mouse cardiomyocytes to identify genes which when disrupted, either protect against or worsen doxorubicin toxicity. Validating our approach, *Top2a* (the doxorubicin target) and *P-glycoprotein* (the doxorubicin efflux pump) were amongst the very top desensitizers and sensitizers, respectively. The screen also uncovered hundreds of other genes which significantly influenced toxicity as well as multiple over-represented gene sets. Our study provides an unbiased, forward genetics-based insight into doxorubicin-induced cardiotoxicity and a springboard to define novel toxicity mechanisms and susceptibility alleles.

## PgmNr 1787: Reducing the sample quality barriers between short and long read platforms to achieve genome comprehensiveness.

### Authors:

H. Doddapaneni <sup>1</sup>; H. Chao <sup>1</sup>; S.N. Jhangiani <sup>1</sup>; S.D. Kalra <sup>1</sup>; Q. Meng <sup>1</sup>; K. Walker <sup>1</sup>; S. Medhat <sup>1</sup>; C. Nessner <sup>1</sup>; Z. Momin <sup>1</sup>; Y. Han <sup>1</sup>; H. Santibanez <sup>1</sup>; V. Vee <sup>1</sup>; S. Lee <sup>1</sup>; M.L. Grove <sup>2</sup>; G. Metcalf <sup>1</sup>; F.J. Sedlazeck <sup>1</sup>; E. Boerwinkle <sup>1,2</sup>; R.A. Gibbs <sup>1</sup>; D. Muzny <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College Medicine, Houston, Texas.; 2) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX 77225

---

Comprehensive whole genome sequencing (WGS) requires more than standard Illumina short-read sequencing. We have tested alternative DNA sequencing platforms, including Pacific Biosciences (PacBio) or Oxford Nanopore (ONT) (long reads), 10x Genomics (short linked-reads) along with improved bioinformatics solutions, and demonstrated impressive phasing lengths (N50 of 67 Mbp). In a pilot study, we 'upgraded' 18 samples selected from a Cardio Vascular Disease (CVD) cohort set of 4,425 samples, based upon predicted Structural variations (SV) representation, by generating ~10x PacBio long read coverage of each. This demonstrated that long read data identifies new SVs and confirms SVs identified in short-read data and illustrated the ready need to establish production readiness for the new platforms.

Each method has particular sample requirements that must be balanced with performance attributes – for example, 10x Genomics platform is susceptible to low DNA quality but needs only ~1ng as library input. ONT provides the longest reads and PacBio platform generates CCS reads of highest read quality, however each require long molecular length DNA as for routine use. To introduce the procedures to a production environment, these DNA quality and DNA quantity demands must be met and balanced with the desired sequence output.

Using the 10x genomics platform for development, we improved both the DNA quality and quantity in the sample preparation workflow. For DNA QC, instead of the Pippin Pulse technology, that can QC 12 samples per run, Bio-Rad Pulse field gel electrophoresis was employed to QC 2x more samples and also provide better size resolution for >50 Kb DNA fragments. A follow-up electro-elution protocol for isolating DNA fragments >50Kb from fragmented DNA samples is under development. A new QC step to accesses the linked molecule length in 10x Chromium libraries through low coverage (~8x) sequencing is in production. These methods are also easily adaptable to PacBio and Nanopore platforms and will collectively reduce the DNA quality barriers between the short and long or linked read platforms and facilitate use of existing DNA for achieving genome comprehensiveness.

## **PgmNr 1788: Using genetic information for the reconstruction of metabolic processes underlying coronary heart disease: An example of the use of metabolomics in systems epidemiology of lipids.**

**Authors:**

F. Drenos

[View Session](#)[Add to Schedule](#)**Affiliation:** Brunel University London, London, United Kingdom

---

Lipids and the particles carrying them in circulation play a crucial role on the development and progression of changes leading to coronary artery disease (CAD). Studying the relationships between these particles and other CAD relevant molecular measures can provide valuable information for the metabolic processes involved and their functional organisation.

I used 230 metabolic measures from a targeted metabolomics approach together with genetic information from two samples of an established European cohort, totalling more than 9,400 individuals. I tested the phenotypic and genetic correlations between the metabolic measure identifying extensive correlations present and an almost identical correlation structure between the phenotypes and their underlying genetic factors. After Mendelian randomisation, I identified causal connections between 9,496 pairs of metabolic measures, though the direction of the effect was not always easy to infer.

Using the data to generate causal correlation networks showed that we could reconstruct well known metabolic processes using very little prior information. Even with no additional information, the causal correlations clearly showed the heterogeneity of function between high density lipoprotein (HDL) particles that can impact their role on CAD risk. I also found that the observed associations of lipoproteins levels with conjugated linoleic acid and glycolysis markers, such as glucose and citrate, are probably not due to an underlying biological mechanism. In contrast, the acute phase a1-acid glycoprotein is likely able to modify the larger triglyceride rich lipoproteins.

The results suggest that the use of genetics in the system epidemiology of lipids is possible but establishing the direction of effect will require additional information. The approach can improve our understanding of interactions between metabolic processes and, in this case, help us to elucidate the role of HDL particles in CAD.

## PgmNr 1789: Determination of expression level of miRNAs circulating plasma in patients with acute myocardial infarction (AMI) early onset.

### Authors:

H.J. Villalpando <sup>1</sup>; S.E. Flores-Martinez <sup>1</sup>; J. Sanchez-Corona <sup>1</sup>; R. Arreola-Torres <sup>2</sup>; I.-R. de Dios-Perez <sup>2</sup>; R. Rojas-Castillo <sup>2</sup>; M.A. Rosales-Reynoso <sup>1</sup>; L.A. Flores-Martinez <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Medicine, Center of biomedical investigation of the western, Phd; 2) Hospital de especialidades, Centro Medico de Occidente, IMSS, MD

### Introduction

MiRNAs are important players in cardiovascular physiology, as some balance control the differentiation and proliferation during cardiogenesis, and the aberrant expression of certain miRNAs has been related to cardiac diseases, including myocardial ischemia, because they act by limiting the expression of some proteins or increasing their expression according to the cellular requirement.

### Aim of the study

To determine the level of expression of miR1, miR16.2, miR21, miR107, miR130, miR133, miR143, miR155 and miR221 in patients with AMI and compare their expression in subjects without AMI.

### Material and methods

In 37 patients with AMI and 32 patients without AMI The extraction was done performed miRNAs circulating plasma with the exoRNeasy kit SerumPlasma. The synthesis of cDNA was done with the miScript kit II RT and the validation of expression of the 9 miRNAs by real-time PCR using the miScript matrix miRNA PCR Arrays. Comparative analysis of expression of miRNAs between patients and control is performed carried off by the algorithm  $2^{-DDCt}$  in RT<sup>2</sup>Profiler™ PCR Array Data Analysis (QIAGEN) software.

### Results

The average age of presentation of the AMI in the patients was 63.56 years. Of the patients studied, 81% were men, 70% had DM2 and 70% had hypertension. The comparison of the level of expression of miRNAs showed a differential expression profile in patients with AMI compared to the group without AMI, since a significant pattern of subexpression was observed in miR16 ( $p = 0.02$ ), miR21 ( $p = 0.023$ ) and miR155 ( $p = 0.0001$ ), with level greater than 2 times of change to the downside.

### Conclusions

The results suggest that there is an expression profile of miRNAs that distinguish patients with AMI

from healthy subjects, highlighting the possible involvement of miR16, miR21, and miR155 in the physio pathology of AMI in the Mexican population and its possible use as molecular biomarkers .



## PgmNr 1790: Genotype correlates with clinical severity in *PIK3CA*-associated lymphatic malformations.

### Authors:

J.T. Bennett <sup>3,6</sup>; C.V. Cheng <sup>1</sup>; D.M. Jensen <sup>3</sup>; A.E. Timms <sup>3</sup>; G. Shivaram <sup>4</sup>; R. Bly <sup>2</sup>; S. Ganti <sup>5</sup>; K. Whitlock <sup>5</sup>; W. Dobyns <sup>1,6</sup>; J. Perkins <sup>2</sup>; K.B. Zenner <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Integrative Brain Research, Seattle Childrens Research Institute; 2) Division of Pediatric Otolaryngology, Department of Head and Neck Surgery, University of Washington; 3) Center For Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute; 4) Department of Radiology, University of Washington; 5) Center for Clinical and Translational Research, Seattle Children's Research Institute; 6) Department of Pediatrics, Division Genetic Medicine, University of Washington

---

Lymphatic malformations (LMs) are congenital, non-neoplastic vascular malformations associated with post-zygotic activating *PIK3CA* mutations. The mutation spectrum within LMs is narrow, with the majority having one of three “hotspot” mutations. Despite this relative genetic homogeneity, clinical presentations differ dramatically. We used molecular inversion probes and droplet digital polymerase chain reaction to perform deep, targeted sequencing of *PIK3CA* in 271 affected and unaffected tissue samples from 81 individuals with isolated LMs and retrospectively collected clinical data. Pathogenic *PIK3CA* mutations were identified in affected LM tissue in 64 individuals (79%) with isolated LMs, with variant allele fractions (VAFs) ranging from 0.1 to 13%. Initial analyses revealed no correlation between VAF and phenotype variables. Recognizing that different mutations activate PI3K to varying degrees, we developed a metric to reflect differences in mutation strength, which we called the **genotype-adjusted VAF (GVAF)**, and found significantly higher GVAFs in LMs with more severe clinical characteristics including orofacial location or microcystic structure. In addition to providing insight into LM pathogenesis, we believe GVAF may have broad applicability for genotype-phenotype analyses in mosaic disorders.

## PgmNr 1791: Profiling of transcriptome and chromatin accessibility dynamics during iPSC-cardiovascular progenitor cell differentiation identifies novel cardiac fate-determining marker genes.

### Authors:

J. Nguyen<sup>1</sup>; W. Greenwald<sup>1</sup>; M. Donovan<sup>1,2</sup>; A. Antonio-Chronowska<sup>3</sup>; K. Frazer<sup>3,4</sup>; M. D'Antonio<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, 92093, USA; 2) Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, 92093, USA; 3) Institute of Genomic Medicine, University of California, San Diego, La Jolla, CA, 92093, USA; 4) Department of Pediatrics and Rady Children's Hospital, University of California, San Diego, La Jolla, CA, 92093, USA

---

The origin of many diseases are rooted in development, rather than in adults, underscoring the importance of identifying and characterizing genes driving cell differentiation. To identify genes underlying differentiation into cardiomyocytes (CMs), it is possible to leverage induced pluripotent stem cell (iPSCs) as there are known experimental protocols for differentiating iPSCs into CMs. However, while the importance of modulating WNT signaling is well established, the set of genes whose expression drives differentiation at each step of the protocol remain unknown. Thus, elucidating which genes are key drivers of the cardiac differentiation protocol could provide insight into cardiac disease etiology.

To identify the genes underlying CM differentiation, we characterized the dynamics of gene expression and chromatin accessibility throughout differentiation. We examined temporal changes in RNA-seq and ATAC-seq profiles at various stages of CM differentiations (day 0, 2, 5, 9, 15) in three iPSC lines (L1, L2, L3) from three subjects in triplicate (45 total samples). We observed a substantial divergence of global gene expression and chromatin accessibility at D5 and beyond in one line (L1) compared to the other two lines, suggesting the samples have different cell fates. We thus examined scRNA-seq data and found that while L1 was primarily comprised of epicardium-derived cells (EPDCs), L2 and L3 were primarily CMs. We leveraged these two different cell types to distinguish the driver from neutral genes that happen to have similar expression patterns. First, we identified genes with A-shaped spikes in their expression patterns at day 5, 9, or 15 in CMs and not in EPDCs. Next, we intersected these genes with the signature CM genes from the scRNA-seq, and utilized a machine learning approach to identify the genes at each time point which explained a large proportion of variation in final cell fate. This process resulted in the identification of 23 putative driver genes across the three time points. The genes from D9 and D15 were functionally enriched for cardiac muscle contraction and CM membrane potential regulation, respectively, and also included known cardiac trait-associated genes (*PRKAG2* and *FBN2*). Of note, 17 are eGenes regulated by cardiac eQTLs. Our results provide a list of driver genes that underlie key time points of cardiac cell differentiation and whose differential expression may contribute to the fetal origin of heart disease.

## **PgmNr 1792: Longitudinal multi-omics profiling in response to acute exercise in healthy and prediabetic individuals.**

### **Authors:**

S. Wu <sup>1</sup>; K. Contrepois <sup>1</sup>; K. Moneghetti <sup>2</sup>; D. Hornburg <sup>1</sup>; S. Ahadi <sup>1</sup>; F. Haddad <sup>2</sup>; M. Snyder <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Genetics, Stanford University, Stanford; 2) Stanford Cardiovascular Institute, Stanford

---

Exercise training has been widely recognized to have multiple benefits including cardiac effects and systemic factors. Insulin resistance (IR) is linked to the development of prediabetes and type 2 diabetes (T2D), all of which are further associated with increased risk for cardiovascular disease and altered immune function. Exercise training is also commonly recommended to control IR and T2D. However, the main molecular mechanism of its cardioprotection and prevention against T2D development is still poorly understood. In this context, we present an integrative multi-omics profiling study, including transcriptomics, proteomics, metabolomics, and lipidomics, on blood-based analytes in response to exercise in healthy and prediabetic individuals at a longitudinal fashion. The acute bout of exercise was associated with a wealth of bio-molecular changes (10,235 molecules at FDR < 0.05) spanning all omic layers. The integrative omic analysis revealed a complex network of molecular associations highlighting expected and novel connections related to energy metabolism, inflammation, oxidative stress, and coagulation/homeostasis. This longitudinal multi-omics study not only provides integrative insights into the mechanisms underlying exercise effects with molecular pathways, but also enables detection of personalized responses to exercise between healthy and prediabetic individuals, further facilitates precision exercise interventions against the development of prediabetes and T2D.

## PgmNr 1793: RNA-based mutation detection approaches identify a novel *PKP1* mutation with unusual splicing pattern in a family with ectodermal dysplasia-skin fragility syndrome.

### Authors:

A.H. Saeidian<sup>1,2</sup>; H. Vahidnezhad<sup>1</sup>; L. Youssefian<sup>1,2</sup>; M. Faghankhani<sup>1</sup>; S. Sotoudeh<sup>3</sup>; S. Zeinali<sup>4</sup>; J. Uitto<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dermatology, Thomas Jefferson University, Philadelphia, PA, USA; 2) Genetics, Genomics and Cancer Biology PhD Program, Thomas Jefferson University, Philadelphia, PA, USA; 3) Department of Dermatology, Children's Medical Center, Pediatric Center of Excellence, Tehran University of Medical Sciences, Tehran, Iran; 4) Biotechnology Research Center, Department of Molecular Medicine, Pasteur Institute of Iran, Tehran, Iran

---

Ectodermal dysplasia-skin fragility syndrome (ED-SFS), an ultra-rare autosomal recessive form of epidermolysis bullosa, manifests with blistering and erosions of the skin and palmoplantar keratoderma, variably associated with abnormal ectodermal development, including hypotrichosis, hypohidrosis, nail dystrophy, and abnormal dentition. Here, we examined a consanguineous family with two siblings with ED-SFS, for mutations by phenotype- and homozygosity mapping (HM)-assisted transcriptome profiling by RNA-Seq. HM reduced the number of candidate genes to 29, and transcriptome profiling identified *PKP1* with reduced expression. In addition, *PKP1* was the most downregulated gene among 94 candidate genes previously associated with the range of phenotypes observed in our cases. *PKP1*, encoding plakophilin 1, an integral component of the epidermal cell adhesion complexes, desmosomes, harbored a sequence variant at the end of the intron 10, *PKP1*: IVS10: c.1835-1G>C. As a result, the canonical acceptor splice site at the intron 10/exon 11 borders was inactivated, and instead, a cryptic splice site eight bp upstream within the exon 11 was activated. Sashimi plot revealed splicing aberrations, including retention of intron 13 sequences. Thus, our results support the hypothesis that whole-transcriptome sequencing assisted by HM and clinical phenotype can be utilized as a first-tier diagnostic approach for finding mutations in patients with Mendelian skin disorders.

## PgmNr 1794: Clinical, genetic and biochemical findings in a cohort of >1,000 patients with Gaucher disease.

### Authors:

C. Beetz; C. Cozma; L. Demuth; M. Werber; G. Oprea; V. Weckesser; A. Rolfs

[View Session](#) [Add to Schedule](#)

**Affiliation:** Centogene AG, Rostock, Germany

Gaucher disease (GD) is a lysosomal storage disorder that is caused by bi-allelic pathogenic variants in *GBA*. It may manifest with organomegaly, bone pain, thrombocytopenia and/or other symptoms, and is clinically very heterogeneous. Genotype-phenotype correlations have been suggested for the two outstandingly frequent mutations c.1226A>G (p.Asn409Ser, “N370S” in old nomenclature and representing a ‘milder’ allele) and c.1448T>C (p.Leu483Pro, “P444L”; ‘sever’ allele). While the activity of the *GBA*-encoded enzyme glucocerebrosidase has classically been determined as a biochemical diagnostic parameter, we recently suggested the sphingolipid Lyso-Gb1 to represent an even more sensitive and potentially dynamic biomarker. We now present Lyso-Gb1 determination as well as clinical and genetic findings in a cohort of >1,000 confirmed GD patients from 52 countries.

We identified a total of 157 pathogenic alleles, 68 (43%) of which are novel. Homozygosity was approximately as frequent as compound heterozygosity, and there were similar numbers of males and females. The clinical records were translated into 170 distinct HPO terms; there were an average of 4.1 terms per patient. Consistent with the cardinal symptoms of the disease, the most frequent terms were *thrombocytopenia* (in 53% of patients), *anemia* (48%), *splenomegaly* (41%), *hepatomegaly* (40%), *hepatosplenomegaly* (34%) and *bone pain* (34%). Plasma Lyso-Gb1 concentrations ranged from 18.7 to 1,600 ng/μl (pathological cutoff at >10 ng/μl), and showed an exponential distribution. Notably, Lyso-Gb1 values correlated with the numbers of HPO terms per patient, and were higher in more severely affected patients (e.g. *thrombocytopenia* + *excessive bleeding* vs. *thrombocytopenia*). Homozygous c.1226A>G patients had lower Lyso-Gb1 levels than homozygous c.1448C>T patients, while c.[1226A>G];[1448C>T] compound heterozygous patients were intermediate. Many additional genotypes differed significantly from each other as regards the associated Lyso-Gb1 values.

Our study widens the mutational spectrum in Gaucher disease, and reveals novel insights into clinical-genetic associations. It also characterizes Lyso-Gb1 as a highly valid biomarker for assessing disease severity and, potentially, monitoring of progression and response to therapy.

## PgmNr 1795: Characterization of heterogeneity in adipogenesis by single-cell RNA-seq.

### Authors:

J. Li <sup>1,2,3</sup>; S. Gustafsson <sup>4</sup>; A. Rao <sup>5</sup>; C. Park <sup>1,2,3</sup>; E. Ingelsson <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA; 2) Stanford Diabetes Research Center, Stanford University, Stanford, CA 94305, USA; 3) Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305, USA; 4) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 5) Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

---

Significant cell-to-cell variability in morphology and function has long been observed in differentiating adipocytes, which cannot be discerned by population-averaged measurements, such as bulk RNA-seq. It is also unclear to what extent the observed variability reflects the presence of phenotypically distinct subpopulations or asynchronous cellular responses to differentiation stimuli. To better characterize the heterogeneity in adipogenesis, we performed single-cell RNA-seq on human Simpson-Golabi-Behmel syndrome (SGBS) cells - a non-immortalized and non-transformed preadipocyte model - before and after 7 days of adipogenic differentiation. By using droplet-based single cell capture which enabled the gentle handling of large and rupture-prone adipocytes, we successfully mapped a single-cell profile including 5,662 cells from day 0 and 3,643 cells from day 7, validated by the correlated transcriptomic profiles of matching bulk samples. Next, by applying K-means clustering on the aggregated and normalized single-cell data of both timepoints, followed by t-SNE visualization, we defined 3 major clusters with distinctive gene expression patterns. Cells in Cluster 1 showed strong enrichment in genes featuring a preadipocyte phenotype, including those involved in proliferation (*TGFB1*, *CCND1*), extracellular matrix remodeling (*CLDN11*, *SERPINE1*) and negative regulation of adipogenesis (*GREM1*, *ID3*). Cells in Cluster 2 and 3 were enriched in genes positively associated with adipogenesis (*FABP4*, *ADH1B*, *CFD*), indicating a cell commitment to adipogenic trajectory. Among these, cells in Cluster 3 were further enriched in genes characteristic of a mature adipocyte function, such as adipokine synthesis (*ADIPOQ*), lipolysis (*PLIN1*, *LIPE*, *GOS2*, *PNPLA2*), fatty acid biosynthesis and desaturation (*FASN*, *SCD*, *FADS1*), representing a subpopulation of lipid-rich and metabolically active adipocytes further along into differentiation. Our single-cell transcriptomic catalog thus serves as a valuable resource to uncover novel molecular events underlying the intrapopulation heterogeneity both within and between timepoints of adipogenesis. Future studies combining single-cell profiling with genetic and metabolic perturbations in human primary (pre)adipocytes can pave a way for deeper understanding of subject-, depot-, and cell-specific regulation of adipocyte formation and function in health and disease.

## PgmNr 1796: Metagenomics of gallbladder stone and bile in Han Chinese in Taiwan.

### Authors:

H. Yang<sup>1</sup>; P. Yang<sup>2,3,4,5</sup>; C. Tsai<sup>2,4</sup>; C. Tsai<sup>2,4</sup>; T. Liu<sup>2,4</sup>; K. Ho<sup>2,4</sup>; J. Lee<sup>2,3,4</sup>; M. Li<sup>2,4</sup>; D. Chu<sup>2,4</sup>; F. Lee<sup>2,4</sup>; Y. Lien<sup>1</sup>; S. Shih<sup>1,3,4,6</sup>; Y. Lee<sup>1,7,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Research, Mackay Memorial Hospital, New Taipei City, Taiwan; 2) Department of Surgery, MacKay Memorial Hospital, Taipei, Taiwan.; 3) Department of Medicine, MacKay Medical College, New Taipei City, Taiwan; 4) MacKay Junior College of Medicine, Nursing and Management, Taipei, Taiwan; 5) Department of Pharmacology and Graduate Institute of Medical Sciences, Taipei Medical University, Taipei, Taiwan; 6) Division of Gastroenterology, Department of Internal Medicine, MacKay Memorial Hospital, Taipei, Taiwan; 7) Department of Pediatric Endocrinology, MacKay Children's Hospital, Taipei, Taiwan; 8) Institute of Biomedical Sciences, MacKay Medical College, New Taipei City, Taiwan; 9) Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

---

**Background and Goal:** Gallstone disease (GSD) confer as common gastrointestinal disease worldwide. Cholesterol-, mixed- and pigmented-stone are major types in GSD. This complex disorder most resulting from interactions between genetics and environmental factors. The incidence of GSD is about 5-14.3% in adult and correlated with age positively in Taiwan. It costs lots of clinical and surgical treatment. Genetic factors are also affecting in pathogenesis of gallstone. Many evidences for bacteria playing a role in the pathogenesis and formation of pigmented gallstones from human. However, the relationship between microbiota and GSD are not well discussed in Taiwan. In this study, we try to find the possible risk factors in bacteria affect in GSD.

**Methods:** We enrolled 78 GSD samples by laparoscopic cholecystectomy which include pigmented-bile, pigmented-stone, cholesterol-bile, and cholesterol-stone. Gallbladder bile and stone total DNA extracted by ZymoBIOMICS DNA miniprep kit. Bacteria V3, V4 region amplified by PCR and sequencing by Illumina Miseq. Sequence data trimmed, alignment and analysis by CLC Bio software.

**Results:** All the GSD sample in microbiota analysis merely bacteria-free. We found that microbiota diversity of pigmented-bile, -stone are far more than cholesterol-bile, -stone. For example, *Proteobacteria*, *Gammaproteobacteria*, *Enterobacteriales*, *Enterobacteriaceae*, and *Klebsiella*. The percentage of bacteria between male and female are different significantly under the same type of stone or bile. Such as *Firmicutes*, *Bacilli*, *Lactobacillales*, *Enterococcaceae*, *Enterococcus* are significant in male than female.

**Summary:** Our study provides a framework to study microbial profiling that predict a potential role in gallstone formation across different types of stones and patient backgrounds. We hope that will provide to clinical use in the future.



## PgmNr 1797: Identification of differentially expressed genes from kidney transplant recipients in Northern Ireland.

### Authors:

K. Kerr; L.J. Smyth; J. Kilner; A.P. Maxwell; H. McAneney; I. Young; B. McGuinness; F. Kee; A.J. McKnight; The Northern Ireland Cohort for the Longitudinal Ageing collaborative group.

[View Session](#) [Add to Schedule](#)

**Affiliation:** Centre for Public Health, Queens University Belfast, Belfast, United Kingdom

Belfast City Hospital is a kidney transplantation centre with the highest living donor transplant rate per million population in Europe. Kidney transplantation is a gold standard treatment for end stage renal disease, however there are significant complications associated with immunosuppression, such as an increased risk of diabetes and cancer. To date there have been limited transcriptomic analyses of patients following kidney transplantation. We aimed to identify differentially expressed genes (DEGs) and pathways associated with kidney disease and those that may be indicative of complications associated with kidney transplantation.

Targeted RNA sequencing was conducted using the Ion AmpliSeq™ Human Gene Expression Core Panel from peripheral blood mononuclear cells, to identify DEGs in kidney transplant recipients (n=37, median age = 54 yrs) compared to individuals with no evidence of renal disease (n=21 controls, median age = 56 yrs). In the kidney transplant group, 15 had a previous cancer diagnoses, one participant had unknown cancer status and three participants had new onset diabetes after transplantation (NODAT).

Analyses comparing transplant recipients to controls, identified 1089 significant DEGs with a 2 fold +/- change in expression ( $P \leq 0.1 \times 10^{-8}$ ). The pancreatic cancer pathway was significantly enriched with identification of six significant DEG in this pathway previously associated with several cancer types and/or diabetes (*RALA*, *JAK1*, *PIK3CA*, *MAPK8*, *JNK*, *SMAD2*). Outside of the pancreatic cancer pathway, *BAP1*, a gene previously associated with renal tumorigenesis, was identified to have significantly increased expression ( $p=7.3 \times 10^{-13}$ ) in kidney transplant recipients compared to controls. Gene ontology analysis identified enrichment in metabolic process, cellular process and biogenesis.

Our analyses have illustrated a significant enrichment in the pancreatic cancer pathway, as well as genes associated with several non-pancreatic cancers and diabetes, in kidney transplant patients approximately 20 years post-transplantation. RNA-seq and additional multi-omic analysis, in a larger case-control study, will help determine if exploration of a transplant recipient's molecular profile can predict risk of developing certain cancers and NODAT following transplant.



## PgmNr 1798: Expression profiling of allograft inflammatory factor-1(AIF-1) in association with cold/warm preservation time and the outcome of kidney graft function.

### Authors:

D. McDaniel <sup>1,2</sup>; J. Neill <sup>1,2</sup>; M. Roseburgh <sup>1,2</sup>; A. Bangale <sup>1,2</sup>; C. Anderson <sup>1,2</sup>; L. McDaniel <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Surgery, Univ Mississippi Med Ctr, Jackson, Mississippi.; 2) School of Medicine; 3) Department of Microbiology and Immunology

---

**Background:** Gene expression profiling of kidney with end stage disease (ESRD) revealed that allograft inflammatory factor-1 (AIF-1) was a highly expressed gene associated with pathologic glomerular nephritis at post allograft transplantation. The mRNA expression pattern of AIF-1 in kidney tissue and plasma protein were elevated in associated with rejection episodes. Our goal was to explore the significance of this genomic marker in early rejection vs. late rejection episodes in association with cold/warm preservation time that may trigger the release of inflammatory signals causing early host response of rejection. Extended cold or warm preservation time during organ procurement has been associated with increased risk of long term graft failure. To elucidate the early state of rejection and establish its therapeutic strategy, it is important to clarify the factors that are influential in the initiation of inflammation.

**Methods:** Clinical data were stratified, including delayed graft function (DGF), in which the patient required kidney dialysis after transplantation, stable graft function (SGF), or rejection episode (RE). The data were analyzed on the basis of cellular and antibody mediated rejection and time points (<6 months, 6-12 months, 13-36 months and >36 months) that rejection occurred. Blood samples from each patient at different time intervals (pre-transplantation, day 3 and day 6 after transplantation) were tested by q-RT-PCR. The impact of cold and warm preservation time were assessed.

**Results:** A total of 400 patients who received kidney transplant between 2007 and 2014 at the UMMC were subject of this study. A rate of 22/66 (5.6%) for rejection episodes within <6 months and 9/66 (2.3%) within >36 months were observed. Patients who received a kidney within a  $\leq 10$  hr. cold ischemia time (CIT) had superior clinical outcome and low AIF-1 expression, indicating that prolongation of CIT causes a delay in graft function. AIF-1 was expressed in podocytes and infiltrating inflammatory cells specifically increased in patient's samples that experienced cellular rejection within the < 6 month after transplantation. An undirected intersection network genes/protein analysis revealed interaction between AIF-1 and multiple immunologic signals involved in both inflammation and rejection.

**Conclusion:** The mRNA profiles of inflammatory markers observed in this study appeared to be predictive of early and long term allograft status after transplantation.

## PgmNr 1799: Applications of immuno-omics strategies to identify molecular features in infectious diseases caused by unknown pathogens.

### Authors:

T.M. Ko <sup>1,2,3,4</sup>; W.C. Chao <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan (R.O.C.); 2) Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsinchu, Taiwan (R.O.C.); 3) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan (R.O.C.); 4) Center For Intelligent Drug Systems and Smart Bio-devices (IDSB), National Chiao Tung University, Hsinchu, Taiwan (R.O.C.); 5) Taichung Veterans General Hospital, Taichung, Taiwan (R.O.C.)

---

Infection source clearance is the most effective way to prevent the progression of sepsis. However, identifying the source of infection in patients with sepsis is a major challenge. Although typical laboratory methods could expand certain microbes for further investigation, many microbes are still difficult to culture. We combined advanced omics platforms to obtain comprehensive data on the diversity of microbes and immunoglobulin (BCR receptor, BCR) that can interact directly with pathogens. We found that 8 patients with acute sepsis also expand the selected strain as in the general infection pattern. In addition, we also observed a dominant clonal expansion of these patients. The clonal expansion of IgM supported that although sepsis patients may not have sufficient immune defenses to combat these sources of infection, B cells in these patients still have the ability to screen for pathogens and expand effector B cells. Interestingly, we also found the consensus motifs of expanded immunoglobulin clonotypes between different patients, which implied that some patients with sepsis may be caused by the same source of infection. Our data suggested that patients with sepsis still have the ability to produce antibodies against pathogens, and applying immuno-omics strategies will be beneficial for identifying sources of infection.

## **PgmNr 1800: Somatic hypermutation (SHM) in the human heavy chain immunoglobulin receptor: Regions of low and high mutational burden.**

### **Authors:**

D. Hospodsky; M. Smit; C. Linkem; B. Howie; A. Sherwood

[View Session](#) [Add to Schedule](#)

**Affiliation:** R&D, Adaptive Biotechnologies, Seattle, WA.

---

The immune system is equipped to fight a multitude of antigens thanks to the diversity of its immune cell receptors. In the heavy chain of the B cell receptor, diversity stems from the antigen independent process called VDJ gene rearrangement, which occurs in Pro B cells. The heavy chain diversity is furthered through antigen-dependent processes during B cell maturation: Naïve B cells undergo maturation in the dark zone of the germinal center, where they are activated through antigens and proliferate into centroblasts. Single point mutations in the variable region are introduced at a high rate by activation-induced cytidine deaminase in this stage of maturation. Some centroblasts will have improved affinity to antigens due to introduced mutations in the variable region while other mutations lead to an inability to bind to antigens. The latter will undergo apoptosis. Centrocyte affinity is tested in the light zone of the germinal center through antigen exposure, which induces class switching. After class switching a subset of formerly naïve B cell leave the germinal center as memory B cells.

The process of SHM on a molecular level is not well understood. For example, it is unknown whether specific regions of the heavy chain are more or less prone to SHM. For this study, regions of high and low SHM were mapped through heavy chain DNA sequencing of naïve and memory B cells. In detail, cells from a leukopheresis blood product were magnetically sorted into memory and naïve B cells. Each pool of cells (naïve or memory) underwent RNA extraction. cDNA was synthesized from RNA using the SMARTer® PCR cDNA Synthesis Kit utilizing oligo dT primer on the 3'end and SMARTer oligonucleotide on the 5'end. Nested PCR resulted in amplicons of 530-550 basepair size, which were sequenced on Illumina DNA sequencing instruments in a paired-end approach. Naïve and memory B cells were sequenced from the same donors and DNA sequences in naïve and memory B cells compared to identify region of low versus high SHM activity.

## PgmNr 1801: Single-cell transcriptomics identifies effector capacity as a critical determinant of human CD4+ T cell responses to cytokines and activation.

### Authors:

E. Cano Gamez <sup>1,2</sup>; B. Soskic <sup>1,2</sup>; D. Plowman <sup>1,2</sup>; D. Wille <sup>3</sup>; N. Nakic <sup>4</sup>; J. Esparza-Gordillo <sup>5</sup>; D. Tough <sup>6</sup>; W. Rowan <sup>5</sup>; C. Larminie <sup>5</sup>; P.G. Bronson <sup>7</sup>; E. So <sup>8</sup>; T. Roumeliotis <sup>8</sup>; J. Choudhary <sup>8</sup>; G. Trynka <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK; 2) Open Targets, Wellcome Genome Campus, Cambridge, UK; 3) Biostatistics, GSK Medicines Research Centre, Stevenage, UK; 4) Functional Genomics R&D, GSK Medicines Research Centre, Stevenage, UK; 5) Human Genetics R&D, GSK Medicines Research Centre, Stevenage, UK; 6) Epigenetics DPU, Immuno-Inflammation and Oncology Therapy Area, GSK Medicines Research Centre, Stevenage, UK; 7) Statistical Genetics and Genetic Epidemiology, Biogen, Cambridge, MA, USA; 8) Functional Proteomics, The Institute of Cancer Research, London, UK

---

**Background:** Immune-mediated diseases are caused by the dysfunction of immune cells and the interactions between them. In particular, risk loci for immune-mediated diseases are enriched in genes specifically expressed in CD4+ T cells. However, CD4+ T cells are a heterogeneous cell population, with different subsets of naïve and effector memory cells. These effector subsets are shaped by the cytokine signals present in the environment. Thus, to disentangle the molecular mechanisms of disease it is crucial to understand how cytokines modulate T cell function.

**Methods:** We exposed naïve and memory CD4+ T cells from healthy individuals to different cytokine cocktails and profiled RNA (RNA-seq) and protein (LC-MS/MS) expression. For a subset of cytokines, we also profiled gene expression at the single-cell level (scRNA-seq).

**Results:** We observed that, in contrast to naïve T cells, memory cells did not respond to Th2-polarising cytokines and did not acquire a regulatory phenotype in response to TGF- $\beta$ . This suggested that naïve and memory cells activate different pathways in response to cytokines. We investigated this at the single-cell level and found that CD4+ T cells form a continuous spectrum, spanning from naïve to effector memory T cells. The position of cells in this spectrum reflects their effector capacity. We observed that the effector capacity of a T cell shapes its response to activation and to cytokines. For instance, Th17 cytokines induced the expression of IL9 and CCL17 only in the cells with higher effector capacity. In contrast, only cells with lower effector capacity responded to Th2-stimulation. In the context of T cell activation, cytokine and chemokine genes were also upregulated in a manner proportional to the effector capacity. To assess the disease relevance of these findings, we generated a time course of CD4+ T cell activation with single-cell resolution and integrated it with risk loci for immune diseases to identify the stages of activation most relevant to disease.

**Conclusions:** In this study we created a reference map of T cell responses to activation and cytokine stimulation with single-cell resolution. Our results shed light on the heterogeneity of CD4+ T cells and suggest that T cells respond to cytokines in a manner dictated by their effector capacity. This study

will serve as a valuable resource for studying the role of cytokines in human health and disease.

## PgmNr 1802: Adaptive immune receptor repertoire determination through probe capture of genomic DNA followed by next-generation sequencing.

### Authors:

Y.C. Lin <sup>1</sup>; P.L. Chen <sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan

---

The collection of B cell receptor (BCR) and T cell receptor (TCR) form the adaptive immune receptor repertoire (AIRR). The antigen binding sites of BCR and TCR are composed of variable (V), diversity (D) and joining (J) segments. When V(D)J recombination occurs, one segment is chosen from each V(D)J types in genomic DNA, which causes the extreme diversity of adaptive immunity. Human leukocyte antigen (HLA) genes have long been implicated in many diseases. *ATD*-induced agranulocytosis (TiA) was found associated with *HLA-B\*38:02* and *HLA-DRB1\*08:03*; anti-IFN- $\gamma$  autoantibodies in adults with disseminated nontuberculous mycobacterial (dNTM) infections were found associated with *HLA-DRB1\*16:02* and *HLA-DQB1\*05:02*. However, only a proportion of people carrying the risk HLA genotypes developed the phenotypes. Here we propose a hypothesis that the development of certain immune related diseases (such as TiA or dNTM) is determined by the combined effect of genomic profile of AIRR and HLA genes. In this study, we established a probe captured-based next-generation sequencing (NGS) method of genomic DNA to determine the AIRR profiles. We first designed probes for every known V alleles of BCR and TCR based on the ImMunoGeneTics (IMGT) database. Each probe was continuous 60 bps sequence by customized service from Roche/NimbleGen, and each V allele was covered by three probes. We verified this method by testing seven reference materials (RMs) from genome in a bottle (GIAB) (<https://jimb.stanford.edu/giab/>), as well as several Taiwanese participants with newly retrieved primary peripheral blood mononuclear cells (PBMCs). We built bioinformatics pipelines for analyzing the NGS data of genomic AIRR genes. In the future, we will apply this method to the genomic DNA samples from clinical TiA and dNTM patients. Through combined analysis of the AIRR and HLA genotypes, we will test the digenic/oligogenic hypothesis of genetic susceptibility of TiA and dNTM.

## PgmNr 1803: TotalSeq™ antibodies standardized antibody-oligo conjugates for CITE-Seq™, or single-cell proteogenomics.

### Authors:

A. Chowdhury<sup>1</sup>; M. Li<sup>1</sup>; B. Yeung<sup>1</sup>; K. Cardenas<sup>1</sup>; A. Rahman<sup>2</sup>; M. Tam<sup>1</sup>; K. Nazor<sup>1</sup>; X. Yang<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) BioLegend, Inc., San Diego, CA.; 2) Icahn School of Medicine at Mount Sinai, New York, NY

---

High-throughput single cell RNA-seq techniques have transformed our understanding of complex cell populations and processes, however, these don't allow additional phenotypic analysis of the same cells. Stoeckius *et al.* (Nature 2017) described the use of antibodies coupled with oligonucleotides to simultaneously study protein and RNA expression at the single cell level. This method, termed Cellular Indexing of Transcriptomes and Epitopes by sequencing (CITE-seq), combines highly multiplexed protein marker detection with unbiased transcriptome profiling and is compatible with scRNA-seq platforms based on a poly-A capture system. A collaboration between CITE-seq developers and BioLegend generated antibody-oligonucleotide conjugates under the brand name TotalSeq™.

As new technologies are developed, it is imperative to demonstrate that reagents deliver reproducible results, and that manufacturing complies with the highest standards possible. Here we demonstrate that our conjugates comply with BioLegend's quality standards, and the results are comparable with traditional flow cytometry. We also demonstrate the utility of CITE-seq using TotalSeq™ conjugates to improve cluster resolution as compared to samples analysed via RNA-seq alone. We also show that the technology can produce equivalent data when compared to CyTOF. Finally we show how TotalSeq™ "Hashtags" are used to efficiently multiplex single cell samples, analyse multiple TotalSeq antibody dilutions simultaneously, or optimize the CITE-seq protocol and reagents.

In summary, we establish the utility of analysing proteins and transcriptomic in the same cell in a highly multiplexed fashion, and show a completely new way to approach biomedical research.

## PgmNr 1804: Unique RNA profiles in different fractions of plasma.

### Authors:

J. Jia; L. Wang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pathology, Medical college of Wisconsin, Milwaukee, WI.

---

Circulating extracellular RNAs (exRNAs) have great potential to serve as biomarkers for a wide range of diagnostic, therapeutic and prognostic medical conditions. Current research has been focusing on the characteristics of the RNA content in extracellular vesicles and its delivery process in various human biofluids. However, exRNAs are also present in other plasma fractions including cell debris, platelets, nucleosomes or free circulating condition. So far, knowledge of the difference or association between these different sources of exRNAs is very limited. To address this issue, we collected 10 plasma samples and performed a sequential physical and biochemical precipitation to receive 4 plasma fractions (cell debris, Thrombin-associated fraction, exosomes and supernatant) from each sample. From total RNAs of 40 fractionated plasma fractions, we made ligation-free libraries and performed RNA-seq in an Illumina HiSeq sequencer. To evaluate full spectrum RNA abundance, we included all RNAs during library preparation without size selection. Due to complicated RNA composition in these libraries, we utilized a successive stepwise alignment strategy to map the RNA sequences to different RNA categories, including mature miRNAs, premature miRNAs, piwi-interacting RNAs, tRNAs, other types of ncRNAs, cDNAs, circRNAs, and exogenous RNAs. We found that each plasma fraction had its own unique distribution of RNA species. Hierarchical cluster analysis showed similarity in samples with the same fraction and significant differences between different fractions. As expected, exosome fraction showed the highest percentage (~4.9% of all mappable reads) of mature miRNAs with average detectable miRNA=398, while premature miRNAs, other types of ncRNAs and cDNAs were predominant in cell debris fraction with the percentage of all mappable reads at ~1.3%, ~61.9%, and ~8.1%, respectively. Interestingly, the supernatant fraction showed the lowest percentage of most small RNAs but relatively higher percentage of premature miRNAs and circRNAs (~2.1% and ~12.6% respectively). In addition, we also observed abundance difference in each unique transcripts. miR-1246 ranked the highest abundance in exosome fraction while miR-3123 ranked the highest in the cell debris fraction. Furthermore, we observed a considerable proportion of exogenous RNAs in all plasma fractions. These results demonstrate that thorough inspection of all plasma fractions is necessary for exRNA-based biomarker study.



## PgmNr 1805: Validating age-associated sex-specific DNA methylation sites in a Bangladeshi population.

### Authors:

R.J. Jansen <sup>1,2,3</sup>; L. Tong <sup>4</sup>; M. Argos <sup>5</sup>; F. Jasmine <sup>4</sup>; M.G. Kibriya <sup>4</sup>; H. Ahsan <sup>4,6,7</sup>; B.L. Pierce <sup>4,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Public Health, North Dakota State Univ, Fargo, North Dakota, USA.; 2) Genomics and Bioinformatics Program, North Dakota State University, Fargo, ND, USA; 3) Biostatistics Core Facility, North Dakota State University, Fargo, ND, USA; 4) Department of Public Health Sciences, The University of Chicago, Chicago, Illinois, USA; 5) Division of Epidemiology and Biostatistics, School of Public Health, University of Illinois at Chicago, Chicago, Illinois, USA; 6) Department of Medicine, The University of Chicago, Chicago, Illinois, USA; 7) Department of Human Genetics and Comprehensive Cancer Center, The University of Chicago, Chicago, Illinois, USA

---

For epigenome-wide association analyses, replication of CpG associations with human traits is critical for identifying robust evidence for associations. In this study, we used an independent validation set consisting of 399 Bangladeshi individuals (167 males) participating in the Health Effects of Arsenic Longitudinal Study (HEALS) to replicate age-associated CpGs identified in a previous study of 400 (189 males) individuals from the Bangladesh Vitamin E and Selenium Trial (BEST). Prevalence of never smoking was 90% and 48% among females and 26% and 4% among males in HEALS and BEST, respectively. All BEST but no HEALS participants had arsenical skin lesions. The mean age in HEALS was 41 (SD=10) for males and 34 (SD=9) for females compared to 43 (SD=9) and 44 (SD=11), respectively, in the BEST sample. DNA was extracted from whole blood in both studies. The Illumina HumanMethylation 450K Beadchip was used for BEST and the Illumina EPIC array (850K) Beadchip was used for HEALS which meant 47780/423604 (11%) 450K CpGs were not present on the 850K chip. Using the HEALS validation set, we conducted a genome-wide search for age-associated CpG sites (among 375,824 overlapping sites) using a reference-free approach for cell type proportion adjustment (R package: RefFreeEWAS) and model adjustment for ever/never smoked. Among the 3,294 overlapping age-associated CpGs ( $P < 5 \times 10^{-8}$ ) observed among females in BEST, 1162 (35% at  $P < 5 \times 10^{-8}$ ) were also associated with age in HEALS in a consistent direction. Likewise for males, among the 716 age-associated CpGs ( $P < 5 \times 10^{-8}$ ) observed in BEST, 316 (44% at  $P < 5 \times 10^{-8}$ ) were also associated with age in HEALS in a consistent direction. Similar to the BEST analysis, in HEALS there is a higher proportion of hypomethylation across individual genome regions (shore, shelf, open sea) and higher proportion of hypermethylation (islands) based on Fisher exact tests among the age-associated CpGs for both sexes. Again consistent with the BEST analysis, hypermethylation enrichment among the age-associated CpGs across individual relationship to gene categories (i.e., TSS1500, TSS200, 5'UTR, First Exon, Body, 3'UTR, and Intergenic) compared to all CpGs tested, only intergenic was not significant. Our results indicate that age-associated methylation tends to be highly consistent across both study samples for both sexes even with different sample age distributions, smoking status, and lesion status.

## PgmNr 1806: The utility of WGS-derived telomere length in population health studies: A proof-of-concept using consumer wearable sleep metrics.

### Authors:

J.X. Teo<sup>1</sup>; S. Davila<sup>1,2</sup>; C. Yang<sup>3</sup>; C.J. Pua<sup>3</sup>; J. Yap<sup>4</sup>; S.Y. Tan<sup>4</sup>; A. Sahlén<sup>4,5</sup>; C.W.L Chin<sup>4</sup>; B.T. Teh<sup>1,6,7,8,9</sup>; S.G. Rozen<sup>1,6,10</sup>; S.A. Cook<sup>1,2,3,11,12</sup>; K.K. Yeo<sup>4</sup>; P. Tan<sup>1,6,9,13</sup>; W.K. Lim<sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) SingHealth Duke-NUS Institute of Precision Medicine, Singapore, Singapore; 2) Cardiovascular and Metabolic Disorders Program, Duke-NUS Medical School, Singapore; 3) National Heart Research Institute Singapore, National Heart Centre Singapore, Singapore; 4) Department of Cardiology, National Heart Centre Singapore, Singapore; 5) Department of Medicine, Karolinska Institutet, Sweden; 6) Cancer and Stem Biology Program, Duke-NUS Medical School, Singapore; 7) Laboratory of Cancer Epigenome, Division of Medical Sciences, National Cancer Centre Singapore, Singapore; 8) Institute of Molecular and Cell Biology, Agency for Science, Technology and Research, Singapore; 9) Cancer Science Institute of Singapore, National University of Singapore, Singapore; 10) Centre for Computational Biology, Duke-NUS Medical School, Singapore; 11) National Heart and Lung Institute, Imperial College London, United Kingdom; 12) MRC Clinical Sciences Centre, Imperial College London, United Kingdom; 13) Biomedical Research Council, Agency for Science, Technology and Research, Singapore

---

Premature telomere attrition is associated with adverse health outcomes, including cardiovascular disease and all-cause mortality. The measurement of leukocyte telomere length (LTL) in population health studies have in the past been carried out using quantitative PCR. However, the growing adoption of whole-genome sequencing (WGS) in population health studies presents an opportunity to estimate LTL using WGS data. To validate the utility of WGS-derived LTL in population health studies, we used a computational tool called telomerecat to estimate LTL in 175 healthy volunteers. We first showed that WGS-derived LTL is significantly correlated with both qPCR-derived LTL as well as volunteer chronological age. We then examined the relationship between WGS-derived LTL and habitual sleep duration obtained from consumer-grade wearables. There was a significant positive association between habitual sleep duration and WGS-derived LTL after adjusting for age, gender and ethnicity. Volunteers that had sufficient sleep (> 7 hrs) had significantly longer WGS-derived LTL compared to those that had insufficient sleep (< 5 hrs). We were able to replicate these associations in a separate validation cohort of 300 individuals, for whom LTL was measured using qPCR. Our findings demonstrate the utility of WGS data in population health studies to estimate LTL, and thus provide insights into lifestyle and clinical factors associated with biological aging.

## **PgmNr 1807: Isolation and characterization of exosomes from mixed cortical neurons derived from Induced Pluripotent Stem Cells (iPSC) from patients with type 2 Gaucher disease and controls.**

### **Authors:**

C. McKinney <sup>1</sup>; B. Behre <sup>1</sup>; K. Baumgarner <sup>2</sup>; E. Sidransky <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Molecular Neurogenetics Section, NIH/NHGRI, State College, PA.; 2) Genetics, Via College of Osteopathic Medicine and Gibbs Research Institute, Spartanburg, SC

---

Exosomes, intracellular vesicles ranging in size between 30-100 nm, originate from many cell types. Originally, thought to contain cellular debris; they are now recognized to contain proteins/RNAs captured from the cytoplasm. Exosomes transduce cell-to-cell signals to adjacent cells and to distant cells by entering the bloodstream and being captured in other locations. Although their physiologic roles remain unclear, exosomes function in intracellular communications and, perhaps, deliver critical proteins to other cellular locations. There is increasing evidence that delivered exosome cargo may mediate neurodegenerative disease progression.

Type 2 Gaucher disease (GD2), a lysosomal storage disorder, has a severe neurodegenerative course resulting in early death. We used patient and control neonatal fibroblasts to derive iPSC lines that were then used to generate neural progenitor cells (NPCs) and then differentiated to mixed cortical neurons. The NPCs stain with Neurofluor Cdr3, a fluorescent probe for early neuron differentiation. The mixed cortical neurons stain with Neurofluor NeuO, a live stain that detects neurons. Both stains identified >90% of the cells in a 10 cm plate. Further characterization of the cortical neurons by PCR showed expression of MAP2, NeuN, Nestin, NSE and ChaT. Some cultures also showed low expression of TH (dopaminergic) and GFAP (astrocytes).

Neurons were grown in serum free Brainphysy medium for >30 days. Exosomes were isolated from culture plate medium collected over 48-72 h. Two methods were used to capture exosomes, ExoQuick and Captur'em columns. Nanosight analyses showed a peak of vesicles in the 50-200 nm size range. The best yields were from ExoQuick preparations. Western blots identified flotillin-1, a marker protein for exosomes. Dot blots showed other exosome markers. GCase protein and a potential tetrameric form of synuclein were found in both control and GD2 samples.

The exosomes will be analyzed by transmission electron microscopy and be evaluated for additional neuronal markers. Proteome analysis will be performed on exosome cargo from controls vs. GD2. Transcriptome analysis for miRNAs will be evaluated. It has been suggested that since exosomes circulate in the blood and cross the blood-brain barrier, they may represent a therapeutic modality to carry drugs to neurons. Our exosomes will be labelled fluorescently and tested for entry into both control and GD2 neural cells.

## **PgmNr 1808: Connect the dots: A method for interpreting multi-metabolite perturbations for the diagnosis of metabolic disease using disease-specific metabolomics networks.**

### **Authors:**

L.R. Thistlethwaite <sup>1</sup>; L.C. Burrage <sup>1</sup>; K. Riehle <sup>1</sup>; F. Scaglia <sup>1</sup>; M.J. Miller <sup>2</sup>; S.H. Elsea <sup>1</sup>; A. Milosavljevic <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Baylor College of Medicine, HOUSTON, TX.; 2) Indiana University School of Medicine, INDIANAPOLIS, IN.

---

Untargeted metabolomics profiling captures a global view of the metabolic state of an individual, and diagnosis using untargeted metabolomics is data-driven and unbiased. One prominent practice in diagnosing Inborn Errors of Metabolism (IEM) is to follow a targeted approach, where a diagnosis is nominated based on clinical symptoms observed, and targeted biochemical testing is performed to either confirm or rule out the nominated diagnosis. Problems with this targeted approach to diagnosis occur when clinical symptoms are undifferentiated across multiple IEM. In these circumstances, untargeted metabolomics profiling has shown to be useful. The challenge with diagnosis using untargeted profiling, however, is quantifying the significance of multi-metabolite perturbations observed in individual profiles using a standardized, automatic and transparent methodology. Currently, interpretation of untargeted metabolomics data is performed manually, where clinicians use their biochemical knowledge to reason about the patterns observed. In effect, metabolite perturbations can be examined by superimposing those perturbations onto biochemical pathway maps, and searching for bottlenecks where metabolite perturbations around an affected enzyme show opposite directionality.

To improve accuracy and transparency and to automate this subjective and manual diagnostic process, we develop CTD, a computational diagnostic method that “**connects the dots**” between metabolite perturbations observed in patient’s metabolomics profiles with patterns observed in disease-specific networks. In this work, we show that disease-specific metabolomics perturbation networks lead to more accurate detection of disease-specific metabolite perturbation patterns when compared to general biochemical pathways. Furthermore, using disease networks for 16 different IEM (e.g., 4 organic acidemia disorders, 2 Zellweger spectrum disorders, 4 urea cycle disorders, 3 nucleotide metabolism disorders, and 3 others), we show that CTD reproduces expert performance in diagnosing IEM disorders.

## PgmNr 1809: Diagnosis of guanidinoacetate methyltransferase (GAMT) deficiency by untargeted metabolomic analysis in plasma.

### Authors:

L. Lizcano-Gil <sup>1</sup>; K. Pappan <sup>2</sup>; D. Toal <sup>2</sup>; B. Wittmann <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurogenética Genética Humana, Bogotá, Colombia; 2) Metabolon Inc, Durham, NC, USA

---

**BACKGROUND:** Guanidinoacetate methyltransferase (GAMT) deficiency causes neurological and muscular alterations, which, being nonspecific and of varying severity, make clinical diagnosis difficult.

**METHODS:** To describe the usefulness of metabolomic analysis in cases of patients with suspected neurometabolic disease, we present the case of a 6 years old patient, with symptoms of ataxia and epilepsy. Features of the autistic spectrum were observed, seizure control had not been achieved, and there was no improvement of his neurodevelopmental process through intervention by therapies. Previously, Angelman syndrome and other chromosomal conditions were ruled out, using FISH, methylation testing and arrCGH. Plasma amino acid profile, urine, qualitative organic acids, and brain MRI spectroscopy were normal. Given the possibility of neurometabolic disease, analysis by non-targeted metabolomics in plasma was performed.

**RESULTS:** We documented reduced creatinine levels and extremely elevated levels of guanidinoacetate, a pattern consistent with a blockage of the creatine synthesis pathway. These findings reflect the deficiency of GAMT. There was also a decrease in molecules in different metabolic pathways that have, in common, pyridoxal phosphate-dependent reactions. Several metabolites of the glutamate and DOPA pathways were decreased. All those changes can contribute to the phenotype. Sequence analysis of *GAMT* gene revealed two variants: Probable pathogenic variant c.64del and a variant of uncertain significance, c.521G>T. Three bioinformatics predictors estimate that the resulting changes in the protein would be pathogenic. We designed a treatment with creatine, restriction of arginine, L-Ornithine, pyridoxine and increase in the dietary contribution of taurine. Crisis control was achieved, normalization of the walking pattern and noticeable improvement in the interaction with the environment and with third parties were observed.

**DISCUSSION:** It is demonstrated that metabolomic analysis is useful for the diagnosis of GAMT deficiency and that also it is a very valuable tool for the study of patients with suspected neurometabolic disease.

## PgmNr 1810: Untargeted metabolomic analysis in plasma for the diagnosis of 3-methylglutaconic aciduria (MGA).

### Authors:

S. Ospina Lagos <sup>1</sup>; L. Lizcano-Gil <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, Genética Humana, Bogotá, DC, Colombia; 2) Neurogenetics, Genética Humana, Bogotá, DC, Colombia

---

*Introduction:* MGA includes a heterogeneous group of metabolic disorders which main biochemical characteristic is the increased levels of 3-Methylglutaconic Acid. Up to date, seven of MGA are recognized based on the gene involved.

*Objective:* To describe the usefulness of metabolomic analysis in cases of patients with suspected neurometabolic disease, but with nonspecific manifestations and normal DNA analysis.

*Materials and methods:* We present the case of a 9-year-old Colombian patient, who was referred to the neurogenetic clinic with cognitive disability, global neurodevelopment delay, hypotonia and complex seizures. Previously, clinical exome sequence was done with normal results (there were no pathogenic variants or variants of uncertain significance), plasma and urine amino acids profiles were also normal. Given the possibility of neurometabolic disease, analysis by non-targeted metabolomics in plasma was performed, which demonstrated a biochemical signature consistent with the diagnosis of MGA, mainly, characterized by elevation of 3-methylglutaconate and 3-methylglutaryl carnitine. We performed sequence and deletion/duplication analysis for genes AUH, TAZ, SERAC1, OPA3, TMEM70 and DNAJC19, with normal results. Based on these findings, diagnosis of 3-methylglutaconic aciduria type 4 was established (NOS 3-MGA-uria).

*Conclusion:* It is demonstrated that metabolomic analysis is useful for the diagnosis of 3-methylglutaconic aciduria and that it also is a very valuable tool for the study of patients with suspected neurometabolic disease.

## PgmNr 1811: Deciphering the role of non-coding genome in mitochondrial function: From long non-coding RNAs to hidden micropeptides.

### Authors:

M. Gusic<sup>1,2</sup>; V.A. Yépez<sup>3,4</sup>; R. Kopajtich<sup>1,2</sup>; C. Mertes<sup>3</sup>; P. Goldberg<sup>3</sup>; C. Ludwig<sup>5</sup>; J. Gagneur<sup>3,4</sup>; H. Prokisch<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany; 2) Institute of Human Genetics, Technical University of Munich, Munich, Germany; 3) Department of Informatics, Technical University of Munich, Garching, Germany; 4) Quantitative Biosciences Munich, Department of Biochemistry, Ludwig-Maximilians Universität, Munich, Germany; 5) Bavarian Center for Biomolecular Mass Spectrometry (BayBioMS), Technical University of Munich, Freising

---

Advances in computational biology and large-scale OMICS approaches reveal that a larger fraction of the genome than previously recognized is transcribed and translated. Most of these novel transcripts are long-noncoding RNAs (lncRNAs). Though mostly functionally uncharacterized, they are emerging as important players in cellular processes during normal development and physiology, but also disease. Noteworthy, genes coding for small peptides (<100 amino acids in length) have been frequently missannotated as lncRNA genes due to their small open reading frame, overlooked by automatic gene annotation. Micropeptides are known to have important roles in fundamental biological processes. In addition, the mammalian mitochondrial proteome is surprisingly enriched in micropeptides, accounting for 5% of its proteins. To gain insight into the role of lncRNAs in mitochondria, we have integrated GTEx data, performed RNA-Sequencing (RNA-seq) and measured oxygen consumption rate of >300 fibroblasts derived from mitochondrial disorder patients. We performed logistic regression on samples from GTEx dataset using MITOCARTA genes as the positive class and all other protein coding genes as the negative class, correlated gene expression of mitochondrial-disease genes and lncRNAs, and correlated lncRNA expression and cellular respiration. By performing these three strategies, we were able to identify several lncRNAs as new candidate genes to be associated with mitochondria. To explore whether these lncRNAs encode micropeptides with a possible role in mitochondria, we analysed their coding potential by looking at possible open reading frames, evolutionary conserved regions and RNA-seq and ribosomal profiling data. As additional supportive evidence of a novel micropeptide discovery, we have applied mass spectrometry (MS)-proteomics on patient-derived fibroblasts. Next to computational approaches, knockout studies are currently ongoing in order to elucidate the function and biological relevance of these genes and overall highlight the importance of small peptides in cellular processes.



## PgmNr 1812: Transcriptional changes in circulating immune cells of patients with Gulf War Illness.

### Authors:

L. Nathanson <sup>1</sup>; D. Van Booven <sup>2</sup>; A.K. Cheema <sup>1</sup>; L. Sarria <sup>1</sup>; V. Beljanski <sup>3</sup>; M.A. Fletcher <sup>1,4</sup>; N.G. Klimas <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Neuro Immune Medicine, Dr. Kiran C. Patel College of Osteopathic Medicine, Nova Southeastern University, Fort Lauderdale, Florida.; 2) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida; 3) Dr. Kiran C. Patel College of Allopathic Medicine, Nova Southeastern University, Fort Lauderdale, Florida; 4) Miami Veterans Affairs Healthcare System, Miami, Florida

---

**Background:** Gulf War Illness (GWI) is a severely debilitating condition with a cluster of chronic symptoms affecting up to 30% of the veterans deployed to the 1990-91 Persian Gulf War. The symptoms associated with GWI include fatigue, headaches, musculoskeletal/joint pain and disturbances of multiple organ systems. The causes and mechanisms of GWI are unknown, although the spectrum of symptoms and organ systems involved suggest that GWI is most likely caused by a combination of genetic predisposition and its interaction with environmental factors such as exposure to toxic chemicals and/or pathogens.

**Objectives:** The main objective of this research proposal is to identify novel mechanisms of transcriptional regulation in GWI, which will help to better understand GWI pathobiology.

**Methods:** We used RNA-seq to evaluate gene expression and alternative splicing in the peripheral blood mononuclear cells (PBMC) of 20 GWI patients and 20 matched healthy controls at T0 (baseline), T1 (VO2max, exercise challenge) and T2 (4 hours of rest after T1). After alignment using STAR and counting exons using HTSeq, we used DESeq2 to evaluate differences in gene expression and DEXSeq to evaluate differences in splicing. We compared the gene expression and splicing events between individuals with GWI and healthy controls at each time point, as well as difference in responses to the exercise challenge (between time points for the same group). Results of RNA-seq analysis were validated using NanoString technology.

**Results:** An increased abundance of differentially expressed genes related to immune disorders, small GTPase mediated signal transduction, protein binding and nucleotide binding was observed. Additionally, the observed changes in gene expression and alternative splicing between time points (in response to exercise challenge) were smaller in individuals with GWI as compared to healthy controls.

**Conclusion:** Our findings provide novel insights into the molecular changes associated with GWI onset and progression enhancing the current understanding of the pathophysiological mechanisms of the disease.

This work was supported by the Office of the Assistant Secretary of Defense for Health Affairs,



through the Gulf War Illness Research Program under Award No. W81XWH-15-1-0163. Opinions, interpretations, conclusions and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense.

## PgmNr 1813: Transcriptional endophenotypes predicting genetic risk of dementia in Mexican Americans.

### Authors:

M. Almeida <sup>1</sup>; J.E. Curran <sup>1</sup>; R. Duggirala <sup>1</sup>; M.P. Johnson <sup>1</sup>; E.K. Moses <sup>2</sup>; H.H.H. Goring <sup>1</sup>; G. Maestre <sup>1</sup>; D. Glahn <sup>3</sup>; J. Blangero <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) STDOL- Department of Genetics, University of Texas Rio Grande Valley- School of Medicine, Brownsville, Texas.; 2) The Curtin/UWA Centre for Genetic Origins of Health and Disease, School of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Western Australia, Perth, Western Australia, Australia; 3) Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut

---

Dementia is an age-associated neurodegenerative disease marked by diminishing cognitive function. There are multiple types of dementia, with Alzheimer's disease being the most common. Alzheimer's disease and related dementias (ADRD) are a major worldwide public health problem. ADRD affects individuals of all ethnicities, but Hispanic individuals show a 1.5-fold higher risk when compared to non-Hispanic whites. ADRD is a complex phenotype with both genetic and environmental causal components. Additionally, precise diagnosis often requires invasive measures. The identification of ADRD biomarkers is a research topic of growing importance. We propose to identify novel non-invasive endophenotypes (a particular type of biomarker that specifically indexes genetic risk) using high dimensional multiomic data in an existing cohort of Mexican American individuals. The participants are members of large Mexican American families of the GOBS (Genetics of Brain Structure and Function Study) project. Extensive multiomic data is available for approximately 2,000 subjects. We applied a mean-based endophenotype discovery method tailored for relatively rare diseases when extended pedigree are available. A set of 70 GOBS subjects were diagnosed as suffering from ADRD. The estimated heritability of ADRD is high ( $h^2 = 0.75$ ) in this population, supporting a major role of genes influencing risk. The ADRD cases have 363 1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> degree relatives currently not suffering from dementia. We contrasted this group of ADRD relatives with a set of 839 unaffected individuals lacking relatives with no ADRD. This approach detects genetic correlation between endophenotypes and disease risk and eliminates the problem of disease state effects. We tested the validity of our proposal by testing a set of 113 genes associated to Alzheimer's disease. We observed a clear enrichment for significant effects over that expected under the null hypothesis ( $p = 3.3 \times 10^{-10}$ ). Several cytokine genes (*TNF*, *IL1B*, and *IL1A*) exhibited reduced expression in relatives of cases versus controls. Notably, the *APBB1* gene showed increased expression in case relatives. This gene regulates the transcription of the amyloid precursor protein that is central for ADRD development. The efficient selection of relevant transcriptomic endophenotypes can aid the identification genes involved in ADRD risk. Quantitative non-invasive endophenotype identification may ultimately aid early detection of increased ADRD risk.

# PgmNr 1814: ‘Dark’ and ‘camouflaged’ genes may harbor disease-relevant variants that long-read sequencing can resolve.

## Authors:

M.T.W. Ebbert <sup>1</sup>; T.D. Jensen <sup>1</sup>; K. Jansen-West <sup>1</sup>; J.P. Sens <sup>1</sup>; P.G. Ridge <sup>2</sup>; J.S.K. Kauwe <sup>2</sup>; V. Belzil <sup>1</sup>; L. Prgent <sup>1</sup>; M.M. Carrasquillo <sup>1</sup>; Y.W. Asmann <sup>1</sup>; N. Ertekin-Taner <sup>1</sup>; S.G. Younkin <sup>1</sup>; O.A. Ross <sup>1</sup>; R. Rademakers <sup>1</sup>; L. Petrucelli <sup>1</sup>; J.D. Fryer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Neuroscience, Mayo Clinic, Jacksonville, Florida.; 2) Biology, Brigham Young University, Provo, UT

**Background.** Researchers have known for years that many genomic regions remain ‘dark’ because short-read sequencing read lengths are inadequate; the extent of the problem has largely remained uncharacterized, however. Here, we systematically characterize ‘dark’ regions, and regions we term ‘camouflaged’—dark regions arising from low mapping quality due to genomic duplications. We assess how well long-read sequencing technologies resolve these challenging regions, and we present a method to rescue variants from camouflaged genes. We applied our method to >10000 whole-exomes and genomes from the Alzheimer’s Disease Sequencing Project (ADSP), rescuing 4200 variants from protein-coding exons, including a rare 10bp frameshift deletion in *CR1*—a top Alzheimer’s disease gene.

**Methods and Results.** We identified 6054 gene bodies containing at least one dark region ( $\geq 10$ bp), many known to drive human disease. Regions are dark for two reasons: (1) low read depth ( $\leq 5$  reads); or (2)  $\geq 90\%$  of reads have a mapping quality  $< 10$ . Dark regions are based on an average across whole-genomes from 10 unrelated individuals, aligned to hg38 by BWA (median 37x coverage). 527 gene bodies are 100% dark and 2128 are  $\geq 5\%$  dark. Nearly 4000 are protein-coding genes, where most dark regions are intronic, followed by exons. 117 protein-coding genes are 100% dark in coding exons while 592 are  $\geq 5\%$  dark. Most dark regions are specifically camouflaged. Oxford Nanopore Technologies best resolves these regions (90%), compared to 10x Genomics (50%) and PacBio (64%).

Many disease-relevant genes are camouflaged, including *CR1* (26% camouflaged), *C4B* (83%), *SMN1* and *SMN2* (95%; spinal muscular atrophy), and *NEB* (10%; associated with 24 diseases). We also determined long-read alignment may not properly represent an individual’s *CR1* structure where Bionano Genomics’ optical mapping succeeds.

To rescue variants, we identified sets of genomic regions with  $\geq 98\%$  similarity (BLAT), masked all but one region, and aligned all reads previously aligned within the set to the unmasked region. We cannot determine upfront from which region in a set the mutation originated, but we can determine it exists within the set. The rare *CR1* 10bp frameshift was found in 5 ADSP cases and 0 controls (confirmed by Sanger). We identified a 6<sup>th</sup> case in a separate dataset.

**Conclusion.** Thousands of genes remain dark, potentially hiding important disease variants. Long-

read sequencing technologies are critical to resolving them.

## PgmNr 1815: Somatic mosaicism and mono-allelic expression across six brain cell types from single-cell RNA-seq in 48 AD individuals.

### Authors:

M. Kousi<sup>1,2</sup>; C. Boix<sup>1,2</sup>; M. Mathys<sup>2,3,4</sup>; K. Galani<sup>1,2</sup>; J. Mantero<sup>1,2</sup>; L.H. Tsai<sup>2,3,4</sup>; M. Kellis<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MIT Computer Science and Artificial Intelligence Laboratory (CSAIL), Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Picower Institute for Learning and Memory, Cambridge, MA; 4) Department of Brain and Cognitive Sciences, Cambridge, MA

Due to discovery power, most genetic variants reported to be associated with complex disorders are common with weak effect sizes, though few cases of rare mendelian events have also been described. For Alzheimer's disease (AD) in particular, the spectrum of causal variation spans common variants in 29 loci and rare familial mutations in three genes. In addition to the heritable genetic mechanisms underlying disease, somatic mosaicism may also account for a fraction of cases that remain with no molecular genetic diagnosis. Here, we seek to interrogate whether mosaicism in cell types that can still differentiate and propagate through the brain, such as astrocytes, oligodendrocytes and/or glial cells, can create focal areas of dysfunction leading to the manifestation of disease-relevant symptoms.

Here, we develop a new method for using single-cell RNA-seq technology to infer exonic mutations across brain cell types and characterize the identity of each cell harboring these mutations. We applied our method to Smart-seq2 data from post-mortem brain samples in a cohort of 24 Alzheimer's and 24 control patients from the ROS/MAP project. We aligned single-cell RNA-seq reads, inferred called genotypes, QC filtered, and rejected germline hits using whole-genome-sequencing. We validated called mutations both through targeted higher depth re-sequencing and by computational methods including linked-read analysis and comparison to low allelic balance calls in brain WGS, resulting in a cell-type specific map of mosaic events per individual.

We next assessed differential mutational burden across healthy and affected individuals, and detected several potential clonal events in glial populations, that further cluster in discrete pathways differing by cell-type. Our mutational and burden analysis uncovered multiple significantly differentially mutated or altered genes for oligodendrocytes, microglia, and excitatory neurons, many of which have been previously implicated in AD. Finally, we observed that 80% of germline heterozygous variation is expressed in a mono-allelic manner in single cells.

Our results show that single cell RNA-seq enables systematic cell type-specific survey of mosaicism in the aging brain, shedding light on the role of clonality and differential allelic expression. We also demonstrate how these observations can be used for relative pathway burden analysis in different cell types and prioritizing mutations, genes and pathways, for functional validation.

## **PgmNr 1816: Investigating oxidative damage: Implications in cognitive and metabolic phenotypes of Alzheimer's disease in the Mexican American population.**

### **Authors:**

D. Reid; G. Pathak; A. Blessing; R. Barber; N. Phillips

[View Session](#) [Add to Schedule](#)

**Affiliation:** Graduate School of Biomedical Sciences, UNT Health Science Center, Fort Worth, Texas.

---

As the elderly population expands, it is predicted that many age-related diseases, such as cardiovascular disease, metabolic disorders, cancer, and neurodegenerative diseases will burden the healthcare system. The Texas Alzheimer's Research and Care Consortium (TARCC) is a collaborative research effort to identify factors in the development and progression of Alzheimer's Disease (AD) in the Mexican American population compared to non-Hispanic whites, and to better understand the role of these factors. In the Mexican American population diabetes, depression, stroke, and obesity are common risk factors for developing cognitive impairment. The reasons for the association between cognitive decline and comorbidities remain unclear. Some studies have shown correlation between common pathological changes that are observed in AD and those that are a result of DNA damage. The mitochondrial genome specifically is particularly vulnerable to DNA damage. Age-associated decrease in mitochondrial function results in accumulation of reactive oxygen species which are capable of damaging DNA and other vital biomolecules. Oxidative damage to DNA takes many forms, but oxidation of guanine (G) to 8-oxoG is one of the most prevalent lesions. Currently, the methods for detection of 8-oxoG are limited and lack reproducibility. We propose nanopore sequencing technology as an improved alternative to the current methods. Here we describe preliminary proof of concept results and discuss the future application of this method for analysis of mtDNA damage in the TARCC cohort. We will investigate if oxidative DNA damage may be implicated in cognitive and metabolic phenotypes observed in Mexican Americans as compared to non-Hispanic whites.

## PgmNr 1817: Enabling precision genomics diagnosis: Accelerating diagnosis of rare diseases in clinical setting by reporting and re-interpreting VUS variants in genes of unknown clinical significance.

### Authors:

I. Thiffault<sup>1,2,3</sup>; M. Cadieux-Dion<sup>1,2,3</sup>; J. Alaimo<sup>1,2,3</sup>; L. Zellmer<sup>1</sup>; E. Farrow<sup>1,2,4</sup>; C. Saunders<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, KC, MO; 2) Department of Pathology and Laboratory Medicine, , Children's Mercy Hospital, KC, MO; 3) University of Missouri-Kansas City School of Medicine, KC, MO; 4) Department of Pediatrics, Children's Mercy Hospital, KC, MO

---

Next-generation sequencing (NGS) has resulted in an exponential number of new gene-disease associations. Assessing the level of evidence available for these assertions is necessary as part of quality variant interpretation for clinical diagnosis. The standards and guidelines for the interpretation of variants developed by the ACMG and the AMP promote consistency of variant interpretation; however these guidelines do not apply to genes of unknown significance (GUS). *De novo* variants are heavily weighted, given their overall rarity in the genome and their contribution to human diseases. Other compelling genotypes include loss of function variants in genes with high pLi scores or compound heterozygous variants in a gene with biological relevance. In our center, such variants in GUS are only included in the report if there is human genetic evidence. Other GUS are not reported, but are submitted to Matchmaker Exchange. As such, inconclusive results for clinical exomes are frequent, particularly for ultra-rare conditions where the phenotypic spectrum is ill-defined. To further the knowledge base for new gene disease associations and phenotypic spectrum of emerging genes, we report our current experience in the diagnosis of rare disease in a clinical laboratory setting by evaluating variants in GUS, particularly *de novo* variants, as well as patients with novel phenotypes associated with known genes. **Methods:** Of ~2000 trio clinical NGS tests performed in a three year period, over 170 novel candidate genes or novel-expanded phenotypes were submitted to GeneMatcher. **Results:** 38% (65/171) candidate genes had more than 5 matches; of these 80% (52/65) had more than 10 GeneMatcher hits, which suggests an eminent publication implicating the gene in human disease. *De novo* variants were detected in 28% (48/171) candidate genes of which 90% (43/48) were clinically reported as GUS. Of these, many have only been previously reported in large autism cohorts where limited phenotypic information is available, albeit with *de novo* variants in some. **Conclusion:** The use of tools such as GeneMatcher, is an essential step for assembling the evidence necessary for a strong gene-disease association in humans, with identification of multiple affected individuals and functional studies being the ultimate goal. Data sharing is imperative for fostering multi-center collaboration for gene-disease relationships. Lastly, the reporting and re-interpretation of VUS and GUS is highly important.

## PgmNr 1818: Concordance of germline and non-germline genotypes in glioblastoma multiforme patients.

### Authors:

P. Dharia <sup>1,2</sup>; M. Scheurer <sup>2</sup>; T. Armstrong <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The University of Texas Health Science Center at Houston, Houston, Texas.; 2) Baylor College of Medicine, Houston, Texas; 3) Neuro-Oncology Branch (NOB), Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH)

---

**Introduction:** Glioblastoma multiforme (GBM) is an aggressive brain tumor with expected mortality within 2 years of initial diagnosis with median survival as low as 12.61 months. Germline and non-germline DNA sources have been historically used for varying purposes for studies in patients with GBM. The objective of this study was to compare genotype results between two sources of DNA – germline and non-germline – in patients with GBM to improve the usefulness of collected data in retrospective studies. **Methods:** We analyzed data from 143 patients included in a cooperative group front-line therapy trial who had been genotyped using both DNA sources – blood (germline) and tumor tissue (non-germline). Samples were genotyped using the Illumina chips, according to manufacturer's protocol. The genotypes were analyzed using GenomeStudio. The concordance was analyzed by two methods: 1) percent concordance per chromosome - calculated by dividing the number of concordant SNPs by total number of SNPs in the germline DNA in a given chromosome; and 2) Kappa coefficient with 95% CI - calculated with two-sided significance level of 0.05. **Results:** After standard quality control measures, germline and non-germline genotypes were determined from 141 patients and for 937,203 SNPs from somatic chromosomes. The mean concordance among germline genotypes compared to non-germline genotypes for autosomal chromosomes in individual samples was 91.78% (range: 88.99 – 93.93). The lowest kappa coefficient was seen in chromosome 22 (0.9152) followed by chromosome 10 (0.9172). We also found that the kappa coefficient and concordance are proportional to the call frequency in germline DNA source. The lowest germline call frequency was seen in chromosome 10. **Conclusion:** For most chromosomes, no significant discordance between germline and non-germline genotypes was found, suggesting either could be used for retrospective studies as a good source of DNA for most SNPs. However, concordance was especially low for chromosome 10. This is of interest for GBM as the MGMT gene is located on chromosome 10, which has been associated with poorer therapy-related outcomes among GBM patients. Methylation of the MGMT promoter region regulates MGMT expression which is correlated with GBM survival. This indicates that tumor tissue derived genotypes may not be suitable for all types of retrospective genomic studies, e.g., those examining factors in areas of known genomic disturbance in the tumor tissue.



## PgmNr 1819: Omic-analysis to discovery genetic biomarker of extreme CNS forms of Incontinentia pigmenti by using IPGB biobank.

### Authors:

M.V. Ursini <sup>1</sup>; A. Pescatore <sup>1</sup>; M.G. Rosano <sup>1,2</sup>; N. Coratella <sup>1</sup>; A. El-Sharkawy <sup>1</sup>; D. Fergola <sup>1</sup>; M.B. Lioi <sup>2</sup>; F. Fusco <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Human Gen, IGB ABT CNR, Naples, Naples, Italy; 2) Dept. of Science, University of Basilicata, Potenza Italy

---

The phenotypic heterogeneity of Incontinentia Pigmenti (IP, MIM308300), a rare X-linked dominant disorder affecting neuroectoderm, suggests to search for molecular genetic markers to aim personalized therapeutic strategies.

Although significant progress has been made over the last few years in the diagnosis of this rare condition, the genetic defect and underlying pathological abnormality remain unknown. Moreover, to date no definitive cure is available for IP, nor are there sufficiently reliable and specific biomarkers to predict the disease severity. This is in part due to the rarity and genetic and phenotypic heterogeneity of IP disease and the lack of access to patient samples.

The availability of the genetic Biobank for IP has addressed this bottleneck and supported IP research. The Incontinentia Pigmenti Genetic BioBank Project (IPGB, <http://www.igb.cnr.it/ipgb>) was launched in 2015 with the aim of providing evidence for the implementation of personalized medicine by constructing a large, patient-based biobank and to enhance research on IP, related to pathophysiology, biomarkers and therapeutic approaches.

The cohort includes 0-60 years old, encompassing 436 clinically-confirmed IP cases (IP female and IP male) and 7 EDA-ID male from 200 cooperative medical institutes worldwide including USA. The Biobank contains blood samples, peripheral blood mononuclear cells (PBMC), red blood cells/granulocyte pellet (totaling 1370 samples).

We will describe the IPGB infrastructure and we will present preliminary results of the project to identify what variations/genes/pathways may affect the severity of IP, we used an OMIC-strategy: Target Exome Sequencing of Metabolic Pathway in 80 IP cases (40 IP-severe, 40IP-mild); Whole Exome and Transcriptome Sequencing in 3 IP-severe. More than 30,000 high quality variants were found across the haloplex samples, with an average of 430 variants/sample, and more than 300,000 across the exome samples, with an average of 40,900 variants/sample. Statistical analysis and gene validation are in progress. All IP patients belonged to our historical collection and to IPGB biobank.

## PgmNr 1820: Impact of exercise rehabilitation on skeletal muscle transcriptional programs in Parkinson's disease.

### Authors:

K. Lavin<sup>1,7</sup>; Y. Ge<sup>8,9</sup>; S.C. Sealton<sup>8,9</sup>; M.N. McDonald<sup>2,3,6</sup>; K. Wilk<sup>8,9</sup>; V.D. Nair<sup>8,9</sup>; P. Lakshman Kumar<sup>2,3,6</sup>; S.T. Windham<sup>4,7</sup>; M.M. Bamman<sup>1,5,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Cell, Developmental, and Integrative Biology, University of Alabama at Birmingham, Birmingham, AL.; 2) Department of Genetics, University of Alabama at Birmingham, Birmingham, AL.; 3) Department of Medicine, University of Alabama at Birmingham, Birmingham, AL.; 4) Department of Surgery, University of Alabama at Birmingham, Birmingham, AL.; 5) Department of Neurology, University of Alabama at Birmingham, Birmingham, AL.; 6) Division of Pulmonary, Allergy and Critical Care Medicine, University of Alabama at Birmingham, Birmingham, AL.; 7) UAB Center for Exercise Medicine, University of Alabama at Birmingham, Birmingham, AL; 8) Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY.; 9) Center for Advanced Research on Diagnostic Assays, Icahn School of Medicine at Mount Sinai, New York, NY.

---

Parkinson's disease (PD) is the most common motor neurodegenerative disease, and neuromuscular function deficits associated with PD contribute to disability. Targeting these symptoms, our laboratory has successfully used 16-wk high-intensity resistance exercise as rehabilitative training (RT) in individuals with PD. We observed significant improvements in muscle mass, strength, power, motor unit activation and total score on the 39-item PD Questionnaire (PDQ-39), providing encouraging evidence of potential symptom reversal through RT. In order to characterize the transcriptome-wide transducers of this intervention in PD, we generated RNA-seq data from skeletal muscle of a subset of these individuals [4M/1W, 67±2yr, Hoehn & Yahr stages 2 (n=3) and 3 (n=2)] before and after 16-wk high intensity RT. Following RT, 302 genes were significantly upregulated, notably related to remodeling and nervous system/ muscle development. Additionally, 404 genes, primarily negative regulators of muscle adaptation, were downregulated. Next, we used Pathway-Level Information Extractor (PLIER) to reveal coordinated gene programs (as latent variables, LV) that differed in skeletal muscle among young (YA), old (OA) healthy adult muscle (n=12 per cohort), and PD subjects pre- vs post-RT. Notably, the LV4 gene program was significantly lower at baseline in PD than YA, tended to be lower in OA, and was significantly increased by exercise. LV4 genes are associated with angiogenesis, axon guidance, and muscle remodeling. The LV16 program was higher in both PD and OA than YA and was reduced by 16-wk RT in PD; functional annotations suggest involvement in denervation, autophagy, and cell death. In summary, this approach enabled identification of two novel skeletal muscle transcriptional programs that are dysregulated by PD and aging, respectively. Encouragingly, RT has a normalizing effect on both programs in individuals with PD. Findings provide direction towards elucidating the molecular mechanisms responsible for RT-induced improvements in symptom severity and optimizing exercise regimens for individuals with PD.

Supported by NICHD T32-HD-071866.

## PgmNr 1821: Elucidating iPSC-derived neuronal phenotypes of Smith-Magenis syndrome.

### Authors:

J. Cross; J. Russ; F.J. Gordovez; S. Detera-Wadleigh; F. McMahon

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland.

Retinoic Acid Induced I (*RAI1*) is highly expressed in the brain and is thought to encode a transcriptional regulator. Smith-Magenis Syndrome (SMS) is caused by a ~3.7 MB deletion on chromosome 17p11.2, which includes *RAI1*, but ~10% of cases have mutations in *RAI1* alone. Clinical features of SMS include obesity, craniofacial abnormalities, sleep cycle disturbance, and intellectual disability. Although several studies have investigated *RAI1* deletions in mice, very few have examined the functional effects in patient-derived cells.

We have generated induced pluripotent stem cells (iPSCs) from SMS patients with mutations in *RAI1* to study the impact at the cellular level. Genes co-expressed with *RAI1* were established using RNAseq data sets from post-mortem brains, and evaluation of these genes indicated that *RAI1* could have an important function at neuronal synapses. Specifically, *SYT5* (calcium-dependent vesicle fusion) and *PHACTR1* (synaptic plasticity) were present. Therefore, we first differentiated iPSCs from three SMS patients and unaffected controls into neuronal precursor cells (NPCs).

RNAseq data from these cells provided 293 differentially expressed genes (122 up-regulated, 171 down-regulated). The up-regulated genes were associated with terms relating to neuropsychiatric disorders, such as schizophrenia and depression, suggesting that *RAI1* may have a broad role in neuropathology. The down-regulated genes provided strong evidence of defects in the extra-cellular matrix (ECM), including genes such as asporin and decorin which are implicated in axonal development. Interestingly, *BICDL1* is the only gene found in both the co-expression and NPC RNAseq data sets, and has a role in vesicle secretion and neurite outgrowth.

As such, it is critical to differentiate our iPSCs into neurons. As transcription of *RAI1* can be influenced by retinoic acid, I am utilizing a direct differentiation approach, in which the human transcription factor, neurogenin-2, is transfected into the desired iPSC lines. Once stable transfection has occurred, doxycycline can induce differentiation directly into excitatory cortical neurons in a matter of weeks. I then aim to ascertain differences in neurite and dendritic growth, with particular focus on proteins involved in the ECM and synaptic vesicle transmission. Such knowledge will provide insights into mechanisms involved in SMS and other neuropsychiatric disorders, ultimately leading us to novel treatments.

## PgmNr 1822: Low coverage RNA sequencing increases power for eQTL studies.

### Authors:

T. Schwarz<sup>1</sup>; K. Burch<sup>1</sup>; I. Mandric<sup>2</sup>; T. Boltz<sup>3</sup>; M. Bot<sup>4</sup>; N. Mancuso<sup>5,6</sup>; R. Ophoff<sup>3,4</sup>; B. Pasaniuc<sup>2,3,7</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA, USA; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 3) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 4) Center for Neurobehavioral Genetics, University of California Los Angeles, Los Angeles CA, USA; 5) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 6) Center for Genetic Epidemiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 7) Department of Computational Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

---

Expression quantitative trait studies (eQTL) have proven to be a powerful approach to identify common genetic variants contributing to regulation of gene expression, and subsequently to complex traits and diseases. Here, we show that low-coverage RNA-seq across an increased number of individuals attains more power for eQTL discovery than high-coverage RNA-seq with a limited amount of individuals, under a fixed budget. We propose a framework to improve gene expression estimation in low-coverage bulk RNA-Seq datasets, by leveraging correlation structure across genes as well as patterns across individuals to correct low confidence gene expression estimates. We show via simulations that reducing coverage while increasing the number of individuals in an eQTL study can boost power to discover eQTLs. This is quantified by effective sample size, or the amount of individuals that would have needed to be sequenced at high-coverage to discover the same number of eQTLs. We find that RNA-Seq data at 5-fold reduction in coverage captures upwards of 40% of the variation of high-coverage RNA-sequencing, on average. Within the context of reducing experimental costs, our results suggest that low-coverage RNA-sequencing in many individuals can yield increased effective sample size of eQTL studies based on high-coverage RNA-seq. We explore an empirical demonstration, by assuming a fixed budget and accounting for all costs, and show that instead of sequencing 690 individuals at 50 million reads per sample, we can achieve an effective sample size of over 800 individuals by sequencing 2000 individuals at 10 million reads per sample. Applying our approach for low-coverage gene expression measurement can further improve these estimates, which provides an additional boost to the effective sample size.

## PgmNr 1823: Identifying and validating cell type specific driver genes in autism-associated copy number loci using cerebral organoids.

### Authors:

E. Lim<sup>1,2</sup>; Y. Chan<sup>1,2</sup>; G. Church<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Harvard Medical School, Boston, Massachusetts.; 2) Wyss Institute, Boston, Massachusetts.

---

Neuropsychiatric disorders have been particularly challenging to study using model organisms, and recently, human-derived cerebral organoids demonstrate great promise for discovering molecular processes that are important in these disorders. However, several challenges remain in achieving robust phenotyping of cerebral organoids to discover cell type specific genes. We perform RNA sequencing on 1,420 cerebral organoids from 25 donors, and describe a framework (Orgo-Seq) to identify cell type specific driver genes, for two major risk loci associated with autism, by leveraging on bulk and single-cell RNA sequence data. We identify stem cells and neuroepithelial cells as critical cell types for 15q11-13 duplications and 16p11.2 deletions respectively, and discover novel and previously reported cell type specific driver genes.

Finally, we describe a CRISPR/Cas9-based framework to differentiate 'mosaic' cerebral organoids harboring synthetic mutations in *KCTD13*, followed by single-cell dissociation for fluorescence activated cell sorting (FACS) and targeted MiSeq sequencing. We validated the observations from our transcriptomics analyses, that deleterious mutations in *KCTD13* in the 16p11.2 locus are significantly enriched in Nestin+ neuroepithelial cells (Wilcoxon  $P=4.6 \times 10^{-3}$ ), depleted in TRA-1-60+ stem cells (Wilcoxon  $P=3.5 \times 10^{-3}$ ), but do not affect the proportion of NeuN+ neurons in the cerebral organoids (Wilcoxon  $P=0.47$ ). Our work presents a quantitative discovery and validation framework for identifying cell type specific driver genes in complex neurological diseases using cerebral organoids.

## PgmNr 1824: Proteomics of addiction: Postmortem brain analyses of cocaine and opioid use disorder.

### Authors:

C. Walss-Bass <sup>1</sup>; R. Hu <sup>2</sup>; L. Stertz <sup>1</sup>; G. Fries <sup>1</sup>; W. Russell <sup>3</sup>; N. Crist <sup>1</sup>; G. Sandberg <sup>4</sup>; Z. Zhao <sup>1,2</sup>; T. Meyer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry and Behavioral Sciences, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX.; 2) Center for Precision Health, School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX; 3) Department of Biochemistry and Molecular Biology, University of Texas Medical Branch at Galveston, Galveston, TX; 4) Harris County Institute of Forensic Sciences, Houston, TX

Substance use disorders (SUD) are a significant global health problem, with drug overdoses being the leading cause of death for Americans under the age of 50, killing more than guns, car accidents, or breast cancer. Drug use leads to changes in the structure and function of the brain by causing alterations in brain proteins. However, the assessment of protein modifications in human brain has been limited by the difficulty in obtaining brain tissue. Here we present preliminary data in our investigation of brain proteome alterations in the dorsolateral prefrontal cortex (BA9) of subjects with cocaine use disorder (CUD, n=10), opioid use disorder (OUD, n=20) and controls (n=12) from the UTHealth brain collection. Importantly, all CUD and OUD subjects died of a cocaine or opioid overdose, respectively.

Approximately 50 mg of tissue was lysed in RIPA buffer and extracted proteins were reduced, alkylated, de-lipidated and digested. NanoLC MS/MS was performed using a nano-LC chromatography system, coupled on-line to a mass spectrometer through a nanospray ion source. Data were acquired using XCalibur and the raw mass spectrometry data files were processed using MaxQuant. Peptide identifications were accepted if they could be established at greater than 95.0% probability. MS/MS spectra were searched against the Swiss-Prot human database. In total, 4584 unique proteins were identified in all the samples. Data was normalized and differential expression analyses were performed in CUD and OUD vs. controls respectively using the protein-wise linear models combined with empirical Bayes statistics implemented in the DEP (Differential Enrichment analysis of Proteomics data) R package and the *limma* R package.

We identified 49 proteins differentially expressed between the CUD and control groups and 40 proteins between the OUD and control groups ( $p.val < 0.05$  and  $|\log FC| > \log_2(1.5)$ ), of which 10 were found in both CUD and OUD. Pathway enrichment analyses identified protein localization to synapse and myelination pathways to be enriched in CUD, and acute inflammatory response and astrocyte development as the main enriched pathways in OUD.

Our results point to unique brain protein alterations induced by cocaine and opioids, suggesting distinct mechanisms of action and neurotoxicity. These results could shed light on the neurobiological mechanisms of SUDs and could lead to development of novel therapeutic approaches to minimize damage induced by these drugs of abuse.

## PgmNr 1825: Approaches to transcriptome analysis of human induced neurons in co-culture with murine glia to model functional synapses.

### Authors:

C. Purmann<sup>1,2</sup>; X. Zhang<sup>1,2</sup>; C. Pak<sup>3,4</sup>; Y. Huang<sup>1,2</sup>; R. Pattni<sup>1,2</sup>; S. Grieder<sup>5</sup>; M. Wernig<sup>5</sup>; D. Levinson<sup>1</sup>; B. Aronow<sup>6</sup>; T. Sudhof<sup>3</sup>; A. Urban<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 3) Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, CA; 4) Department of Biochemistry & Molecular Biology, University of Massachusetts Amherst, Amherst, MA; 5) Department of Pathology and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA; 6) Department of Pediatrics, Computational Medicine Center, University of Cincinnati College of Medicine, Cincinnati, OH

Neurons derived *in vitro* from human iPS cells carrying known genetic variants are an increasingly important model for studying the molecular mechanisms and neurophysiological consequences of neurogenetic disorders. To make use of the full potential of this approach, the neurons have to reach maturity and display functional synaptic properties. Achieving neuronal maturity often requires co-culture with glial cells derived from mouse brains. This is a robust and well-established procedure for neurophysiological analyses, but it creates cell cultures that contain both mixed cell types and from mixed species. Therefore functional genomics analyses of these cultures require either experimental or computational separation of cells, but the cultures contain dense and inter-connected networks and are difficult to dissociate.

We developed a dissociation protocol that generates intact single cells from cultures of human induced neurons (iNs) that had matured on mouse glia.

Using this dissociation protocol we carried out droplet-based single-cell (sc) RNA-Seq analyses of mature human iN cultures. We accompanied this with comparative analyses between scRNA-Seq, bulk RNA-Seq (i.e. of RNA extracted from the neural cultures without dissociation and separation of cells), and bulk RNA-Seq after dissociation and separation of cell types by immunopanning (i.e. of RNA extracted separately from neurons and glia).

Bulk RNA-Seq analysis that includes computational separation of murine and human reads, using a Bowtie2+RSEM pipeline, already performed unexpectedly well in distinguishing between the two species. This was confirmed by RNA-Seq of immunopanned cells. For scRNA-Seq where single cells are analyzed, the co-culture of murine and human cells does not interfere with the analysis. Both the bulk and scRNA-Seq detected very similar levels of human versus murine sequencing read distributions. However, unlike bulk RNA-Seq, the scRNA-Seq approach revealed varying degrees of cellular heterogeneity within cell types from the two species. The single-cell approach also provides an increased ability to gauge variability between biological replicates and enables functional genomics studies of, for example, the molecular basis of neuron induction and maturation and reciprocal signalling effects on heterogeneous cell types. Taken together, we present a comprehensive comparative overview of approaches to transcriptomic analysis of heterogeneous human stem cell

based neuronal cultures.



## PgmNr 1826: Using a novel “audiomic” approach to understand hearing loss of unknown etiology.

### Authors:

A. Koleilat <sup>1</sup>; L. Schultz-Rogers <sup>2</sup>; J.A. Morales-Rosado <sup>2</sup>; L.A. Schimmenti <sup>2,3</sup>; E.W. Klee <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Mayo Clinic, Mayo Clinic Graduate School of Biomedical Sciences, Clinical and Translational Sciences Track, Mayo Clinic, Rochester, Minnesota, USA Rochester, MN.; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota, USA; 3) Department of Clinical Genomics, Mayo Clinic, Rochester, Minnesota, USA

---

Whole exome strategies are now the mainstay for genomic diagnoses in patients who have multisystem disorders and lack a unifying reason for their condition. The Mayo Clinic Genomic Odyssey Board has performed analysis for 1451 patients who have undergone whole exome sequencing. Of this group 88 patients were identified to be deaf or hard of hearing of unknown etiology as part of their systemic condition. In order to find the cause of their condition, we are developing novel strategies in this subset of patients. Previous tiered strategies evaluated target genes for 1) *de novo* variants, 2) homozygous recessive variants, and 3) compound heterozygous gene variants. Our novel strategy utilizes the evaluation of the transcriptome of zebrafish and mouse inner ear hair cells in order to identify novel genes that could be implicated in hearing loss that would have otherwise been ignored through previous strategies. We used a large dataset of known inner ear hair cell and supporting cell transcripts obtained from published literature (Barta et al. 2018, Elkon et al. 2015) and performed what we call an “audiomics” analysis. Using the publically available adult zebrafish RNAseq data from Barta et al., of the 442 uniquely expressed genes in the sensory hair cells and 2152 uniquely expressed genes in the non-sensory hair cells 1273 genes had human orthologues. Using publically available P0 mouse microarray expression data from Elkon et al. 2015 human orthologues for 1185 genes were identified. Between the zebrafish and mouse data only 95 duplicates between species were noted. This “audiomics” analysis includes referencing these two lists (zebrafish and mouse) to a gene list created from phenolyzer using the term “hearing loss”. Phenolyzer returned a list that includes 711 putative genes with various degrees of association to the term “hearing loss”. Therefore, of the 2393 genes uniquely expressed in zebrafish sensory and non-sensory supporting cells and mouse hair cells, 2262 were not captured by the current phenolyzer gene list. This strategy demonstrates our unique approach to identifying novel gene-disease associations for hearing loss and could possibly identify the cause of disease in this subset of patients.

## PgmNr 1827: Identification of novel limbal stem and progenitor cell types in human corneal epithelium using single-cell transcriptomics.

### Authors:

J. Choi <sup>1</sup>; S. Kim <sup>2</sup>; R. Chen <sup>2,3</sup>; D.-Q. Li <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX.; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.; 3) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 4) Ophthalmology, Baylor College of Medicine, Houston, TX.

---

Limbal stem cells (LSCs), also known as corneal epithelial stem cells, are located at the basal epithelial layer of the corneal limbus and serve an important function in maintaining the homeostasis of the corneal epithelium. The inherent regenerative properties of LSCs, therefore, hold the promise for treating various degenerative eye diseases and LSC-based cornea transplantation. Several putative molecular markers of LSCs have been previously identified. Nonetheless, the specificity of these markers remains largely controversial. To address this gap in the current understanding of LSCs, we performed a transcriptome profiling of heterogeneous corneal limbal population using single-cell transcriptomics technology to identify LSCs and their exclusive markers. Using the 10x Genomics platform, we obtained the transcriptome of about 18,000 individual single-cells. Furthermore, principal component analysis (PCA) was used to reduce the dimensionality of the dataset, followed by clustering unique cell types through the graph-based visualization method Uniform Manifold Approximation and Projection (UMAP). A total of 11 unique clusters of cells were identified and each cluster was assigned to a putative cell type based on previously published biomarkers. As a result, 5 differentiated cell clusters, 3 progenitor clusters, 2 melanocyte clusters, and 1 putative stem cell cluster were identified. Differentiated cell clusters were defined by well-known differentiation markers such as *KRT12* and *GJA1*. Interestingly, the differentiated cell clusters were divided into two main subtypes as characterized by presence of *KRT3* gene expression. Two subtypes of differentiated cells likely represent different stages of differentiation or spatial location they reside. Progenitor clusters were defined by known basal limbal markers such as *KRT14* and *KRT19*. One progenitor cluster was characterized by the increased expression of cell cycle related genes and, therefore, has been assigned as transient amplifying cells (TACs). Two melanocyte clusters were defined by many melanocyte related genes and likely represent different stages of cell maturation. Lastly, one putative stem cell cluster was defined based on the expression of several novel stem-cell related markers. To further validate each cell cluster and determine their location in human corneal limbus tissue, a combination of tissue RNA in situ hybridization (ISH) assay using RNAscope and antibody staining is currently underway.

## PgmNr 1828: Best approaches and methods for saliva-derived genomic DNA extractions.

### Authors:

T. Brunetti <sup>1</sup>; M. Campbell <sup>1</sup>; M. Boorgula <sup>1</sup>; A. Franklin <sup>1</sup>; T. Beaty <sup>3</sup>; R. Mathias <sup>1,2</sup>; M. Daya <sup>1</sup>; K. Barnes <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Colorado Center for Personalized Medicine, Univ Colorado, Denver, Colorado.; 2) Department of Medicine, Johns Hopkins University, Baltimore, MD, 21224, USA; 3) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, 21205, USA

---

Technological advancements have led to the development of minimally invasive and easily accessed (e.g. clinic, home, field) biospecimen collection methods and specimen stability for DNA. Saliva is a popular alternative to whole blood for sequencing of germline DNA, due in part to the ease, portability, and lower cost. This is of particular importance when conducting large-scale population genetics studies, especially in challenging field settings where human subjects risks associated with venipuncture and access to appropriate infrastructure for the collection, processing and storage of whole blood may be burdensome or impossible. With the increased demand for saliva biospecimens there has been an influx of genomic DNA extraction kit options from various commercial entities; however, there is a dearth of unbiased, peer-reviewed literature for objectively assessing which collection method and extraction kit should be used, balancing the considerations of cost and time. In the absence of these resources, investigators may compromise sample collections for a more expensive kit both in time and money, at the expense of under-powering the study. As part of the Consortium on Asthma among African-ancestry Populations the Americas (CAAPA), for which a major goal is to study the genetics of asthma in populations of African ancestry where prevalence is disproportionately high, we have identified the following logistical questions: 1) Which extraction kit is the most-efficient and yields high quality results; and 2) Are DNA extracted from saliva and whole blood comparable for variant calling?

Thirty participants with saliva and whole blood specimens were recruited into a pilot study to address these questions. Four different saliva extraction kits were used and the following four metrics were assessed: DNA quality, DNA yield, time for extraction, and cost. Because time is a significant factor for large-scale studies, we also considered using a kit that can be automated versus kits that require labor intensive, hands-on extraction processes. Quality was assessed by assessing A260/280 readings, DNA fragmentation levels, and bacterial contamination ratios. A TaqMan 16S rRNA gene expression kit was used to measure bacterial contamination and a TaqMan RNaseP probe was used to measure human control DNA. Paired sample collections of whole blood and saliva were extracted and genotyped on a custom-build of Illumina's MEGA chip where concordance was assessed across ~2M SNPs.

## **PgmNr 1829: Primate brain retinal ganglion cell (RGC) topographic map revealed by ultra high DSI 2D hexagonal pattern (Polyak, 1953) translates faithfully from optic nerve to 3D thalamic arrays.**

### **Authors:**

J. Korenberg <sup>1</sup>; F. Yeh <sup>2</sup>; L. Dai <sup>1</sup>; A. Van Hoek <sup>1</sup>; E. Hsu <sup>3</sup>; O. Abdullah <sup>4</sup>; S. Joshi <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Pediatrics, Univ of Utah, Salt Lake City, Utah.; 2) Univ of Pittsburgh, Pittsburgh, PA; 3) Bioengineering, Univ of Utah, Salt Lake City, Utah.; 4) Univ. Abu Dhabi, Abu Dhabi, United Arab Emirates

---

The complex circuitry of the primate visual system is an elegant testament to the central role of light processing in survival throughout the past 550 my of evolution. Organized as the integration of the evolutionarily older “primary” visual system, that then maps on to the visual cortex, a striking feature is the general “retinotopic” organization that is found in most target regions of both systems. How this organization arises from the planar organization (Polyak 1953) seen as a plane of hexagons for retinal ganglion cells (RGC) is known for only a few basic rules. The fundamental need has been for a technology capable of tracing multiple neighboring axonal bundles from the retina or optic nerve through the complex crossings of the optic chiasm, to target regions. Diffusion spectrum imaging (DSI) provides a theoretical framework to do this but has needed internal systems of known biological structures to be able to evaluate its validity. We report the creation of a heretofore unprecedented dataset of the macaque fascicularis brain that utilized >500 directions, multiple b values of up to 40,000, and 140 h of acquisition in a Bruker 7T with a 20 cm coil. Using the DSI Studio algorithm (Yeh), we generated a series of tests that required successful navigation of multiple DSI challenges including closely spaced crossing fibers in the primate visual system. Further to dimensionally decrease the complex patterns, we developed novel approaches using subvoxel seeding equivalent to micro-voxels of 16-32 micron cubed regions. To align these with high resolution neuroanatomic architecture, we used diffeomorphic transformations to integrate the tracks with a 3D atlas generated from the same animal using block face acquisitions of the complete brain (2300 slices) sectioned at 30microns. We report here two major findings. Axon patterns, presumably represented by the diffusivity of multiple clustered axons, correctly tracked from the right versus the left optic nerve, to reveal a closely apposed pattern of more than 1,000 red versus green signals in the left visual cortex. Second, the findings revealed a pattern of hexagonally arranged planes that could be followed from each of the optic nerves, through the chiasm and to the lateral geniculate nucleus where the patterns were maintained. These results strongly suggest that DSI combined with histological data, may be suitable for tracing the micro and mesocircuitry of the primate brain.

## PgmNr 1830: Gene Curator: A multi-source, data-driven approach to disease-based gene ranking.

### Authors:

P. Singh; L. McAuliffe; M. Fasnacht; S. Ng; J. Delaney; J. Hakenberg; M. Jung

[View Session](#) [Add to Schedule](#)

**Affiliation:** Illumina Inc, 5200 Illumina Way, San Diego, CA 92122.

Targeted sequencing based on data-driven disease content strategies is a powerful and cost-effective application of high throughput sequencing technologies. However, the data supporting evidence for correlations between genes and diseases are spread across several isolated databases with no single tool currently available that integrates the information for the purpose of assay design, thus necessitating better integration models.

Here we present Gene Curator, a tool that optimizes gene content for disease phenotypes using a multi-omics data driven approach where each type of information encompasses a separate module. Currently, four different modules are implemented namely, gene expression, GWAS catalog, scientific literature, and curated databases. The **gene expression** module is based on the data in Illumina Correlation Engine with more than 22K studies and over 500K individual experiments based on RNA-seq and microarrays. It embraces experimental heterogeneity by analyzing multiple cohorts that are diverse in their population, sample type, assays used and other factors. Our robust statistical pipeline allows us to find reproducible and tissue-specific gene expression signals in the “noise” of multiple studies. The **GWAS module** extracts data from the GWAS catalog and combines across multiple studies to bump up p-values based on study sizes. It then maps the SNPs to genes and derives gene p-values using a simulation approach that takes the length of the gene and linkage disequilibrium into consideration to rank genes with SNPs that have the lowest p-values for that gene. The **curated databases** module is Illumina’s gene panel app that starts with a phenotype supported by ontology backbone (e.g. Omim, SNOMED, MedGen), generating a list of disease-associated genes. The **scientific literature** module utilizes 30 million articles in PubMed with scores reflecting the number of text pieces supporting the gene-disease relation.

Gene Curator results can be utilized to tailor gene panels for a disease of interest, enhance current targeted content, understand the implications of sequencing results by intersecting with disease-gene correlations or to identify new biomarkers for complex disease models. Taken together, our modular approach enables streamlined design procedures for each gene panel application, simplifying the selection from CNV panels over tissue specific RNA expression screenings to genotyping panels.

## PgmNr 1831: Differential gene expression between the Turner and Klinefelter syndromes.

### Authors:

A. Rojas <sup>1</sup>; M. Manotas <sup>1</sup>; J. Calderon <sup>2</sup>; O. Moreno <sup>1</sup>; F. Suarez <sup>1</sup>; L. Lopez <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá, Colombia; 2) Universidad Nacional de Colombia, Bogotá, Colombia

---

**Background:** Analysis of patients with chromosomal abnormalities, including Turner syndrome (45,X) and Klinefelter syndrome (47,XXY), has highlighted the importance of X-linked gene dosage as a contributing factor for disease susceptibility. Escape from X-inactivation and X-linked imprinting can result in transcriptional differences between normal men and women as well as in patients with sex chromosome abnormalities

**Objective:** identify differentially expressed genes among patients with Turner and Klinefelter syndrome.

**Methodology:** Two sets of gene expression data of Turner and Klinefelter syndrome were obtained from the Gene Omnibus Expression (GEO) database of the National Center for Biotechnology Information (NCBI) with accession number GSE46687 and GSE42331. The statistical analysis of the data was performed using the R software, the data processing was done using different Bioconductor libraries. Differentially expressed genes were determined using significance analysis of microarray (SAM), for the data corresponding to Turner syndrome a critical value of  $d_i = 3.0649$  was selected and for the data of Klinefelter syndrome a value  $d_i = 0.9579$  was selected, corresponding respectively to a false positive percentage that approaches zero (FDR - False Discovery Rate). The functional annotation was performed in DAVID v6.8 (The Database for Annotation, Visualization and Integrated Discovery)

**Results:** There are no genes over expressed simultaneously in both diseases. When crossing the list of repressed genes for Turner syndrome and the list of over expressed genes for Klinefelter syndrome, there was a coincidence in 16 genes: *SLC25A6*, *AKAP17A*, *ASMTL*, *KDM5C*, *KDM6A*, *ATRX*, *CSF2RA*, *DHRX*, *CD99*, *ZBED1*, *EIF1AX*, *MVB12B*, *SMC1A*, *P2RY8*, *DOCK7*, *DDX3X*, of which eight genes encode to regulators of gene expression through epigenetic mechanisms.

**Conclusion:** our results suggest that global epigenetic and RNA expression changes play a central role for the Turner and Klinefelter phenotype.

## PgmNr 1832: Resonance tunneling integrated omics sequencing.

**Authors:**

V. Vatannia; S. Vatannia; M. Yavi

[View Session](#) [Add to Schedule](#)

**Affiliation:** Resonance HRT, Palo Alto, California.

---

Resonance HRT has developed a novel approach utilizing resonant tunneling (RT) to achieve identification of biopolymers, DNA, RNA, proteins and metabolites on the same device. This technology enables single molecule long reads, which has not yet been fully achieved by the currently available platforms. In addition, Resonance tunneling-based sequencing approach is inherently preserving the secondary modifications on molecule. Lastly, this process spares the multi-step sample preparation and is independent of PCR-based sequencing technology.

The Resonance HRT device utilizes coupling between energy states of a custom designed quantum well structure (Chip) with the electronic state (HOMO and LUMO) of individual molecule components (e.g. nucleotides, peptides, methylations, sugars, etc.). The system (emitter, insulator, base, molecules and collector) is designed such that when the energy levels of the Emitter and Quantum Well are aligned, the electrons will pass through the insulator layer by way of resonance tunneling effect to the Base. These electrons are highly coherent in energy spectrum and will travel through the Base and will be collected by the Collector, only if the energy level of the molecules (in this case the DNA bases) align with the eigen energy of Quantum Well and Fermi level of the Emitter.

Results show a sharp Lorentzian Peak around -1.7V, which is indicative of RT in solution at room temperature over large distances. We have ruled out electrochemical reactions since the voltage is kept constant at -0.5V, below the level of electrochemical reactions for Pt in an aqueous. The current stops after withdrawing the tip more than 5nm from the original setpoint (for a total of 7nm between electrodes), within the range of electrochemical reactions but beyond the range of RT.

Our current projections indicate orders of magnitude improvement in standard measures over the current state of the art sequencing (e.g. Q40 raw accuracy, 1 base/ms/channel speed, and much lower total cost of sequencing) and entirely new possibilities for direct sequencing RNA, precise methylations mapping, non-destructive, single-molecule protein sequencing, and ultra-precise metabolomic sensing. Enabling multi-omics sequencing will broadly advance biomedical research, with commensurate impacts on a variety of applications such as displays and sensors.



## PgmNr 1833: A 12 minute, single tube, nanogram input library prep for WGS, FFPE, WES, and RNAseq.

### Authors:

J. Dunham <sup>1</sup>; B. Akins <sup>1</sup>; C. Mason <sup>2</sup>; C. Angermayer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) SeqOnce Biosciences, Pasadena, CA.; 2) Weill Cornell Medicine, New York, NY

---

Next Generation Sequencing (NGS) library construction is a workflow bottleneck. The processes of library construction for DNA and RNA sequencing are complex, error prone, costly, and in general are time consuming. Current commercial products offer multistep, multi-hour sample processing that can be cost and labor prohibitive. These issues have created a demand for a simple, rapid, and cost-effective library construction product that offers users application flexibility while minimizing construction complexity for low to ultra-high throughput sample processing. SeqOnce has developed a novel library technology that is rapid and minimizes sample processing complexity. The five tube kit is comprised of formatted master mixes for a simple and stable user workflow. The 12 minute library construction for Human WGS, FFPE, and WES, uses a single master-mix that when combined with fragmentation and PCR steps produces libraries in less than 50 minutes. A single size selection step occurs after PCR and the PCR free protocol is less than 20 minutes. The kit contains three master mixes and includes adapters, which maximally promotes product flexibility across multiple applications. With master-mixes that are stable for multiple days at ambient temperatures, automation on liquid handling platforms is effortless due to the simple workflow. The technology has been validated on human samples for shallow and deep whole genome sequencing, FFPE, WES, and RNAseq with input ranges of 10ng - 200ng. The sequencing data is either equivalent or superior to other competing library products, with high mapping rates and excellent performance across variable GC content. The technology is currently optimized for the Illumina platform and alternative NGS platforms are under evaluation for future development.



## PgmNr 1834: Best practices for handling genomic DNA during NGS library preparation.

### Authors:

K. Luttgeharm <sup>1</sup>; C. Pocernich <sup>1</sup>; S. Borg <sup>2</sup>; S. Siembieda <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Agilent Technologies, Ankeny, Iowa.; 2) Agilent Technologies, La Jolla, California

---

Superior NGS library preparation and sequencing results start with quality genomic DNA (gDNA). Quality control analysis provides a great deal of information that is required before beginning NGS library preparation including size, smear distribution, and concentration. Quality of the gDNA starting material can be modified by the dynamic environment and handling processes it encounters. To determine how common laboratory practices such as pipetting, vortexing, and freeze-thaw cycling impact gDNA quality, commercially available gDNA samples were analyzed using the Agilent 5200 Fragment Analyzer system. Both the Agilent Genomic DNA 50 kb kit and HS Genomic DNA 50 kb kit were used to determine the average size, smear distribution pattern, concentration, and percent of degradation of the gDNA sample. It was found that different sample handling practices have varied impact on gDNA sizing, with some, such as high-speed vortexing having a greater effect on sample integrity compared to gentler handling practices, such as slow pipetting with a wide-bore pipet tip. From this study, we have developed a set of best practices for handling gDNA. Quality control analysis of gDNA and gentle handling practices ensure gDNA integrity and build confidence in the success of downstream applications such as NGS library preparation.

## PgmNr 1835: IFN- $\gamma$ -dependent transcriptome altered by human cytomegalovirus-encoding UL23 protein on human foreskin fibroblast cells(HFF) using RNA-seq.

### Authors:

H. Li; S. Wu; L. Feng; H. Chai; X. Yang; Z. Deng

[View Session](#) [Add to Schedule](#)

**Affiliation:** College of Life Science and Technology, Jinan University, Guangzhou, Guangdong, China

Human cytomegalovirus (HCMV) is a human pathogenic herpesvirus belonging to the subfamily of beta-herpesvirus, which causes global epidemics and complications in AIDS patients and organ transplant recipients, and is a major factor of birth defects. HCMV infection is also associated with inflammatory and proliferative diseases such as certain cardiovascular diseases and cancers. Interferons (IFNs) are pleiotropic cytokines produced and secreted upon pathogen encounter to induce a remarkable change of the cellular gene expression profile, leading to an enforcement of intrinsic immunity, the induction of innate immunity and the stimulation and recruitment of adaptive immunity. Interferon responses are vital for a host to combat infections of many human viruses including human herpesviruses. HCMV has a large dsDNA genome of approximately 230 kb, encoding a multitude of proteins mediating immune evasive properties, targeting both innate and adaptive immune responses.

HCMV-encoding UL23 is a virion protein found in the tegument and is expressed in the cytoplasm in HCMV-infected cells. UL23 was found to facilitate viral immune escape from IFN- $\gamma$  responses and enhancing viral resistance to IFN antiviral effects in our previous studies. Here, we profiled the expression of mRNAs in UL23-ectopic expressing HFF cells using the RNA-seq to investigate the transcriptional changes during IFN- $\gamma$  treatment. At 12 hours IFN- $\gamma$  incubation, a total of 84,709,994 sequence reads was obtained. 18,585 protein-coding genes were observed in Refseq database. Differential gene expression analysis identified 257 differentially expressed genes (DEGs) between HCMV-infected and mock-infected HFF cells, including 28 up-regulated genes and 229 down-regulated genes. These regulated genes were involved in chemokine signaling pathway and NF-kappa B signaling pathway, all of which may be implicated in host defence against HCMV infection. These findings have provided a dynamic scenario of differentially expressed candidate genes at the virus-host interface and clearly warrant further experimental investigation associated with HCMV infection.

## PgmNr 1836: The value of long read amplicon sequencing for clinical applications.

### Authors:

L. Aro <sup>2</sup>; A. Mensenkamp <sup>1</sup>; L. de Bruijn <sup>1</sup>; E. Askar <sup>1</sup>; S. van der Heuvel <sup>1</sup>; E. Hoenselaar <sup>1</sup>; R. Derks <sup>1</sup>; M. van der Vorst <sup>1</sup>; M. Nelen <sup>1</sup>; M. Ligtenberg <sup>1</sup>; I. McLaughlin <sup>2</sup>; R. M. de Voer <sup>1</sup>; C. Heiner <sup>2</sup>; K. Neveling <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Radboudumc, Nijmegen, Netherlands; 2) Pacific Biosciences, Menlo Park, CA

NGS is commonly used for amplicon sequencing in clinical applications, to study genetic disorders and detect clinically actionable mutations. But this approach can be plagued by limited ability to phase sequence variants and makes interpretation of sequence data difficult when pseudogenes are present. Long read high fidelity (HiFi) amplicon sequencing can provide very accurate, efficient, high throughput (through multiplexing) sequences from single molecules, with read lengths largely limited by PCR. The data is much easier to interpret; phased variants and breakpoints are present within high fidelity individual reads.

Here we demonstrate SMRT sequencing of the *PMS2* and *OPN1MW* genes. In both cases there is a second very similar gene, *PMS2CL* and *OPN1LW*, which can make NGS and MLPA results very difficult to interpret. For *PMS2*, 8 amplicons ranging in size from 3.2 kb to 8.2 kb were designed using primers unique to this gene, covering the entire 38 kb. SMRT sequencing produced HiFi reads with coverage ranged from 200x to 1500x; data clearly indicated 2 deletions >1000kb with precise breakpoint mapping.

For *OPN1MW*, two 14 kb amplicons can be generated covering the regions of interest; *OPN1LW* produces one 15.5kb amplicon. These amplicons can be sequenced in their full length or digested with a restriction endonuclease to produce to two fragments, each about half the original length. We show results of sequencing these genes from 3 patient samples with different mutations, including sequencing the full-length amplicons. In 2 of the 3 cases, the mutations were confirmed by orthogonal technologies.

These results demonstrate several areas of added value from long amplicon sequencing:

- Efficiency
- Less PCR, no nesting
- No added tests (like MLPA)
- Multiplexing for high throughput
- Easier data interpretation and analysis; improved results
- Ability to distinguish between genes and pseudogenes
- Variant phasing within long reads
- Precise breakpoint detection

## PgmNr 1837: Highly scalable linked read library technology enables single tube whole genome sequencing for haplotype phasing and *de novo* sequencing.

### Authors:

Z. Chen <sup>1</sup>; L. Pham <sup>1</sup>; D. Wu <sup>1</sup>; G. Mo <sup>1</sup>; Y. Xia <sup>1</sup>; T. Phan <sup>2</sup>; T. Tran <sup>2</sup>; T. Truong <sup>2</sup>; H. Che <sup>2</sup>; S. Pham <sup>2</sup>; Y. Wang <sup>1</sup>; M. Lei <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Universal Sequencing Technology Corporation, Carlsbad, CA.; 2) Bioturing Inc, San Diego, CA.

---

Genome scale haplotype phasing and structural variant detection, and *de novo* assembly of novel genomes are major hurdles for short read based next generation sequencing (NGS) platforms. Long sequence reads are essential to overcome the significant sequence homology on many regions of the genomes. Several NGS library technology breakthroughs recently have demonstrated barcode linked-read sequencing method can effectively generate long read like information and successfully applied for human genome phasing, structural variation detection or *de novo* assembly of other genomes. However, they either require expensive capital expenditure on specialized instruments or are not scalable for commercial adoption yet due to sophisticated barcode generation. We have developed a simple and scalable NGS library technology, Transposase Enzyme Linked Long-read Sequencing (TELL-Seq<sup>TM</sup>), to use short NGS reads for genome scale haplotype phasing and/or *de novo* genome assembly. Millions of uniquely barcoded beads are used to generate linked reads, which could be linked as long as a hundred kilobases, by strand transfer reactions using transposase in a PCR tube in a standard NGS laboratory setting without any specialized instruments. TELL-Seq library procedure takes approximately 3 hours and multiple samples can be easily processed parallelly in a 96-well format when needed. The library protocol can be adjusted and used for various sizes of genomes from bacteria to human. Using TELL-Seq technology we are able to generate comparable and excellent haplotype phasing results on GIAB human samples, and successfully *de novo* assembly on microbial genomes with various GC content. In addition, *de novo* assembly for human size genome is expected to become achievable routinely soon with improvement on assembly algorithm to overcome the dependency on supercomputing capacity. More applications and analysis solutions are being developed for TELL-Seq library technology.

## PgmNr 1838: Bionano Genomics “Sample to Answer” workflow for single molecule analysis of variation in genome structure.

### Authors:

S. Bocklandt; H. Sadowski; E. Lam; A. Pang; T. Anantharaman; A. Hastie; G. Pljevaljcic; S. Way; P. Lynch; W. Willemse; M. Borodkin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bionano Genomics, San Diego, CA.

---

Complex structural changes are often observed in cancer and other diseases. Bionano optical mapping of very long single DNA molecules provides critical structural information for comprehensive genome analysis. The extracted DNA molecules are labeled at specific motifs and analyzed in massively parallel nanochannels. The single-molecule maps can be assembled into chromosome arm-length maps. The maps can be used for structural variation (SV) detection and elucidation of complex rearrangements.

Bionano presents an optimized workflow for cancer and constitutional disorders that includes DNA isolation and labeling, DNA imaging, and genomic data analysis. The workflow starts with the isolation of ultra-high molecular weight (UHMW) genomic DNA using Bionano Prep SP. This couples solution-based lysis with a purification step that leverages a novel process to bind, wash and elute UHMW genomic DNA. This entire protocol can be conducted with 3 hours of hands-on time on a batch of 6 samples, allowing 12 samples to be processed in one day. The eluted material is ready to label by day 2 and contains high quality UHMW DNA, that includes Mbp-sized molecules. After DNA isolation, an enzymatic labeling approach, Bionano Prep DLS (Direct Label and Stain), preserves the integrity of the DNA while labeling sequence motifs across the whole genome. The labeled DNA molecules flow through the Saphyr Chip and are imaged by the Saphyr system. Up to 3.9 Tbp of coverage can be collected per Saphyr Chip, allowing for the processing of multiple samples and/or very high depth for cancer analysis. The data analysis tools provided with Bionano Access take advantage of this data for genome assembly and SV analysis applications. At standard coverage, we provide unprecedented sensitivity to heterozygous SVs. At high coverage, the single molecule pipeline also uncovers SVs that occur at a 5% allelic frequency. These analyses can be performed on local resources or on Bionano Compute On Demand, an economical hosted offering. Combining all these elements, allows cancer researchers, for example, to get whole genome low allelic SVs from multiple samples, in less than 5 days, sample-to-SVs.

## **PgmNr 1839: Single enzymatic DNA fragmentation workflow simplifies library preparation from FF/FFPE samples of varying quality.**

### **Authors:**

M. Borns <sup>1</sup>; B. Rogers <sup>1</sup>; K. Chen <sup>1</sup>; G. Amparo <sup>2</sup>; C. Pabon <sup>2</sup>; G. Miles <sup>2</sup>; B. Arezi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Agilent Technologies, Genomics Solutions Division (DGG), La Jolla, CA; 2) Santa Clara, CA

---

Mechanical fragmentation of DNA is the most commonly-used method for library construction. Current protocols for mechanical shearing are labor intensive, time consuming, expensive, and therefore unsuitable for clinical customers. Because of these drawbacks, enzymatic shearing methods are becoming more popular. However, these methods also require additional optimization for different input amounts or when used with samples of varying quality. Furthermore, the enzymes are sensitive to commonly-used buffers for DNA storage such as TE because of EDTA presence, and/or generate more false SNPs or fusions compared to mechanical shearing. Here, we address these challenges by developing an endonuclease-based enzymatic double stranded-DNA fragmentation method which generates random DNA fragments from a variety of fresh frozen tissue, and FFPE samples of varying quality and input (in the range of 10-200ng) with a single protocol (i.e., no optimization other than adjusting the PCR cycles). Furthermore, the enzyme is not sensitive to commonly-used high/low TE DNA storage buffers.

The DNA fragments generated have 5'-phosphate and 3'-hydroxyl termini and can be directly used in end-repair/dA tailing without further modification/purification. Unlike transposase shearing, it is suitable for library preparation from any FFPE tissue derived DNA including highly-damaged FFPE samples. As we will demonstrate, fragmentation and library preparation experiments using three bacteria of varying GC content (30%-67%) and human DNA shows high coverage uniformity and minimal bias across all GC content. The correlation of coverage (Rs<sub>q</sub>) by targeted region is >0.9 (between enzymatic and Covaris replicates), and similar percentages of known and unknown variants are called by both methods. Furthermore, we observe more complex libraries (~50%) across samples of medium to high quality (TapeStation DIN >3) compared to mechanical shearing. This is likely due to lower loss of DNA since it is a single-tube process, and/or the presence of more DNA ends that are repairable (as opposed to mechanical shearing).

## **PgmNr 1840: Rapid and accurate quantification of Illumina NGS libraries using sparQ Fast Library Quant Kit on the Q real-time qPCR instrument.**

### **Authors:**

E. Kolossovski; P. Bartholomew; H. Liu; B. Komorous; D. Schuster

[View Session](#) [Add to Schedule](#)

**Affiliation:** Quantabio, 100 Cummings Center Suite 407J, Beverly, MA 01915

---

Accurate quantification of the number of amplifiable library molecules is a critical factor for obtaining high quality read data with NGS technologies. The high sensitivity, broad dynamic range, and specificity of qPCR to quantify library molecules that are suitable for bridge amplification provide significant advantages over methods for total DNA quantification. However, these advantages are often offset by the time to result, complexity of execution, and costs associated with qPCR run. Here we describe application of the sparQ Fast Library Quant Kit and Q real-time qPCR system to simplify accurate and reproducible library quantification with 50% shorter run times compared to traditional protocols.

The sparQ Fast Library Quant Kit consists of a set of six prediluted standards, an optimized library dilution buffer, and a ready-to-use master mix containing Illumina P5 and P7 primers that is optimized for fast cycling on the Q qPCR system. The Q qPCR instrument is a compact and portable system that exploits a novel magnetic induction technology to rapidly heat reactions held in a unique spinning aluminum rotor. This provides superior temperature uniformity of  $\pm 0.05^{\circ}\text{C}$  that eliminates well position effects and cross-instrument variability associated with traditional peltier block-based real-time cyclers. Robust optics with rapid data acquisition times are coupled to an intuitive data analysis interface to deliver highly accurate and reproducible qPCR results.

To test the efficacy of this new library quantification kit and novel qPCR instrument we measured libraries of various fragment sizes and GC contents over multiple trials using a universal fast cycling protocol. Results for sparQ on the Q were highly correlated to those obtained with a leading library quant kit following the manufacturer's recommended protocol. These data demonstrate highly precise measurements across multiple trials and production lots of reagents and standards, high and consistent amplification efficiencies for libraries of varying GC content and fragment sizes, with completed run times in under 40 minutes. Consequently, in addition to simplifying and streamlining reaction assembly and providing faster time to result, the Q and sparQ Fast Library Quant Kit offer the potential to reduce costs and improve operational efficiencies for both independent investigators and moderate throughput NGS labs.

## **PgmNr 1841: Consistent nucleic acid extraction is the key to reliable genomic data - and the Fluent® NAP workstation is the ideal tool to support your efforts.**

### **Authors:**

B. Marg-Haufe

[View Session](#) [Add to Schedule](#)

**Affiliation:** Tecan Switzerland AG, Maennedorf, Zurich, Switzerland

---

Having the right samples is crucial to the success of any genetic research and testing workflow. Reliable and unbiased processing ensures clean nucleic acid samples and the best possible results, generating data you can trust without the need to repeat experiments. Converting from manual to automated extraction is the best way to relieve the common bottlenecks and pain points in sample preparation, helping to improve performance and consistency – from run to run, operator to operator, and lab to lab. This provides higher quality samples for downstream applications, ranging from qPCR to various types of arrays or NGS.

The Fluent NAP workstation offers reliable processing of up to 96 individual samples, with full tracking throughout the process, plus optional quantification and normalization ready for downstream processes. The system's universal set-up allows you to perform a variety of magnetic bead-based extraction workflows without rearranging the worktable, while the intuitive interface provides on-screen prompts to guide user interactions and allow straightforward process customization. All of these features seamlessly work together to help give you consistent results, while easing the time and energy required to transition your lab to an automated workflow.



## **PgmNr 1842: A highly multiplexed target enrichment approach for sample identification and tracking using the NEBNext Direct Genotyping Solution.**

### **Authors:**

A.B. Emerman <sup>1</sup>; B.S. Desmond <sup>1</sup>; K.M. Patel <sup>1</sup>; S.K. Bowman <sup>1</sup>; J.S. Dunn <sup>1</sup>; S.M. Adams <sup>1</sup>; S.E. Corbett <sup>1</sup>; C.E. Elfe <sup>1</sup>; A.J. Barry <sup>2</sup>; T.B. Davis <sup>2</sup>; E. Mauceli <sup>1</sup>; C.L. Hendrickson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Directed Genomics, Ipswich, Massachusetts; 2) New England Biolabs, Ipswich, Massachusetts

---

Next-generation sequencing is increasingly being adopted for genetic screening and clinical diagnostics. To prevent false reporting of results, it is imperative that patient samples are tracked throughout sample processing and data analysis. A reliable method to track sample identity throughout a workflow is to monitor single nucleotide polymorphisms (SNPs) that are highly discriminatory across individuals. In order to incorporate a routine sample tracking method into diagnostic workflows, the method should be reliable, high-throughput, and cost-effective.

To address the need for high-throughput genotyping assays, we developed the NEBNext Direct Genotyping Solution. This approach enables multiplexing of up to 96 samples in a single hybridization reaction that targets between 100 to 5000 SNPs. Here we demonstrate the power of this approach to distinguish 24 unique human samples from each other using a sample identification panel of highly discriminatory SNP targets. Using this approach, minimal sequencing reads were required per sample to obtain sufficient data for germline variant calling. With a one day target enrichment and library preparation protocol and an approximately 12 hour sequencing strategy, we went from DNA samples to data in less than 24 hours. Our approach offers a convenient and reliable method to ensure that data integrity is maintained in a diagnostic workflow.

## PgmNr 1843: Comparative purification methodologies and synthetic long-read sequencing for fecal microbiome identification.

### Authors:

C. Knox; D. Wieczorek; D. Horejsh; C. D'Jamoos; C. Cowles; T. Schagat

[View Session](#) [Add to Schedule](#)

**Affiliation:** Promega, Madison, Wisconsin.

---

Study of the human microbiome is rapidly expanding as we further our understanding of the relationship between human health and gut microbiota. Thus, there is an increased need for protocol optimization to more accurately see what organisms are present. Here, we study two aspects of the protocol: comparison of purification methods and the use of a synthetic long-read sequencing library prep method for improved species level identification.

DNA was purified via three extraction methods from replicate fecal samples and a control sample. Samples were prepared for 16S microbiome sequencing using a novel library prep method that provides comprehensive profiling of all 9 variable regions while incorporating barcode-based molecular counting designed to algorithmically combine short reads into a full-length molecule using linked-read de novo assembly. Two separate chemistries for size-selection of libraries were also compared. Illumina sequencing was performed and results across methods were compared.

In fecal samples, variation in microbiome levels were virtually indistinguishable between size-selective purification chemistry, but showed moderate distinction depending on purification method, with two methods (fecal protocols from Promega Maxwell® RSC PureFood GMO and Authentication kit and ReliaPrep™ gDNA Tissue Miniprep System) in close agreement but diverging from the Qiagen QIAmp PowerFecal purification kit. When comparing results utilizing a microbial community standard, the various purification methods were indistinguishable, but all three methods deviated significantly from the theoretical mix as reported by the manufacturer. These results were likely due to the control being comprised of pre-lysed organisms, negating differences in pre-processing between methods. All methods produced synthetic read lengths approaching the full length 16S maximum of 1500bp when combined with the novel library prep method on standard sequencing platforms, with most organisms identified at the species or strain level.

Divergence from theoretical mixes in a community standard demonstrates the need for careful selection of proper controls when evaluating purification methods. High prevalence microbes are easily detectable in all methods, but rare organisms can be potentially lost depending on purification methods. Full length 16S sequences can be achieved without the need for long-read sequencing platforms, providing greater detail and discrimination within a microbiome.

## PgmNr 1844: The SCALLOP INF Consortium: Mapping the genetic determinants of 91 inflammation-related plasma proteins in 15,335 individuals.

### Authors:

J. Peters<sup>1,2</sup>; J. Zhao<sup>1</sup>; E. Macdonald-Dunlop<sup>3,4</sup>; N. Eriksson<sup>5</sup>; L. Folkerson<sup>6</sup>; A. Kalnapenkis<sup>1,7</sup>; A.K. Hedman<sup>8,9</sup>; P.K. Joshi<sup>3,4</sup>; M. Nagle<sup>10</sup>; J. Marten<sup>1</sup>; Y. Wu<sup>10</sup>; E. Ingelsson<sup>11,12,13</sup>; A. Siegbahn<sup>5,14</sup>; T. Esko<sup>7,15</sup>; M. Landen<sup>16</sup>; S. Enroth<sup>17</sup>; L. Padyukov<sup>9</sup>; A. Johansson<sup>17</sup>; O. Hansson<sup>18</sup>; L. Wallentin<sup>5,14</sup>; J. Danesh<sup>1,2</sup>; A. Mälarstig<sup>10,19</sup>; J.F. Wilson<sup>3,4</sup>; A.S. Butterworth<sup>1</sup>; on behalf of the SCALLOP Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, United Kingdom; 2) Health Data Research UK, United Kingdom; 3) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom; 4) Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, United Kingdom; 5) Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden.; 6) Institute of Biological Psychiatry, Copenhagen, 2000, Denmark; 7) Estonian Genome Center, University of Tartu, Estonia; 8) Pfizer Worldwide Research and Development Stockholm, Sweden; 9) Division of Rheumatology, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden; 10) Pfizer Inc, USA; 11) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA 94305.; 12) Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305.; 13) Stanford Diabetes Research Center, Stanford University, Stanford, CA 94305.; 14) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 15) Broad Institute of MIT and Harvard; 16) University of Gothenburg, Sweden; 17) Uppsala University, Sweden; 18) Lund University, Sweden; 19) Karolinska Institute, Stockholm, Sweden

---

Inflammation is a key host response to infection or injury. However, pathological chronic inflammatory responses occur in immune-mediated diseases, such as rheumatoid arthritis or inflammatory bowel disease (IBD). In addition, there is growing recognition of the role of chronic inflammation in disorders not classically considered autoimmune, such as cardiovascular disease. Understanding the effects of genetic variation on inflammation-related proteins will therefore provide valuable insights into disease aetiology and inform development of novel therapies. In addition, there is now a growing armamentarium of biological therapies capable of targeting specific inflammatory proteins, providing opportunities for drug repurposing.

Knowledge of the genetic determinants of the human plasma proteome has until recently been limited by lack of suitable high-throughput proteomic assays, but technological advances now allow protein measurement at population scale. Here, we report data from the SCALLOP consortium, a multi-national collaboration which seeks to combine genomic and proteomic information. We performed genome-wide protein quantitative trait locus (pQTL) mapping for 91 inflammation-related plasma proteins measured using Olink immunoassays in a meta-analysis of 15,335 individuals. We identify at least 1 pQTL for 71 out of 91 proteins; 12 had *cis* pQTLs only, 14 *trans* only, and 45 both *cis* and *trans* ( $P < 5 \times 10^{-10}$ ). The majority of the 20 proteins with no detectable pQTLs had low abundance in

plasma. We identified up to 8 pQTL regions per protein, with conditional analysis revealing multiple additional independent variants. We use colocalisation and Mendelian randomisation techniques to integrate these data with disease susceptibility information from genome-wide association studies to provide new insights into disease aetiology. For example, we found 1 *cis* and 6 *trans* pQTLs for IL12B, including at IBD- and psoriasis-associated variants, mirroring the clinical efficacy of IL12/23 blockade (ustekinumab) in these conditions. This study provides a resource for understanding complex disease traits and an example of the application of novel bioassay technologies to population-scale cohorts.

## PgmNr 1845: Improving transcriptome analysis by incorporating unique molecular identifiers in RNA-sequencing.

**Authors:**

D. Posfai; K. Krishnan; C. Song; P. Liu; G. Naishadham; B. Langhorst; E. Dimalanta; T. Davis

[View Session](#) [Add to Schedule](#)

**Affiliation:** NEBNext, New England Biolabs, Ipswich, Massachusetts.

---

RNA-sequencing is a powerful tool for the study of gene regulation and function. Over the years, applications of RNA-sequencing (i.e. single-cell, rare cell type characterization) have vastly expanded by improving library preparation methods that require lower RNA inputs. Although analyzing samples with low RNA inputs allows for a more diverse probe of biological questions, generating these libraries can also introduce greater biases. Generating sequencing libraries with sufficient yields also requires a greater number of PCR amplification cycles after adaptor ligation. During this process, however, transcripts are not amplified equally, leading to an over-representation of certain transcripts.

To resolve true transcript abundance, unique molecular identifiers (UMIs) can be used to detect bias that is introduced during the amplification process. UMIs are molecular tags typically of 8-12 random base pairs that allow researchers to distinguish transcript duplicates that originate from a single molecule. Here we introduce UMI-containing barcoded adaptors optimized across various inputs (10 ng - 1 mg) for the NEBNext Ultra II Directional RNA Library Prep for Illumina. Ligation of these adaptors followed by subsequent PCR enrichment produced high-quality library yields and metrics (i.e. insert size, 5'-3' coverage, and GC bias) consistent with established protocols. In addition to these metrics, use of the newly developed UMI-containing adaptors allowed for detection of duplication rates and correction of differential amplification artifacts. This workflow enables a more quantitative transcript level assessment that can be applied across various studies and applications for a more biologically meaningful quantification of transcripts.

## PgmNr 1846: Effect of RNA editing on molecular phenotypes through a multi-omics analysis.

### Authors:

A. Belkadi; G. Thareja; A. Halama; S. Zaghloul; Y. Mahmoud; J. Malek; K. Suhre

[View Session](#) [Add to Schedule](#)

**Affiliation:** Research, Weill Cornell medicine in Qatar, Doha, Qatar

---

RNA editing (RNAe) is a post-transcriptional process consisting in the enzymatic modification of one nucleotide encoded by the genomic sequence resulting in another nucleotide in the transcriptomic sequence. Hundreds of thousands of RNAe sites are well established in humans and other species. It has been proposed that the function of RNAe in vertebrates is to edit endogenous long double stranded RNAs to prevent the innate immune from misidentifying them as viral RNA. We used a bioinformatics approach to detect RNAe sites and estimated RNAe levels in two independent RNA-seq datasets: the Geuvadis dataset composed of 421 Europeans and Africans and original data from the QMDiab study, composed of 320 participants of Arab, Filipino and Indian ethnicities. We conducted a *cis*-association study with RNAe levels in Geuvadis and replicated 534 associations (RNAe-QTLs) in QMDiab. We observed a clear allele-specific editing associated with the two alleles of an RNAe-QTL. Moreover, the RNA structure prediction shows the potential of RNAe-QTL in generating a hairpin loop in the double stranded RNA structure. A total of 83 RNAe-QTLs overlapped with QTLs for immune cell count of eosinophils, neutrophils and lymphocytes, and also with risk of inflammatory diseases including Asthma, Crohn's disease and inflammatory bowel diseases. We investigated the correlation of RNAe with the levels of over 1,100 blood circulating proteins and discovered that the most correlated proteins were related to the immune response. We then assessed the correlation between the RNAe and the white blood cell count. We found that RNAe was positively correlated with neutrophil count and negatively correlated with lymphocyte count. We further tested for association of RNAe level with the DNA-methylation derived estimates of Natural killer cells (NK), T helper cells (CD4), Cytotoxic T cells (CD8) and B cells. We found that most of RNAe positively correlated with NK and B-cells and negatively correlated with CD4 and CD8, or inversely. We finally assessed the correlation of RNAe level for RNAe belonging to an editing enriched region and detected the presence of negatively correlated RNAe sites. We explored these cases in detail and found that they were controlled by either the same RNAe-QTL with opposite Beta or a potential CpG site. Our study is the first to study the RNAe-QTL in two independent datasets and to shed light on the correlation between RNAe and other phenotypes in a multi-omics approach.

## PgmNr 1847: Semi-automated RNA extraction from platelets for disease panel expression profiling.

### Authors:

S. Teter <sup>1</sup>; C. Knox <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Scientific Applications, Promega, Madison, WI.; 2) Promega, Madison, WI

---

Platelets are anucleate cell fragments in blood, which have long been recognized for their importance in blood clotting and wound healing. Beyond these curative roles, platelets also function in disease pathology. As a result of interactions with stimuli encountered during their circulation, the RNA expression patterns in platelets are altered in ways that are characteristic of certain diseases, including inflammatory diseases, like Alzheimer's disease and cardiovascular disease, sickle cell anemia, and cancer. Additionally, by taking up vesicular RNA from solid tumors, platelets have also been shown to become "tumor-educated," carrying cancer-derived RNA biomarkers. There is growing interest in using platelet biomarkers and expression patterns for liquid biopsy rather than more invasive surgical biopsy procedures for disease diagnosis and monitoring. With this in mind, we developed a semi-automated protocol for total RNA extraction from platelets on a unique robotic platform utilizing single-use cartridges. We analyzed RNA expression by RNA-sequencing with a custom-designed panel of RNA targets whose expression is implicated in disease pathology. We compared expression profiles for RNA extracted from platelets and white blood cells collected from two healthy individuals. Clear differences in RNA expression profiles of white blood cells and platelets demonstrate that the platelet expression profile in our study is not masked by that of the RNA-rich white blood cells. While the platelet RNA expression generally looked similar between the two healthy individuals, there are some notable differences for specific genes. This work demonstrates that an RNA-sequencing workflow for platelet expression studies can be simplified by automation of upfront RNA extraction.

## PgmNr 1848: High-throughput encapsulation of single cells in agarose microdroplets for 3D culture.

**Authors:**

T. Ayers; N. Shakir

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dolomite Bio, Royston, Hertfordshire, United Kingdom

---

Hydrogel droplets are picolitre-volume spherical scaffolds. They represent a potent solution for many single cell applications owing to their properties which allow the diffusion of nutrients and dissolved gases to circulate and reach encased cells. Encapsulating biological material within hydrogels allows cells to be grown within three-dimensional scaffolds, closely mimicking their native physiological environment, whereas conventional plate-based cell culture methods grow cells in a two-dimensional plane with relatively low throughput. This property of hydrogels is useful in the field of 3D *in vitro* cell and tissue culture. Cells can be grown in individual hydrogel beads or seeded onto 3D scaffolds for larger scale tissue synthesis. Importantly, 3D *in vitro* tissue culture requires monodisperse hydrogel beads to enforce their predictable stacking into a geometric tetrahedral 3D structure. The ability of the Dolomite Bio Nadia product family to generate highly monodisperse aqueous droplets in an oil phase makes this platform ideal for this application.

Experiments were carried out to, firstly, define run conditions such as pressures, stirrer speeds and temperatures that allowed for stable encapsulation of agarose taking advantage of the versatility of the Nadia Innovate platform. Secondly, we aimed to demonstrate the high-throughput encapsulation of live cells in agarose and the recovery of agarose beads from the resultant emulsion produced on both the Nadia Innovate and the Nadia Instrument.

Agarose hydrogels were successfully produced after adjusting pressures to accommodate the viscosity of a 1.5% agarose solution. The Nadia Innovate Temperature Control Unit enabled the system to be kept at 37°C during droplet formation to ensure that agarose remained molten and could flow freely into the Nadia Innovate microfluidic chip. Following optimisation of the run conditions, live cells were encapsulated in agarose and recovered from the generated emulsion. In conclusion, the Nadia Innovate and Nadia Instrument were used to produce solid agarose beads containing cells that were subsequently recovered and imaged. The beads were monodisperse and their diameter could be changed by altering pressures to ensure compatibility with cells of different sizes. The flexibility afforded by the Nadia product range therefore enhances the potential of hydrogel cell encapsulation in the field of tissue engineering.



## PgmNr 1849: Targeted interrogation of genetic variants in *CFTR* with NEBNext Direct.

### Authors:

B.S. Desmond <sup>1</sup>; A.B. Emerman <sup>1</sup>; K.M. Patel <sup>1</sup>; S.K. Bowman <sup>1</sup>; S.M. Adams <sup>1</sup>; J.S. Dunn <sup>1</sup>; S.E. Corbett <sup>1</sup>; C.D. Elfe <sup>1</sup>; E. Mauceli <sup>1</sup>; A. Barry <sup>2</sup>; T.B. Davis <sup>2</sup>; C.L. Hendrickson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Directed Genomics, Ipswich, MA.; 2) New England Biolabs Inc., Ipswich, MA.

---

The screening and detection of variants in the cystic fibrosis transmembrane conductance regulator (CFTR) are critical for the effective management of cystic fibrosis. More than 2,000 *CFTR* mutations have been reported worldwide and refined sequencing technologies may provide improved genotype-phenotype correlations and better-informed genetic counseling.

Here, we introduce an NEBNext Direct target enrichment panel for the interrogation of genetic variants in the *CFTR* gene. The NEBNext Direct technology uses a novel approach to selectively enrich nucleic acid targets directly from genomic DNA. The approach rapidly hybridizes both strands of genomic DNA to biotinylated baits, captures the targets on streptavidin beads, enzymatically removes off-target sequence, and directly converts captured molecules into Illumina-compatible sequencing libraries in a single day protocol.

The NEBNext Direct CFTR Panel consists of a single pool of baits targeting both strands of DNA across 27 exons (including padding to detect splice mutations) and 7 intronic sites, for a total captured territory of 6.751 kilobases. The panel covers all 346 CF-causing variants listed in the March 11, 2019 release of CFTR2. Here we demonstrate the utility of this panel to detect germline variants in the *CFTR* gene in clinical samples with documented mutations.

## PgmNr 1850: High performance multiplexed target enrichment sequencing from FFPE tissues.

### Authors:

R. Gantt<sup>1</sup>; L. Arbiza<sup>1</sup>; K. Butcher<sup>1</sup>; S. Chen<sup>1</sup>; H. Chilton<sup>1</sup>; M. Consugar<sup>1</sup>; Y. Hasin-Brumshtein<sup>1</sup>; S. Gude<sup>1</sup>; J. Laugharn<sup>2</sup>; J. Simon<sup>2</sup>; U. Thomann<sup>2</sup>; C. Thompson<sup>1</sup>; M. Werner<sup>2</sup>; R. Zeitoun<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Twist Bioscience Corporation, San Francisco, California; 2) Covaris Inc, Woburn, Massachusetts

---

Library construction for Next Generation Sequencing (NGS) using formalin-fixed paraffin-embedded (FFPE) samples offers unique challenges in acquiring high-quality sequencing data due to wide distribution of sample quality. Specifically, differences in formalin fixation methods often lead to crosslinked and/or degraded nucleic acid and inconsistent extraction yields. Hence, FFPE extraction and library construction methods must be carefully considered for target enrichment applications. In collaboration, Covaris and Twist Bioscience demonstrate a complete library preparation and target enrichment solution that generates ready-to-sequence multiplexed libraries directly from FFPE tissue.

This workflow leverages the Covaris truXTRAC FFPE total Nucleic Acid Plus Kit and Adaptive Focused Acoustics (AFA) technology with the world-class performance of Twist Bioscience's Target Enrichment Solutions. In this FFPE-specific application, the Covaris truXTRAC FFPE total Nucleic Acid Plus Kit and shearing in AFA-TUBE consumables enable full emulsification of paraffin and disaggregation of tissue for highly efficient nucleic acid extraction and generation of size specific DNA libraries. With the Twist Bioscience Human Core Exome kit, the resulting libraries are indexed, pooled, and target enriched with uniquely optimized DNA probes to generate ready-to-sequence high quality multiplexed libraries.

Using the aforementioned workflow, results from processing FFPE tissue types with a wide distribution of quality (Q305/Q41 ratios ranging from 0.06 to 0.35) are presented. Sequencing results demonstrate substantial improvements in general Picard metrics that include uniformity (Fold<sub>80</sub> < 1.8), sequencing depth (30X coverage >85% with 150X downsampling), and duplication rates (<10%) when compared to similar published studies. These results demonstrate a validated solution for library preparation and targeted exome sequencing of FFPE samples that can be integrated into automated workflows. The truXTRAC kit and AFA technology from Covaris generate size specific DNA libraries from FFPE samples which, when paired with Twist Bioscience's superior target enrichment workflow, deliver multiplexed libraries for high performance targeted sequencing.

# PgmNr 1851: Resolution of a highly complex rearrangement of an SV40 insertion in an immortalized cell line using a combination of short and long read sequencing technologies.

## Authors:

A. Sivachenko; E. Wilson; B. Tabak; J. Conte; H. Valley; A. Bell; K. Bukis; E. Wong; N. Allaire; M. Mense

[View Session](#) [Add to Schedule](#)

**Affiliation:** Cystic Fibrosis Foundation Therapeutics Laboratory, Lexington, MA.

16HBE14o- cells are an immortalized human bronchial epithelial cell line that was first described by Cozens and colleagues (AJRCMB, 1994). This cell line is widely used in CF research as it expresses *CFTR* mRNA and protein, possesses essential epithelial morphology, and is suitable for electrophysiological studies in Ussing chambers or similar systems.

Recently, in house deep sequencing of the *CFTR* locus of the parental 16HBE14o- cells revealed an insertion of SV40 sequence in intron 6 of one *CFTR* allele. The cell line was originally immortalized by transformation with SV40 and whole genome sequencing showed that in fact the only insertion of SV40 sequence is in the *CFTR* allele. All isogenic cell lines created by CRISPR/Cas9 gene editing of the parental cells also contain the same SV40 insertion, suggesting that the insertion is stable and required to maintain the 16HBE14o- immortalization. Given the widespread use of the parental and gene-edited 16HBE cell lines, a complete characterization of the genomic locus and quantification of the mRNA isoforms induced by the SV40 insertion was needed.

To elucidate the full sequence of the insertion, two novel target enrichment strategies were employed: (1) capture of SV40/*CFTR* using biotinylated baits and (2) Cas9 digestion ~500bp up and downstream of the SV40 insertion followed by Oxford Nanopore Minion long read sequencing. Additionally, RNAseq using Illumina-based short read sequencing, as well as a *CFTR* strand switching isoform assay followed by long read sequencing were used to fully characterize the expressed mRNA isoforms.

Initial analysis relying on available automated hybrid assembly tools using short and long reads failed to completely resolve the insertion and resulted in a stuffed gap assembly. We completed the assembly manually utilizing the information provided by paired end and long reads and determined the inserted sequence to be 16 kB and to contain highly rearranged and repeated SV40 viral genome and vector backbone sequences. Using short read RNAseq data, normal *CFTR* mRNA and fusion products of *CFTR* intronic/coding and SV40 sequence were identified. We are currently using long read data to quantify expressed isoforms. However, allelic discrimination analysis and western blot analysis of a gene-edited heterozygous 16HBE cell line suggest that the allele carrying the SV40 insertion is non-functional and therefore the 16HBE14o- parental and derived cell lines are functionally monoallelic for *CFTR*.

## PgmNr 1852: Clariom™ GO Screen Assay: Transcriptional profiling from direct cellular lysates using 384-well array plates.

### Authors:

P.G. Rack; C. Diaz; M. Staples; J. Hubbard; A. Schweitzer; M. Shapero

[View Session](#) [Add to Schedule](#)

**Affiliation:** Thermo Fisher Scientific, Santa Clara, CA.

---

Genome wide profiling of cellular transcription can currently be done with a variety of lab methods. However challenges still exist for cost effective, high throughput workflows. Here an overview of recent advances related to the Clariom GO Screen Assay is presented.

The Human Clariom™ GO Screen array is a 384-well array plate with ~60,000 probes used to measure ~20,000 genes associated with GO (Gene Ontology Consortium). There are in addition array processing controls, housekeeping genes, gender linked genes and human viral content. The array targets the most constitutively expressed exons which are least affected by alternative splicing. Each gene is assayed with three probes selected for specificity and uniqueness, as well as performance across empirical data. The assay is configured to use cellular lysates or total RNA, precluding the need for RNA extraction but allowing the flexibility for both sample types. GO Screen data can be analyzed in a streamlined fashion utilizing Transcriptional Analysis Console (TAC) Software which allows a user to perform exploratory grouping analysis to find gene expression differences between clusters and underlying structures in data sets.

Following target amplification and labeling, sample hybridization to the 384-array plate, and subsequent wash/stain steps are performed on the Gene Titan Multichannel instrument minimizing hands-on time. Array performance characterized using cell line lysates demonstrates that lysates ranging from 10-1000 cells perform in a comparable manner to total RNA. Furthermore, lysate samples are stable at 4°C for at least eight hours which is useful in applications where sample collection may take course over time (e.g. Fluorescence Activated Cell Sorting). For larger studies where multiple 96- or 384- cell culture plates may be needed, lysates can be stored at -80°C for greater than two months with no measurable decrease in performance.

In summary, the Clariom™ GO Screen Assay provides a complete workflow from sample to answer for high-throughput interrogation of gene-level expression, precluding the need for RNA extraction and the reliance on imputation from a limited number of transcriptional events. Results presented here suggest utility in a large range of applications including, but not limited to, secondary compound library screening to prioritize candidates for lead optimization.

## PgmNr 1853: Exome re-analysis approaches for solving a large number of rare diseases in Solve-RD.

### Authors:

S. Beltran-Agullo <sup>1,2</sup>; R. Joshi <sup>1</sup>; A. Hoischen <sup>3,4,5</sup>; S. Laurie <sup>1</sup>; D. Piscia <sup>1</sup>; B. Zurek <sup>6</sup>; S. Hermanns <sup>6</sup>; A. Papakonstantinou <sup>1</sup>; L. Matalonga <sup>1</sup>; H. Díez <sup>1</sup>; C. Garcia <sup>1</sup>; M. Fernández <sup>1</sup>; C. Gilissen <sup>3,4</sup>; S. Ossowski <sup>6</sup>; M.A. Swertz <sup>7,8</sup>; A. Rath <sup>9</sup>; R. Schüle <sup>10</sup>; M. Synofzik <sup>10,11</sup>; L. Vessiers <sup>3,12</sup>; R. de Voer <sup>3,4</sup>; A. Topf <sup>13,14</sup>; N. Hoogerbrugge <sup>3,4</sup>; A. Verloes <sup>15</sup>; T. Evangelista <sup>16</sup>; G. Bonne <sup>17</sup>; H. Scheffer <sup>3,18</sup>; A.J. Brookes <sup>19</sup>; H. Brunner <sup>3,12,18</sup>; O. Riess <sup>6,20</sup>; H. Graessner <sup>6,20</sup>; Solve-RD Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics Analysis Group, Centro Nacional de Análisis Genómico (CNAG), Barcelona, Catalonia, Spain; 2) Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands; 4) Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands.; 5) Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, the Netherlands.; 6) Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany.; 7) Department of Genetics, Genomics Coordination Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.; 8) Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.; 9) INSERM, US14 ? Orphanet, Plateforme Maladies Rares, 75014 Paris, France.; 10) Department of Neurodegeneration, Hertie Institute for Clinical Brain Research (HIH), University of Tübingen, Tübingen, Germany.; 11) German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany.; 12) Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, the Netherlands.; 13) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK.; 14) John Walton Muscular Dystrophy Research Centre, MRC Centre for Neuromuscular Diseases, Newcastle University, Newcastle upon Tyne, UK.; 15) Department of Genetics, APHP-Robert DEBRE University Hospital, Denis Diderot School of Medicine, Paris University, France; INSERM UMR1141, Paris, France.; 16) Institut de Myologie, CHU Pitié-Salpêtrière, Paris, France.; 17) Sorbonne Université, INSERM UMRS\_974, Center of Research in Myology, 75013 Paris, France.; 18) Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, the Netherlands.; 19) Department of Genetics and Genome Biology, University of Leicester, Leicester, UK.; 20) Centre for Rare Diseases, University of Tübingen, Tübingen, Germany.

---

Solve-RD – “solving the unsolved rare diseases” – aims to solve a large number of rare diseases with unknown molecular cause and improve diagnostics of patients. The project, funded by the European Commission, is applying a sophisticated combination of ‘omics approaches coupled with shared knowledge about genes, genomic variants and phenotypes. As a first milestone, Solve-RD has already collated over 5,000 exomes and genomes from undiagnosed families across Europe through contributions from the four “core” European Reference Networks (RND, EURO-NMD, ITHACA, GENTURIS) and other associated networks, such as UDPs from Italy and Spain. Collation of clinical and phenotypic information is done with a PhenoTips instance linked to the RD-Connect Genome-Phenome Analysis Platform (GPAP) using standards such as HPO, ORDO and OMIM. Standard processing and analysis of the genomic data is done with the GPAP, while custom and innovative approaches are

being setup on the Solve-RD Sandbox, an HPC-like cloud environment enabling collaboration between project partners. Each of the ERNs has setup a Data Interpretation Task Force (DITF), and has proposed several Use Cases for analysing the data. The prioritised Use Cases for exome re-analysis have been distributed across 5 Working Groups within the Data Analysis Task Force (DATF), dealing with individual's relatedness, runs of homozygosity, consanguinity, SNVs, InDels, CNVs, *de-novo* mutations, mosaicism and mutational burden. In the first screening of the data we are prioritising specificity to "solve" the easier cases. We leverage on all the genotypes being annotated and indexed in the GPAP, allowing for real-time queries to a high number of samples. We have tested several combinations of filters on 147 individuals with confirmed causative variants. The filters tested include, amongst other, ClinVar classification, gnomAD frequency, internal GPAP frequency, SNPeff classification and association with HPO terms linked to the specific case. We have settled on a combination with a specificity of 39% and a sensitivity of 35%. The same combination of filters applied to the undiagnosed cases from Solve-RD has yielded a similar number of variants per case as in the test group. The outcome of this approach will be presented at the ASHG meeting. Generation of novel 'omics data from selected cohorts and cases is expected to start at the end of 2019, enabling more comprehensive and innovative approaches.

## PgmNr 1854: Comparing whole exome sequencing probes at the sample and exon level for the human genome.

**Authors:**

C. Vaccaro; F. Mafra; J. Garifallou; J. Smiler; N. Williams; A. Caro; D. Li; L. Tian; P. Sleiman; A. Santani; R. Pellegrino; H. Hakonarson

[View Session](#) [Add to Schedule](#)

**Affiliation:** The Children's Hospital of Philadelphia 3615 Civic Center Boulevard Abramson Research Center Philadelphia, PA 19104

---

**Introduction:** The ample use of next generation sequencing across laboratories has pushed this technology to become a normality in most research and clinically-based applications today. Due to the higher cost of whole genome sequencing, whole exome sequencing has been a good alternative to still generate the detailed data that sequencing can provide, but at a lower expense. A sufficient capture must be comprehensive, while keeping the overall capture size low enough to keep sequencing costs low. In addition, the capture must produce high on target results, with low duplication rates, and coverage uniformity throughout the targeted region.

**Materials/Methods:** In our study, we tested a total of twenty unique samples against six different captures from three vendors. Amongst the cohort, there were HapMap, positive controls, and FFPE samples. The samples were processed using vendor-specific library preparation kits and were prepared manually or by automation. All samples were sequenced on the Illumina HiSeq 2500 or the NovaSeq. To reduce bias amongst sequencing data, all alignment files were down-sampled between the range of 50-65X mean sample coverage. To compare the captures, we focused on the percent of target bases at 20X across three different interval sets. These sets included the intersected region between all captures in the study, HGMD variants, and an in-house curated gene list. We analyzed further by comparing the per sample statistics including percent on target, duplication rates, and capture uniformity. Tools used to generate the data include Illumina DRAGEN Bio-IT Platform, GATK, and BEDTools.

**Results:** Of the six captures, the Twist Human Core Exome probes performed the most efficient in all aspects of the study. Specifically, we calculated this capture to produce the lowest number of regions below 100% at 20X for all three interval sets (less than 6% for all data sets). In addition, the Twist Human Core Exome kit displayed the highest percent on target values at >70%, lowest duplication rates below 10%, and greatest uniformity across the target region.

**Conclusion:** The Twist Human Core Exome probes offer a versatile capture that can be used for multiple applications. The sequencing data shows promising results at all levels of the gene, especially for on-target, duplication rates, and capture uniformity.

## **PgmNr 1855: BCH Connect: Supporting patient diagnoses with a streamlined CLIA-compliant research sequencing pipeline.**

### **Authors:**

S. Rockowitz <sup>1</sup>; P. Sliz <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Boston Children's Hospital, Boston, Massachusetts.; 2) Harvard University, Cambridge, MA

---

Though researchers and clinicians both generate and analyze whole exome and whole genome sequencing data for the purposes of understanding patient disease, there exists an iron curtain between research and clinical pools of genomic data. Research sequencing is generally not done to clinical standards and raw sequencing data remains siloed to the research environment where it might be shared with other researchers, but not clinicians. Findings from research sequencing must be confirmed and cannot contribute directly to clinical reports, and patients rarely, if ever, are returned results or raw research sequencing data.

Clinical sequencing data is paradoxically both more and less accessible. As clinical findings are collected in compliance with Clinical Laboratory Improvement Amendments (CLIA), they are regularly used in the process of generating clinical reports that are returned to the electronic health record and can inform precision medicine. Though the clinical reports are readily available, raw clinical sequencing data is only available for re-analysis (outside of the sequencing company) when approved by the patient. Thus, re-analysis of clinical sequencing data by clinicians or researchers occurs rarely.

Here we describe mechanisms, processes, and technologies implemented at our institution, Boston Children's Hospital (BCH), to enable researchers and clinicians to accelerate rates of discovery and patient diagnoses. Through strategic investment in clinical-grade research sequencing of disease specific cohorts, we have harmonized the sample collection, ordering, and data return processes for research groups across our institution with a prudent, CLIA-compliant workflow. We have integrated research sequencing with clinical follow up; aligned, refined, and operationalized our earlier patient consenting approaches; and streamlined processes for migrating clinically collected genomic data into repositories where BCH clinical and research staff can re-analyze patient data.

By merging research data, electronic health records, and genomic data into repositories with cutting edge technologies, we are empowering researchers and clinicians to access highly curated candidate variants and collaborate with one another more rapidly to discover genes and diagnose patients.



## PgmNr 1856: A large library of UAS-human cDNA constructs and transgenic *Drosophila* stocks to facilitate molecular diagnosis and translational research.

### Authors:

S. Yamamoto <sup>1,2,3</sup>; K.H. Wan <sup>4</sup>; T. Johnson <sup>5</sup>; H.K. Graves <sup>1</sup>; V.H. Bhavana <sup>1</sup>; H. Pan <sup>6</sup>; D. Bei <sup>1</sup>; O. Kanca <sup>1</sup>; S. Jangam <sup>1</sup>; S.I. Park <sup>4</sup>; W.W. Fisher <sup>4</sup>; Y. Takashima <sup>4</sup>; S. Ravani <sup>4</sup>; M. Tomaru <sup>7</sup>; T. Ohsako <sup>7</sup>; M.F. Wangler <sup>1,2</sup>; C. Warr <sup>8</sup>; T. Takano-Shimizu <sup>7</sup>; H.J. Bellen <sup>1,2,3,6</sup>; S.E. Celniker <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine (BCM), Houston, TX; 2) Department of Neuroscience, BCM, Houston, TX; 3) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX; 4) Berkeley Drosophila Genome Project, Lawrence Berkeley National Laboratory, Berkeley, CA; 5) School of Biological Sciences, Monash University, Melbourne, Australia; 6) Howard Hughes Medical Institute, Houston, TX; 7) Drosophila Genetic Resource Center, Kyoto Institute of Technology, Kyoto, Japan; 8) School of Medicine, University of Tasmania, Hobart, Australia

---

Whole-exome and whole-genome sequencing technologies have revolutionized human genomics research and clinical diagnosis, leading to identification of hundreds of new disease causing genes and pathogenic variants. At the same time, these assays are revealing thousands of variants of uncertain significance (VUS) in known disease genes that require functional assessment. Moreover, variants in genes of unknown significance (GUS) are uninterpretable due to lack of gene function information in human.

To facilitate functional annotation of human genes and variants, we are generating a large collection of UAS constructs and transgenic *Drosophila melanogaster* (fruit fly) strains based on sequence validated full length human cDNAs using the  $\phi$ C31 transgenesis system. So far, we have generated >3,000 constructs and established >1,000 transgenic lines that are available from the *Drosophila* Genomics Resource Center (DGRC) and Bloomington (BDSC)/Kyoto (DGGR) Stock Centers, respectively. These reagents can be used to study the function of a genetic variant linked to a human disease by "humanizing" the fly ortholog through the *T2A-GAL4* gene trap system, or through tissue specific over-expression of reference and variant human cDNAs. By further probing the *in vivo* function of these genes in flies and translating these findings to mammalian systems, *Drosophila* studies can facilitate the understanding of biological mechanisms underlying genetic disorders in human to develop effective therapies.

# PgmNr 1857: The impact of DNA source on genetic variant detection from human whole-genome sequencing data.

## Authors:

B. Trost<sup>1</sup>; S. Walker<sup>1</sup>; S.A. Haider<sup>1</sup>; W.W.L. Sung<sup>1</sup>; S.L. Pereira<sup>1</sup>; C.L. Phillips<sup>1</sup>; E.J. Higginbotham<sup>1,2</sup>; L.J. Strug<sup>1,3,4</sup>; C. Nguyen<sup>1,2</sup>; A. Raajkumar<sup>1</sup>; M.J. Szego<sup>4,5</sup>; C.R. Marshall<sup>6,7</sup>; S.W. Scherer<sup>1,2,8</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 3) Department of Statistical Sciences, University of Toronto, Toronto, Canada; 4) Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 5) Department of Family and Community Medicine, University of Toronto, Toronto, Canada; 6) Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Canada; 7) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada; 8) McLaughlin Centre, University of Toronto, Toronto, Canada

---

## Background

Whole blood is currently the most common DNA source for whole-genome sequencing (WGS), but for studies requiring non-invasive collection, self-collection, greater sample stability, or additional tissue references, saliva or buccal samples may be preferred. However, the relative quality of sequencing data and accuracy of genetic variant detection from blood-, saliva-, or buccal-derived DNA need to be thoroughly investigated.

## Methods

Matched blood, saliva, and buccal samples from four unrelated individuals were used to compare sequencing metrics and variant-detection accuracy among these DNA sources.

## Results

We observed significant differences among DNA sources for sequencing quality metrics such as percentage of reads aligned and mean read depth (P-value < 0.05). Differences were negligible in the accuracy of detecting short insertions and deletions; however, the false positive rate for single nucleotide variation detection was slightly higher in some saliva and buccal samples. Sensitivity for copy number variation (CNV) detection was up to 25% higher from blood samples, depending on CNV size and type, and appeared to be worse in saliva or buccal samples with high bacterial concentration. We also show that methylation-based enrichment for eukaryotic DNA in saliva and buccal samples increased alignment rates, but also reduced read-depth uniformity, hampering CNV detection.

## Conclusion

For WGS, we recommend using DNA extracted from blood rather than saliva or buccal swabs; if saliva or buccal samples are used, we recommend against using methylation-based eukaryotic DNA enrichment. The data used in this study are fully available for further open-science investigation.

## PgmNr 1858: Turbolase: A rapid, high-throughput sample prep to sequencing workflow for targeted and whole-genome sequencing.

### Authors:

N. Abdel-Magid; D. Masser; R. Stedtfeld; J. Lenhart; V. Makarov; L. Kurihara

[View Session](#) [Add to Schedule](#)

**Affiliation:** Swift Biosciences, Ann Arbor, MI.

---

Streamlined NGS library preparation saves time and reduces errors allowing for cost-efficient, high-throughput sample processing. We have developed a workflow that combines a rapid library prep, a novel library normalization methodology, and optimized hybridization capture panels for an efficient targeted sequencing library prep solution. Swift 2S Turbo kits comprise two enzymatic steps and a single purification to complete DNA fragmentation, end repair, and adapter ligation. This is followed by standard library amplification using Normalase PCR primers followed by a single purification and two 15-minute Normalase incubation steps that enzymatically select the desired amount of each library and normalize each library to an equimolar concentration within a single pool. Normalase is an enzymatic library normalization method that eliminates manual concentration adjustment of each sample prior to library pooling. When combined with Swift 2S Turbo library kits, a highly streamlined 'Turbolase' workflow is created that is readily automated, where simple bulk processing improves throughput and reduces cost. The pools can then be target-enriched using Swift Hybridization Capture panels or directly sequenced without further purification. Swift offers a 39Mb Exome Panel, a 0.8Mb Pan-Cancer panel covering 127 oncology-related genes, and a 11.1Mb Inherited Disease panel covering genes and SNPs associated with inherited diseases for targeted sequencing. Use of Normalase in this workflow allows for greater than 10-fold variation in input quantity while generating less than 10% variation in sample representation within a pool resulting in optimal sample balance for multiplexed target enrichment or whole-genome sequencing. Together this workflow simplifies NGS library prep minimizing clean-ups and pipetting steps for a rapid, cost-efficient workflow from library generation to sequencing.

# PgmNr 1859: NGS target enrichment: Automation of IDT xGen hybridization capture of DNA libraries on Beckman Coulter i7 Hybrid Genomics Workstation with KAPA libraries and exome panel.

## Authors:

S. Parakh <sup>1</sup>; N. Roseman <sup>2</sup>; N. Sponer <sup>3</sup>; A. Jackson <sup>1</sup>; N. Downey <sup>2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Automation Biotechnology Business Unit, Beckman Coulter, Indianapolis, Indiana.; 2) Integrated DNA Technologies, Redwood City, CA; 3) Integrated DNA Technologies, Coralville, IA

## Introduction:

Next Generation Sequencing (NGS) has enabled collection of exome-wide information faster than any previous technology. Hybridization capture is an enrichment method that can provide disease specific information tuned to the user needs. From small target spaces used to detect rare variants to exome-wide data, the IDT hybridization system can confidently provide the flexibility needed by researchers. This workflow is compatible with NGS libraries prepared using ligation-based techniques, such as TruSeq<sup>®</sup> library kits, KAPA library kits and tagmentation-based library kits such as Nextera<sup>®</sup>. Automated library preparation allows for higher throughput and less hands-on time for complex NGS assays.

## Method:

The xGen Hybridization Capture automated protocol on Biomek i7 NGS workstation has a throughput of 96 samples and can perform concentration via bead based or dry down method. The deck configuration provides components for plate processing, incubation and thermocycling that enable efficiency and minimal hands-on time. The automated method is written in modular design with safe stop points that enable labs to customize how they work. The method provides flexibility to run any number of samples, ease of setup including reagent calculation and decreased costly setup errors using DeckOptix Final Check software.

## Experiment and Result:

Libraries were prepared with 1ug Coriell NA12878 input into KAPA Hyperprep. Eight libraries were captured using IDT's xGen Hybridization and Wash Kit and xGen Exome Research Panel. 500ng of each library was dried down per *xGen Hybridization Capture of DNA Libraries for NGS Target Enrichment* with an overnight hybridization. Captured libraries were quantified with Qubit<sup>™</sup> ds HS DNA Kit, analyzed on TapeStation HS D1000 tape, and sequenced by IDT on a NextSeq 500 (Illumina). 2x150bp, paired end reads were generated. The flanked-on target percentage was 94 % with a mean insert size of 295 bp. The duplicate percentage was <3% and GC skew 60/40 was 1.2.

## Conclusion:

Automation of xGen Hyb Capture of DNA libraries on the Biomek i7 Hybrid Genomics Workstation is a fast and efficient process that delivers quality results, providing a flexible and scalable solution for any size lab.



## **PgmNr 1860: *Whole exome sequencing is “still” a valuable platform for variant detection in research: A 2019 overview of capture methods and improvements.***

### **Authors:**

R. Pellegrino <sup>1</sup>; F. Mafra <sup>1</sup>; C.K. Vaccaro <sup>1</sup>; J. Garifallou <sup>1</sup>; J. Smiler <sup>1</sup>; N. Williams <sup>1</sup>; A. Caro <sup>1</sup>; D. Li <sup>1</sup>; A. Santani <sup>1,2</sup>; L. Tian <sup>1</sup>; H. Hakonarson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Applied Genomics, Childrens Hospital of Philadelphia, Philadelphia, Pennsylvania.; 2) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

---

**Introduction:** A decade has passed since Next Generation Sequencing (NGS) technologies emerged and became powerful to the research field and later to the clinics. New approaches permitted whole-exome sequencing (WES) or targeted panels to be sequenced faster leading to a higher sensitivity and significant decrease in costs making the technology feasible for a diverse areas in research. The WES has evolved through the years, and the captures have improved exponentially in terms of content (better gene annotations) and chemistry. The ideal capture nowadays must be cost-effective, but also versatile in terms of DNA input, offer better on target and representations of relevant genes with great uniformity.

**Materials/Methods:** We tested a total of twenty unique samples against 6 different captures from three vendors. Amongst the cohort, there were HapMap, positive controls, and FFPE samples. The samples were processed using vendor-specific library preparation kits and were prepared using Perkin Elmer Sciclone automation. All samples were sequenced on the Illumina HiSeq 2500 or the NovaSeq. We aimed for specific metrics for quality control (QC) purposes: capture yield after PCR, traces distribution and molarity. Bioinformatics analysis was performed using DRAGEN (Illumina) and GATK. These sets included the intersected region between all captures in the study, HGMD variants, and an in-house generated curated gene list.

**Results:** We found that of the 6 captures, the Twist Human Core Exome probes performed the most efficient in all aspects of the study when compared to other vendors. Precisely, we calculated this capture to produce the lowest number of regions below 100% at 20X for all three interval sets (less than 6% for all data sets). In addition, the Twist Human Core Exome exhibited the highest percent on target values at >70%, lowest duplication rates below 10%, and greatest uniformity across the target region. Additionally, the time for the capture was significantly reduced and the entire protocol was completed within 2 days.

**Conclusion:** The sequencing data demonstrated significant improvements in uniformity, on-target, duplication rates, and flexibility with DNA yield in comparison with other platforms. The Center for Applied Genomics (CAG) has expanded the sample evaluation for other research areas and we continue to work on the most cost-effective and highest quality capture methods for the WES data generation offered by the center.

## PgmNr 1861: Sample identity quality control in Next Generation Sequencing (NGS) routine diagnostics using KASP.

### Authors:

F. Scharf; J. Romic-Pickl; T. Wohlfrom; A. Benet-Pagès; B. Neitzel; A. Abicht; E. Holinski-Feder

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medizinisch Genetisches Zentrum MGZ, Munich, Germany

---

NGS applications comprise complex protocols with parallel processing of multiple samples. Thus, consideration of possible sample swaps is part of the laboratory best practices as they can occur at every step of the process. We present a quality control approach based on the genotyping of single nucleotide polymorphisms (SNPs) with Kompetitive Allele Specific PCR (KASP) technology and the comparison to NGS data for the verification of sample identities. Based on genotypic frequencies and KASP assay performance, 16 SNPs covered in our NGS capture kit (Agilent SureSelectXT) were selected. To reliably detect sample swaps two criteria have to be fulfilled: first, haplotype overlaps between samples need to be excluded, i.e. no sample has been pooled twice (uniqueness); second, the identity of each sample needs to be confirmed based on the comparison of KASP and NGS data per sample (identity). Due to possible dropouts or erroneous KASP results, some comparisons might not always be possible. Therefore, cutoffs were defined based on a risk-benefit assessment to minimize the probability of missed sample swaps and simultaneously the frequency of unnecessarily rejected samples. The minimal number of successful comparisons and maximum number of mismatches for a reliable confirmation of uniqueness and identity was determined by statistic calculations, taking into account the average dropout and error rate of KASP signals. To address the question of uniqueness, the probability of encountering two samples with the same haplotype over a certain number of SNPs in a certain number of samples was calculated; the cutoffs for identity confirmation are based on the probability of a sample showing one specific haplotype over a certain number of SNPs. Results showed that at least two differing SNPs in all pairwise comparisons of the samples exclude overlaps with >99% certainty. Further, with a cutoff of maximally four dropouts of SNPs or three dropouts and one mismatch in the identity confirmation step the correct sample assignment can be verified with a probability of >99%. The identity confirmation is integrated in our NGS analysis pipeline. Samples failing the criteria are automatically identified and held back from further processing (variant calling, annotation and assessment) until the case is resolved. Performance of SNPs in KASP is continuously tracked for quality management and the list of SNPs can easily be adapted, making the method flexible and highly scalable.

## PgmNr 1862: Photodamage of single dye-stained DNA during in vitro fluorescent imaging.

**Authors:**

J. Chen

[View Session](#) [Add to Schedule](#)**Affiliation:** Department of Chemistry & Biochemistry, Ohio Univ, Athens, Ohio.

---

Under illumination, dye-stained DNA generate reactive oxygen species (ROS) that cause photodamage of DNA molecules. ROS are a class of oxygen-containing molecules that, as the name suggests, are highly reactive and have been the focus of many studies. Common ROS are superoxide, singlet oxygen, hydroxyl radicals, and hydrogen peroxide. ROS are believed as a cause of ailments and cancer. The photocleavage of DNA by ROS is a problem for DNA imaging when DNA is cut into smaller pieces. It is particularly problematic for single-molecule DNA imaging such as in a genome optical mapping measurement. Thus, monitoring the generation of ROS and measure its ability of photocleavage is an important step towards finding a solution to this problem. In this report, measure the photodamage of dye-stained DNA molecules, and we use a commercial group of ROS sensing dye CellRox to in situ monitor the ROS generation near single DNA molecules that have been stained with the YOYO and POPO dyes. We observed the correlation between the photobleaching of the DNA staining dyes and the generation of the ROS. This opens an opportunity to directly test ROS suppression reagent at the single-molecule level.



## **PgmNr 1863: A versatile microfluidic platform for NGS library preparation enables multi-omics studies for a broad range of research samples and laboratory settings.**

### **Authors:**

Y. Chen <sup>1</sup>; P. Kumar <sup>1</sup>; E. Carvajal <sup>1</sup>; J. Yoo <sup>1</sup>; W. Chow <sup>1</sup>; E. Cervantes <sup>1</sup>; P. Young <sup>1</sup>; S. Jung <sup>1</sup>; S. Margeridon <sup>1</sup>; M. Jebrail <sup>1</sup>; M. Costello <sup>2</sup>; F. Christodoulou <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Miroculus Inc, San Francisco, California.; 2) Broad Institute Genomics Platform, Cambridge, Massachusetts

---

Advances in sequencing technologies have made NGS available to a broad range of researchers, yet the number of labs with dedicated assay development and automation teams remains limited. Here we describe a novel microfluidic platform, the Miró Canvas™ system, for the deployment, optimization and standardization of new and existing NGS workflows such as library construction for whole genome, exome/targeted panels and RNA sequencing.

Library construction was performed on a simple, disposable cartridge, utilizing proprietary Miró technology to automatically drive samples and reagents through the workflow. Miró technology utilizes electromechanical forces for manipulating (i.e., move, merge, mix and dispense microliter volumes) fluids across a surface of patterned electrodes in an automated fashion. To demonstrate the platform's versatility across different workflows and reagent types, Miró technology was used to automate: 1) the KAPA Hyper Prep kit for PCR-Free Human Whole Genome Sequencing (WGS), 2) the Twist library preparation and Fast Hybridization target enrichment kit of Human Core Exome for Exome Sequencing (WES) and 3) the NEBNext Ultra II Directional RNA Library Prep Kit for RNA-Seq. All sample manipulation steps including reagent mixing, incubations, thermocycling and magnetic bead-based clean ups were done on-chip with the output being a sequence-ready library.

We constructed DNA libraries with varied inputs (75-200ng sheared gDNA for WGS, 50-200ng unsharded gDNA for WES) using reference gDNA (NA12878) and DNA from varied, limited and degraded samples. For RNA-Seq, we used 5-50ng of Poly(A) enriched mRNA from different human tissues as input. Libraries made using Miró technology are equivalent or better in terms of turnaround time, reproducibility, yield, and data quality than current manual or plate-based automated methods. Miró technology improves the stoichiometry of reactions by introducing active mixing steps during incubations and shrinking reaction volumes, leading to higher adapter ligation conversion rates. This enables lower input, lower cost and improved sequencing quality across different sample types.

We believe the Miró Canvas can empower labs lacking sophisticated infrastructure to seamlessly tackle genomics and transcriptomics focused research projects. It combines the advantages of automated, hands-off library construction (alleviating user error and increasing reproducibility) while reducing sequencing costs and improving results.

## PgmNr 1864: Automation of the Twist exome library preparation and capture on the PerkinElmer Sciclone liquid handler for high-throughput, scalable NGS application.

### Authors:

J. Garifallou <sup>1</sup>; F. Mafra <sup>1</sup>; C. Siebert <sup>2</sup>; M. Consugar <sup>3</sup>; C. Vaccaro <sup>1</sup>; R. Gantt <sup>3</sup>; J. Smiler <sup>1</sup>; N. Williams <sup>1</sup>; P. Frere <sup>2</sup>; P. Sleiman <sup>1</sup>; H. Hakonarson <sup>1</sup>; R. Pellegrino <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics - The Children's Hospital of Philadelphia, Philadelphia, PA; 2) PerkinElmer, Waltham, MA; 3) Twist Bioscience, San Francisco, CA

---

**Introduction:** Advances in next-generation sequencing technologies involving sequencer instruments have allowed for faster processing of libraries via reduced on-instrument sequencing time. Updated bioinformatics pipelines have also decreased the analysis time of whole exome libraries from hours to minutes. However A potential limiting factor in any approach is the processing time required to generate libraries from the source and capture the desired regions of the DNA for sequencing. Liquid handler-based automation allows for the simultaneous processing of large sets of samples (up to 96) while limiting manual pipetting mistakes or sample to sample variation that can result. To optimize and expedite processing, liquid handler automation paired with Twist library preparation allow for the simultaneous processing of large sets of samples (up to 96) while limiting manual pipetting mistakes and sample to sample variation.

**Methods:** The Twist core exome kit was used for library preparation of 24 HapMap DNA samples with an input of 50ng. Samples were fragmented enzymatically on the PerkinElmer Sciclone instrument and proceeded immediately into library preparation. Libraries were examined for sizing and yield then normalized-pooled into 4 "capture pools." The hybridization preparation for capture was performed using two methods: on the instrument deck and utilizing a vacuum centrifuge. Capture pools were hybridized overnight in a thermal cycler and then returned to the Sciclone for the automated capture washes. Libraries were assessed for QC, then proceeded to Illumina-based sequencing.

**Results:** Our results indicated that the automated version of the library prep produced similar library sizes, yields, and general sequencing QC metrics when compared with the manual version of the preparation and at a multifold higher throughput. There was a minor drop in the percent on target statistic upon a comparison between the two approaches, which required optimization for improved performance. Results between the on deck hybridization preparation and the vacuum centrifuge based approach were comparable and did not indicate that either method was preferred.

**Conclusion:** The automation capacity of the Sciclone and rapid, high-scalability aspects of the Twist core exome kit allow for minimal processing times of NGS exomes. This leads to decreased turn-around times, and higher overall throughput of samples while maintaining high quality libraries and subsequent NGS data.

## PgmNr 1865: The Library Kit comparison in whole genome sequencing.

**Authors:**

K. Park; S. Lee; D. Seo

[View Session](#) [Add to Schedule](#)

**Affiliation:** Psomagen, Inc., Rockville, Maryland.

---

Various types of library kits are available for conducting whole genome sequencing (WGS) experiments. Each library kit has its own advantages and disadvantages. In this study, we performed a comparison of 10 different types of library kits that are currently available in the market. By using Illumina NovaSeq platform, these libraries were sequenced and analyzed with Illumina's Issac pipeline. For a fair comparison, all samples were randomly downsampled to have the raw throughput size of 100 GB. We compared the percentage of mappable reads, the percentage of de-duplicate reads, mappable mean depth, and the ratio of mappable to throughput mean depth. Among the 10 libraries, one library outperformed the others in most statistics measured in this comparison, two library kits had a relatively lower percentage of duplicated reads, and three library kits showed the relatively better result in uniformity. These results would be a helpful guideline for selecting library kit type and preparation for the best outcome of the WGS experiment.

## PgmNr 1866: Detection and phasing of small variants in Genome in a Bottle samples with highly accurate long reads.

### Authors:

W.J. Rowell; S. Chakraborty; P. Baybayan; R.J. Hall

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pacific Biosciences, Menlo Park, CA.

**Introduction:** Long-read PacBio SMRT Sequencing has been applied successfully to assemble genomes and detect structural variants. However, due to high raw read error rates of 10-15%, it has remained difficult to call small variants from long reads. Recent improvements in library preparation, sequencing chemistry, and instrument yield have increased length, accuracy, and throughput of PacBio Circular Consensus (CCS) reads, resulting in 10-20 kb “HiFi” reads with mean read quality above 99%.

**Materials and Methods:** We sequenced 11 kb size-selected libraries from the Genome in a Bottle (GIAB) human reference samples HG001, HG002, and HG005 to approximately 30-fold coverage on the Sequel II System with six SMRT Cells 8M each. The CCS algorithm was used to generate highly accurate (average 99.8%) reads of mean length 10-11 kb, which were then mapped to the hs37d5 reference with pbmm2. We detected small variants using Google DeepVariant and compared these variant calls to GIAB benchmarks. Small variants were then phased with WhatsHap.

**Results:** With these long, highly accurate CCS reads, DeepVariant achieves high SNP and Indel accuracy against the GIAB benchmark truth set for all three reference samples. Using WhatsHap, small variants were phased into haplotype blocks with N50 from 82 to 146 kb. The improved mappability of long reads allows detection of variants in many medically relevant genes such as *CYP2D6* and *PMS2* that have proven 'difficult-to-map' with short reads. We show that small variant precision and recall remain high down to 15-fold coverage.

Sample	Coverage Depth	SNP Recall	SNP Precision	INDEL Precision	INDEL Recall	WhatsHap Phase Block N50
HG001	29-fold	99.89%	99.87%	98.09%	98.08%	100 kb
HG002	32-fold	99.94%	99.97%	98.86%	99.10%	146 kb
HG005	31-fold	99.92%	99.92%	99.12%	99.04%	82 kb

**Conclusions:** These highly accurate long reads combine the mappability of noisy long reads with the accuracy and small variant detection utility of short reads, which will allow the detection and phasing of variants in regions that have proven recalcitrant to short read sequencing and variant detection.

## **PgmNr 1867: Extraction of High Molecular Weight (HMW) genomic DNA for long-read NGS applications.**

**Authors:**

A. Zegers; D. Horejsh; D. Smith

[View Session](#) [Add to Schedule](#)

**Affiliation:** Promega, Fitchburg, WI.

---

Third-generation sequencing platforms, such as those from Oxford Nanopore Technologies and Pacific Biosciences, have in recent years offered researchers dramatic improvements in sequencing read length. Long, high-quality reads have become a meaningful tool in resolving structural variants, de novo genome sequencing or genome finishing. These advances are leading to a better understanding of human diseases and conditions, as well as aiding in microbiome and agricultural research.

While the sequencing technologies have advanced, researchers are often limited by the difficulties in obtaining high-quality DNA of sufficient size to realize the full potential of these platforms. This study describes an optimized approach to isolating intact HMW gDNA that eliminates the need for laborious gel plug-based methods and enables isolating larger fragments than existing commercial rapid prep kits. PFGE analysis shows that DNA obtained by this method is larger than from existing kits, but more importantly, significant performance improvements are seen in nanopore sequencing results (mean read length and N50), suggesting the DNA is of superior quality and thus greater utility for genomics applications.

## PgmNr 1868: Application of lean manufacturing methodologies in high throughput genomic sequencing.

**Authors:**

T. Howd; P. Trefry; A. Hollinger; T. DeSmet; S. Dodge

[View Session](#) [Add to Schedule](#)

**Affiliation:** The Broad Institute, Cambridge, Massachusetts.

---

The utility and application of genomics to understand disease, and the continuing trend to utilize genomics in healthcare, results in an ever increasing demand for greater sequence data generation. Despite the significant reductions in per-base sequencing cost over the last decade, the infrastructure, capital, and reagent costs are still relatively expensive. Top of the line sequencers can cost over 1 million dollars per instrument, and sequencing run costs can still be tens of thousands of dollars. With such high fixed cost associated with genome data generation, it is important to maximize capacity utilization and reduce the non-value add and wasteful workflow process steps. We demonstrate the application of lean manufacturing methodologies and visual management techniques to the genomic sequencing workflow, which results in achieving a sequencer utilization rate of around 90%, while three fold scaling our library preparation process to over 300,000 samples destined for exome and whole human genome sequencing annually.

By combining the sample preparation methods for both exome and whole genome sequencing into a unified, modularized workflow, samples and reagent supply chains can be optimized resulting in more efficient, and cost effective processing. Additional benefits include reductions to work in process and overall cycle times. Here, we illustrate the methodologies that enable low cost per base sequence data generation applicable across large sequencing cores, and modest sized data generation groups.

## **PgmNr 1869: NGS DreamPrep™ : A fully automated NGS library preparation solution from sample to quantified, pooled, and sequence-ready libraries.**

### **Authors:**

K. Modi; A. Saraiya; E. Neumann; J.D. Heath

[View Session](#) [Add to Schedule](#)

**Affiliation:** Tecan, Redwood City, CA.

---

Illumina NovaSeq S4 flow cells have the ability to generate 10 billion single reads in a single run. To maximize this level of throughput and achieve the optimal cost savings, hundreds of samples may be pooled for each run. Manual library preparation of hundreds of samples can result in long turnaround times, increased errors, poor reproducibility and may require multiple lab personnel. The NGS DreamPrep is a walkaway solution designed and optimized for completely automated NGS library preparation and quantification. It combines Tecan's library preparation reagents, automation and fluorescence plate readers to produce sequence-ready, normalized and pooled libraries. Innovative library preparation technology minimizes workflow time and user interaction during automation. Integration of NuQuant®, a novel library quantification technology, into the library preparation allows automated library quantification, normalization and pooling on the instrument with no sample loss to QC procedures. The NGS DreamPrep system includes a Fluent® liquid handling platform optimized for NGS library preparation with integrated Infinite® F Nano+ plate reader for fast library QC. The NGS DreamPrep enables flexible library preparation for up to 96 samples in less than 4 hours for DNA and approximately 9 hours for RNA. Preparing 96 DNA samples using identical inputs generated libraries with no adaptor dimers, no dropouts and no edge effects. The mean yield (as measured by NuQuant) of all 96 libraries was 249.3 nM with a standard deviation of 24.6 nM showing high sample-to-sample reproducibility across the entire plate. Library molar concentrations were quantified by the integrated plate reader and subsequently pooled. Sequencing results showed a %CV of <15% between the number of sequence reads, demonstrating the accuracy of NuQuant library quantification. The NGS DreamPrep is an innovative solution for high-throughput NGS library preparation and quantification with walk-away automation.

## PgmNr 1870: Ultra-high-throughput PCR-free Whole Genome Sequencing (WGS) workflow.

### Authors:

Z. Momin; M. Mueller; R. Glenn; K. Walker; G. Savery; Y. Meiheerguli; D. Bandaranaike; R. Gibbs; D. Muzny; H. Doddapaneni

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas.

---

Next generation sequencing has revolutionized family and population-based research and clinical applications. Baylor College of Medicine's Human Genome Sequencing Center (HGSC) generates whole genome sequencing (WGS) data in both research and clinical settings. To date, HGSC has sequenced over 62,000 WGS samples prepared as PCR-Free libraries to >35x coverage across several NIH large scale projects on Illumina HiSeq X Ten and NovaSeq platforms. NovaSeq supports ~3x greater throughput for WGS samples compared to HiSeq X Ten platform and has been tested and integrated in both pipelines. Therefore, the current practice of preparing libraries in batches of 96 becomes a rate limiting step for the NovaSeq WGS workflow. To address this need for delivery of high quality NGS libraries, HGSC is developing a robust sample preparation (libraries) in a 384-well format. Processing of libraries in batches of 384 also provides significant cost savings due to reaction miniaturization and 4x labor savings. However, there is no readily available commercial product or workflow that supports PCR-Free, library prep, in a 384-well format. Five liquid handlers (Labcyte Echo, Formulatrix Mantis/Tempest, Biomek iSeries, Tecan Fluent, Hamilton Star) were evaluated for pipetting accuracy and consistency. These platforms were also assessed for their ability to mix reagents, solvents, and beads as required at various steps during library preparation. Based on this evaluation, the Hamilton platform, configured with a 384 multi-channel was selected. A specially designed magnet and thermocycler customizations were also necessary. Illumina test PCR-Free libraries prepared on 384-well plates were successfully sequenced to generate 110Gb (37x coverage) of unique data with 97% covered to 20X depth sufficient for clinical interpretation. These metrics are comparable to WGS libraries currently being prepared on Beckman FXP robot in a 96-well format. Comparison of high quality variants showed a high concordance rate of 98.9% with NIST reference calls. A single 384 liquid handler can generate 9,168 libraries per month (~110k/yr), enough to support ~11 NovaSeq 6000 instruments. By increasing the library throughput, ensuring delivery of high quality sequencing data can be ensured for both research and clinical applications.



## PgmNr 1871: Utilizing phenotypic similarity analyses to distinguish and identify multiple molecular diagnoses.

### Authors:

J. Punetha<sup>1</sup>; X. Song<sup>1</sup>; J.E. Posey<sup>1</sup>; D. Pehlivan<sup>1,2</sup>; Z. Coban-Akdemir<sup>1</sup>; J.A. Rosenfeld<sup>1</sup>; E. Karaca<sup>3</sup>; Y. Bayram<sup>4</sup>; S.N. Jhangiani<sup>5</sup>; T. Mitani<sup>1</sup>; J.M. Fatih<sup>1</sup>; D.M. Muzny<sup>5</sup>; R.A. Gibbs<sup>1,5</sup>; P. Liu<sup>1,6</sup>; C.A. Shaw<sup>1,6</sup>; J.R. Lupski<sup>1,2,5,7</sup>; Baylor-Hopkins Center for Mendelian Genomics.

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Texas Children's Hospital, Houston, Texas, USA; 3) Department of Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA; 4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA; 6) Baylor Genetics Laboratories, Houston, Texas, USA.; 7) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA

---

With the availability of next-generation sequencing technologies, the field of human genetics has seen an unprecedented rise in molecular diagnoses and discovery of novel genes linked to rare diseases. Genome-wide data has given us the opportunity to uncover dual or multiple molecular diagnoses - pathogenic variants at two or more loci resulting in blended phenotypes. Re-analyses of exome sequencing (ES) data from a clinical diagnostic laboratory revealed that approximately 5% (101/2076) of cases that received a molecular diagnosis by ES had more than one molecular diagnosis. Development of phenotypic similarity scores (PSS) to assess the degree of similarity between two disease traits and their corresponding sets of human phenotype ontology (HPO) terms can be used to classify blended phenotypes as either 'distinct' or 'overlapping'.

In order to test the potential utility of PSS to identify subjects for whom an additional molecular diagnosis could more parsimoniously explain the patients' clinical features, we systematically evaluated genotype-phenotype associations for 92 patients previously reported to have dual molecular diagnoses by clinical ES. We first recapitulated our previously published analyses of identifying distinct and overlapping disease pairs using an updated set of PSS with HPO terms mapped to the current OMIM MorbidMap. We then used PSS to compare the observed set of clinical features for each patient (Pt) to the set of expected features for each identified molecular diagnosis (Pt-DzA, Pt-DzB) as well as the set of expected features for the dual molecular diagnoses (Pt-DzA+B). Using a 'goodness of fit' model to compare this new set of PSS, we observed that multiple molecular diagnoses that were classified as either distinct or overlapping were able to better explain the patients' clinical features compared to a single diagnosis (distinct,  $p < 2e-16$ ; overlapping,  $p = 2.3e-15$ ). The findings remained consistent even when the more explanatory molecular diagnosis was detected first (distinct,  $p = 2.9e-11$ ; overlapping,  $p = 1.3e-14$ ). We observed similar results using PSS in an independent cohort of 16 patients with multiple molecular diagnoses. In conclusion, PSS remain a robust method for classification of distinct vs. overlapping blended phenotypes. We further demonstrate that PSS can potentially be used as a framework to assess the potential for more than one molecular diagnosis, paving the way to advance our goal towards precision health.

## **PgmNr 1872: High-throughput, automated, magnetic bead-based bisulfite conversion of human DNA for methylation analysis.**

### **Authors:**

M.M. Weinstein <sup>1,2</sup>; B. Janssen <sup>1</sup>; D. Brecklin <sup>1</sup>; H. Jamal <sup>1</sup>; R. Kemp <sup>1</sup>; L. Jia <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Bioinformatics, Zymo Research, Irvine, California.; 2) Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California

---

The ability to detect and quantify DNA methylation efficiently and accurately has become essential for the study of cancer, gene expression, genetic diseases, and many other important aspects of biology. To date, a number of methods have been developed to detect/quantify DNA methylation including high-performance capillary electrophoresis and methylation-sensitive arbitrarily primed PCR. However, the most common techniques used today still rely on bisulfite conversion.

Treating DNA with bisulfite chemically modifies non-methylated cytosine into uracil, methylated cytosine remains unchanged. Once converted, the methylation profile of the DNA can be determined using the desired downstream application. For single locus analysis, the region of interest is generally amplified following bisulfite conversion (i.e., bisulfite PCR) and then sequenced. However, recent advances in methylation detection allow the investigation of genome-wide methylation patterns, technologies include array-based methods, Pyrosequencing®, reduced representation bisulfite sequencing (RRBS), and whole genome bisulfite sequencing.

To this point all bisulfite conversion products have been dependent on manual manipulation of spin plates and columns or been of limited throughput. Here, we present an adaption of the clean-up of bisulfite converted DNA to a magnetic bead-based procedure on an automated extraction platform. We find that magnetic bead processed samples better and more consistently than established spin-column based procedures. This is demonstrated by consistent recovery, amplification and profile of the converted DNA from human samples. This novel workflow opens the door to reliable, high-throughput bisulfite conversion of human DNA samples for methylation analysis.

# PgmNr 1873: Evaluation of single-molecule long-read whole-genome shotgun sequencing on two reference microbiomes.

## Authors:

Y. Hu <sup>1</sup>; L. Fang <sup>1</sup>; C. Nicholson <sup>1,2</sup>; K. Wang <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Raymond G. Perelman Center for Cellular and Molecular Therapeutics, The Children's Hospital of Philadelphia, Philadelphia, PA.; 2) Department of Biology, University of Pennsylvania, Philadelphia, PA; 3) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Currently, 16S ribosomal RNA (rRNA) gene amplicons and shotgun meta-genome sequencing on Illumina platforms are two dominant approaches for analyzing microbiomes. Using the Oxford Nanopore platform, here we evaluated the feasibility to characterize complete microbial genomes and populations using single-molecule long-read whole-genome shotgun sequencing without any DNA amplification.

We first generated ~500X coverage Nanopore sequencing data on HM-276D Mock Community sample from the Human Microbiome Project. HM-276D is an evenly mixed DNA sample of 20 bacterial strains (each with 5% abundance). We assembled the 20 bacterial genomes into 60 contigs using canu, with contig N50 length of 2.98 Mb. Other computational tools such as wtdbg2 and OPERA-MS generated similar results. All 3 tools consistently achieved high accuracy (~99%) and completeness (~98.5%). In comparison, genome assembly on 160X coverage Illumina sequencing data generated N50 of 0.17 Mb with 4,575 contigs, suggesting that long-read sequencing can generate much less fragmented genome assemblies on microbiomes than short-read sequencing. Next, we generated ~1,000X coverage data of the HM-277D Mock Community sample, which is an unevenly mixed DNA sample of the 20 bacteria strains. For bacteria with higher relative abundances (>4%), at least 99.93% reference genome sequences were covered by assembled contigs. The fractions of exonic coverage of 16S rRNA genes were ~100% for 18 bacterial strains. Estimation of strain abundance from sequence data has a correlation of 0.895 (log scale) with known abundance in the reference sample. Furthermore, taxonomic binning accuracy was evaluated based on both HM-276D and HM-277D mock community, showing that ~99.99% bases can be correctly assigned to known bacterial species with a false positive rate close to zero.

Overall, our results demonstrated the technical feasibility to characterize complete microbial genomes and populations from error-prone Nanopore sequencing data in the absence of DNA amplification. We propose that future metagenomics studies will benefit from this approach to assemble complete microbial genomes, while maintaining the theoretical ability to detect DNA methylations and base modifications, infer repetitive elements and structural variants, and achieve strain-level resolution within microbial communities.

## PgmNr 1874: Integration of molecular chemistries supporting a full-length, mRNA sequencing library preparation method on a microfluidic circuit.

### Authors:

M. Phelan; J. Alipaz; J. Brockman; S. Chamnongpol; B. Fowler; J.A. Geis; T. Goralski; C. Kubu; R. Kung; B. Lacar; C. Park; D.A. King

[View Session](#) [Add to Schedule](#)

**Affiliation:** Fluidigm Corporation, South San Francisco, CA.

---

RNA sequencing (RNA-seq) is the gold standard for hypothesis-free profiling of the transcriptome and an essential biological tool. Because of its functional utility and widespread use, there is a need to simplify the complex steps of RNA-seq library preparation for next-generation sequencing (NGS) applications and to substantially reduce the costs associated with it. We are developing a library preparation solution that combines reagents, a microfluidics device (approximately the same size and shape as a standard 96-well plate), and an instrument to generate full-length, RNA-seq libraries. This microfluidic device is a new integrated fluidic circuit (IFC) that will enable automation of solid-phase sample enrichment, as well as multistep chemistries, after a single pushbutton script to activate the pneumatics and thermal controls on the instrument. The workflow will simplify the tedium and hands-on steps of library preparation while reducing costs because of the nanoliter volumes required on the IFC.

Our RNA-seq library preparation method supports simultaneous processing of up to 48 total RNA samples from many organisms, and it is compatible with Illumina® sequencing instruments. The method generates full-length, stranded libraries from random-priming of the polyadenylated RNA (polyA RNA) present in the total RNA. RNA samples, polyA selection beads, and common reagents are loaded in the carrier inlets on the IFC. The IFC is placed in the instrument and processing scripts are initiated. These scripts control the nine different reaction steps that occur sequentially, without user intervention, from polyA bead column formation through index PCR. Following sample index PCR, the libraries are delivered to individual sample-specific carrier outlets on the IFC for harvesting.

To demonstrate performance, we processed 48 total RNA samples from a wide variety of tissues, cell lines, and RNA from different organisms. We generated high-quality libraries from as low as 10 ng of total RNA. Our mRNA-seq libraries had less than 10% rRNA reads, replicate correlations greater than 99%, and high gene-level and transcript-level detection rates.

In conclusion, mRNA-seq library preparation can be simplified using automated microfluidics technology enabling solid-phase enrichment and multistep chemistries. This solution will help reduce costs while maintaining high-quality RNA-seq libraries.

## PgmNr 1875: Clonal analysis, genetic screening, and expression profiling at a single-cell level.

### Authors:

A. Chenchik<sup>1</sup>; M. Makhanov<sup>1</sup>; R. Darst<sup>1</sup>; D. Deng<sup>1</sup>; D. Tedesco<sup>1</sup>; P. Diehl<sup>1</sup>; L. Kobzik<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cellecta, Mountain View, CA; 2) Harvard Medical School & Harvard School of Public Health, Cambridge, MA

---

Novel methods for profiling clonal composition and phenotypes of primary and metastatic lesions at the single-cell level are required for understanding tumorigenesis mechanisms and development of anti-cancer drugs with unique mechanisms of action. In order to facilitate these studies, we developed a panel of lentiviral barcoded libraries to label and monitor the cancer cells in time course experiments *in vitro* and in mouse xenograft models. The barcodes are transcribed from lentiviral constructs and could be detected by NGS in a single-cell RNA expression profiling assay to identify the subpopulation of descendant cells derived from a single barcoded progenitor cell. Furthermore, cell barcodes were incorporated in conjunction with genetic effector libraries, such as sgRNA libraries, to identify clonal phenotypic changes induced by specific genetic disruptions in progeny cells derived from the single progenitor cell. However, while cell barcodes provide an effective way to group cells based on clonal origin in heterogeneous cell populations, expression profiling at a single-cell level remains challenging. Data will be presented showing how genetic screening and targeted RNA expression profiling of human cells combined with cell barcoding could significantly improve phenotyping of distinct cell populations.

## PgmNr 1876: Cap-dependent linker ligation increases specificity for full-length products compared to template switch reaction.

### Authors:

D. Daujotyte; M. Tsujita; F. Kabinger; T. Dozd; M. Ante; A. Tuerk; P. Moll; T. Reda; A. Seitz

[View Session](#) [Add to Schedule](#)

**Affiliation:** Lexogen GmbH, Campus Vienna Biocenter 5, 1030 Vienna, Austria

---

Nanopore enables full-length sequencing of RNA or cDNA. A fast and easy way to obtain full-length cDNA is the commonly used template switch reaction. The reverse transcriptase (RT) adds non-templated nucleotides (preferentially C's) at the end of a transcript which hybridize to abundant template switch oligos. However, non-templated nucleotides can also be added to fragmented RNAs or premature termination sites of the RT. Artificial Spike-in transcripts are an essential part to monitoring the quality of an RNA-Seq experiment, to control NGS sample preparation, base callers and algorithms by adding a ground truth to the NGS experiment. We used Lexogen SIRV™ set 3 containing 69 Spike-In RNA Variant controls, simulating alternative splicing of 7 SIRV genes plus antisense transcription, plus 92 ERCC Spike in controls (External RNA Controls Consortium Spike-In controls, Thermo Fisher Scientific Inc.). ERCCs are monoexonic but cover a concentration range of 6 orders of magnitude. TeloPrime is a full-length cDNA preparation kit offered by Lexogen. Exceptional 5'-Cap specificity is achieved with the proprietary CAP dependent linker ligation. We capped SIRV set 3 using the Vaccinia Capping Enzyme and protocol from NEB (M2080). Universal Human Reference RNA (UHRR) was spiked in with the capped SIRV set 3 before the sample was subject to a controlled degradation. Nanopore sequencing libraries were made either by using the template switching protocol or a modified TeloPrime protocol (v3) from intact and degraded RNA aliquots. The new TeloPrime v3 Nanopore libraries contain a 12 nt unique molecular Index (UMI) that is introduced with the RT primer, enabling to account for sequencer and PCR errors in high coverage NGS data. Degraded RNA resulted in shorter libraries for the less Cap-sensitive template switching protocol, but little delay in the PCR cycles, while for TeloPrime v3 degraded RNA libraries resulted in significantly less amplifiable library. The analyses of apparent transcript start site distributions by Nanopore sequencing showed a higher cap specificity for TeloPrime v3 than for the template switching protocol. Hence, TeloPrime v3 enables an increased accuracy for transcript 5' end detection.

## PgmNr 1877: Evaluation of RNA sequencing methods for low input samples.

**Authors:**

H. Munger <sup>1</sup>; C. Snow <sup>1</sup>; H. Tjong <sup>2</sup>; C.-L. Wei <sup>2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) The Jackson Laboratory, Bar Harbor, Maine.; 2) The Jackson Laboratory, Farmington, Connecticut

---

RNA-seq is a sensitive and accurate tool for measuring gene expression across the transcriptome, but can be difficult to carry out for rare cell populations, small tissue samples, and other limited amounts of starting material. Here we compare three library construction methods for their performance in transcriptome analysis, including read alignment quality, strand specificity, rRNA contamination, coverage bias, and workflow. We titrated the input RNA down to as low as 100pg and tested each method using both technical and biological replicates using cells from mouse brain and granulocyte-monocyte progenitor cells. Sequencing was carried out on the Illumina NextSeq500, 75bp paired end, to a depth of 20-30 million reads per sample. Of the libraries that were sequenced, we observed similar alignment rates amongst the three methods, while other analysis metrics were variable. A key evaluation parameter for RNA-seq is strand specificity and only one protocol maintains strand information. Overall, high consistency of transcriptome counting was observed among different biological and technical replicates, while all methods yielded robust transcriptome counting and survey of the complexity of the expressed genes as well as uniformity in coverage. In summary, we found all three protocols to be robust and reproducible with low input RNA, and each protocol has applications depending on the researcher's experimental needs.

## PgmNr 1878: NuQuant: A fluorescence-based method for accurate quantification of NGS libraries in minutes.

**Authors:**

A. Saraiya; C. Parikh; B. Li; J. Heath

[View Session](#) [Add to Schedule](#)

**Affiliation:** Tecan, Redwood City, CA.

---

Accurate quantification of NGS libraries is critical for a successful sequencing run. Currently used methods of quantification are time-consuming, costly, and can be highly variable. We have developed NuQuant, a novel method to accurately quantify NGS libraries, that can be performed with a simple fluorescent measurement. NuQuant is compatible with the common red fluorescence excitation/emission filter set (650/670 nm), making it compatible with a wide range of bench top fluorometers and fluorescent plate readers. We have developed a custom library quantification application for the Qubit fluorometer that directly provides the molar concentration of a library. Utilizing this application, we have demonstrated that NuQuant has excellent reproducibility across users from multiple sites. We now demonstrate the compatibility of NuQuant with standard fluorescence plate readers, enabling quantification of libraries in a high-throughput fashion. We have tested NuQuant on a variety of commonly used plate readers such as the Tecan Infinite 200 Pro and the Promega GloMax. Libraries in a 96-well format can be measured in a matter of minutes, without the need for sample dilution. Molar concentration of libraries was easily determined by utilizing a standard curve. We tested libraries with various input from 10ng to 500ng and insert size from 200bp to 500bp, and found good agreement between NuQuant values and a qPCR based quantification method. Most importantly, we observe good correlation between NuQuant library concentration and total number of sequenced reads. In conclusion, scientists with access to commonly used fluorescent plate readers can now use NuQuant to achieve rapid and cost-effective quantification of NGS libraries, generating highly uniform sequence reads in multiplex runs.



## PgmNr 1879: Global analysis of transcriptional start sites using Lexogen's Corall total RNA-seq library prep kit.

**Authors:**

M. Sauert; Y. Göpel; G. Wiktorin; T. Reda; A. Seitz

[View Session](#) [Add to Schedule](#)

**Affiliation:** Lexogen GmbH, Campus Vienna Biocenter, Vienna, Austria

---

The actual transcriptional start sites (TSS) of processed as well as non-processed transcripts are of particular interest to many researchers. Cap-analysis gene expression sequencing (CAGE-Seq) is one favored method to globally assess TSS in given RNA sample. This method, however, only works for mature mRNA with an intact 5' cap and provides only limited information about the full-length transcript.

CORALL is Lexogen's new stranded total RNA library prep kit with excellent whole transcriptome coverage and unprecedented representation of 5' ends. Lexogen's proprietary methodology enables a particularly efficient coverage of 5' ends without a complete drop of sequencing reads towards the 5' ends - unlike most other whole transcriptome library prep kits. CORALL's comprehensive coverage thereby delivers improved transcript start and end site representation.

Here, we show in a proof of principle that RNA-Seq libraries prepared with CORALL excellently represent the actual 5' ends of transcripts. Read coverage was analyzed using the ERCC spike-in controls, which feature precise, known transcription start and end sites (TSS and TES, respectively). CORALL reads map more accurately to the exact ERCC TSS than competitor libraries, which fail to cover the true start sites. Additionally, CORALL provides elevated coverage at TES.

CORALL enables streamlined generation of Illumina-compatible libraries within 4.5 hours, featuring seamless integration of Unique Molecular Identifiers (UMIs) and exceptional protocol-inherent strand specificity (above 99 %). The fragmentation-free protocol uses Lexogen's proprietary Strand Displacement Stop and Ligation technologies to deliver complete transcript representation, and CORALL libraries can be prepared from as little as 1 ng of total RNA input. Flexible input types are supported including, rRNA-depleted, poly(A)-enriched, or total RNA from a wide variety of species, as well as degraded and FFPE RNA samples.

## PgmNr 1880: Stranded RNA-Seq next-generation library preparation kit for low inputs.

**Authors:**

A. Wood; K. Peck; S. Sandhu; V. Makarov; L. Kurihara

[View Session](#) [Add to Schedule](#)

**Affiliation:** R&D, Swift Biosciences, Ann Arbor, Michigan

---

RNA-Seq technology allows for transcriptome profiling and the accurate measurement of transcript expression levels. This is important for the detection of differentially expressed genes across different treatments or biological conditions. Additionally, RNA-Seq has applications for identifying novel transcripts, novel splice junction sites, and gene fusions among others. However, one major challenge is the production of RNA-Seq libraries from a limited amount of starting material. Most RNA-Seq kits require >100 ng of total RNA in order to reliably generate high-quality RNA-Seq libraries. To overcome this challenge, we present the Swift RNA-Seq stranded library preparation workflow. Based on Adaptase™ technology, Swift RNA-Seq produces quality libraries from as little as 10 ng of starting total RNA. The high-efficiency Adaptase reaction successfully adds a truncated Illumina-compatible adapter following first-strand cDNA synthesis, avoiding the requirement for synthesis of the second-strand. Following Adaptase, ligation adds the second truncated adapter and indexing PCR completes the library preparation by incorporation of full-length indexed adapters. Swift RNA-Seq stranded libraries result in high mapping rates and maintain >98% strandedness, even at inputs as low as 10 ng. Additionally, Swift RNA-Seq libraries have low duplication rates, even coverage across transcripts, and high expression profile efficiencies. Compared to two leading RNA-Seq kits, the Swift RNA-Seq stranded kit identifies more transcripts and genes at low total RNA inputs (< 100 ng), with comparable metrics at high inputs. Further, Swift RNA-Seq libraries show negligible adapter dimers with no adapter titration required, does not require a PhiX spike-in for sequencing, and is compatible with gene-specific primers. Overall, the Swift RNA-Seq stranded library preparation kit provides an optimal library preparation kit for next-generation sequencing and excels particularly when total RNA inputs are limited.

## PgmNr 1881: Library preparation miniaturization: Effects of miniaturization on next generation sequencing data.

**Authors:**

R. Obermoeller; J. Kinman; P. Frere; T. Rawlins; J. Atwood; A. Kothandaraman

[View Session](#) [Add to Schedule](#)

**Affiliation:** PerkinElmer, Austin, TX.

---

Gene expression of a single cell differs from the population from which it originates. This makes the identifying complex gene expression patterns associated with cellular functions and disease a challenge. Thus, funding requests for studies aimed at sequencing hundreds to thousands of single cells have been growing steadily. However, the inherent complexity of the chemistry required to generate cDNA from nanogram scale inputs of RNA can make the approach cost prohibitive.

This study demonstrates a fully-automated, high-throughput workflow for single-cell RNA sequencing at a miniaturized scale. This workflow from cell lysis through cDNA synthesis and library construction enables an economic method to prepare high-quality NGS libraries. The study compares critical factors like the quality of sequencing data, enzyme consumption, and costs, to characterize the degree of miniaturization possible without compromising data quality. For research use only. Not for use in diagnostic procedures.

## PgmNr 1882: Cell type specific differential expression from bulk tissue with single cell reference.

### Authors:

Wang <sup>1</sup>; N. Zhang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania.; 2) Department of Statistics, Univ Pennsylvania, Philadelphia, Pennsylvania.

---

Bulk tissue sequencing is often used to probe genes that have tissue-level expression changes between biological cohorts. However, tissue are usually a mixture of multiple distinct cell types and the tissue-level changes are due to shifts of cell type proportions as well as cell type specific expression changes. Single-cell RNA sequencing (scRNA-seq) allows the investigation of the roles of individual cell types during disease initiation and development. Deconvolution methods have been developed to estimate and compare the cell type proportions between tissues with scRNA-seq data as reference. We focused on the "reverse" estimation of cell-type-specific expression and cell-type-specific differential test. Here we present a method MuSiC-DE to detect cell-type-specific differential expression with estimated cell type proportions by MuSiC, which utilizes not only cell-type specific gene expression from scRNA-seq data but also cross-cohort variation to characterize cell type compositions from bulk RNA-seq data in complex tissues, where ignoring the cross-cohort variation leads to attributing all variation to cell type proportions changes. With MuSiC-estimated cell type proportions, we propose a reverse estimation procedure that can detect cell type specific differential expression, allowing for the elucidation of the roles of genes and cell types, as well as their interactions, on disease phenotypes.

## PgmNr 1883: Long-read sequencing and phasing of mutations in *TP53* using Xdrop targeted enrichment.

### Authors:

A. Ameur <sup>1</sup>; I. Höijer <sup>1</sup>; E. Madsen <sup>2</sup>; Z. Li <sup>1</sup>; A. Bajalan <sup>1</sup>; T. Kvist <sup>2</sup>; T. Soussi <sup>3</sup>; M. Mikkelsen <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of immunology, genetics and pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 2) Samplix, Herlev, Denmark; 3) Department of Oncology-Pathology, Karolinska Institute, Cancer Center Karolinska (CCK), Stockholm, Sweden

---

Targeted enrichment of long DNA-fragments is advantageous if your focus is limited to specific regions in the genome. To obtain long-read sequencing information might be crucial for resolving structural variations or phasing of variants and/or mutations. Common approaches, such as long-range PCR and hybridization-based methods, have difficulties in enriching for very long fragments (>10kb) and may introduce chimeric molecules and other PCR-related artefacts during amplification. Here, we have evaluated the Xdrop technology, a novel low-input method for targeted enrichment of long DNA-molecules, to analyze mutations in *TP53*. The Xdrop technology utilizes short-range sequence information and only nanograms of input DNA to obtain long-range sequence information by isolation of large DNA-molecules in double-emulsion droplets. Locus-specific droplet PCR of short fragments (<100bp) is used for fluorescent tagging of the droplets, which could then be sorted by FACS. After sorting, the enriched DNA is released and emulsified into single-emulsion droplets and amplified by MDA. By only having one DNA-molecule per droplet the production of chimeras can be avoided during MDA. By using the Xdrop technology and PacBio SMRT sequencing we were able to produce continuous sequence up to 40 kb with reads over 10 kb in length. This allowed us to analyze the entire *TP53* gene in Jurkat cells and in chronic lymphocytic leukemia (CLL) patient samples. With a newly developed analysis pipeline we were able to phase the cancer mutations in the samples. Moreover, we are optimizing the protocol to obtain even longer reads and are evaluating the Xdrop technology for nanopore sequencing.

## **PgmNr 1884: Electrical detection of single base incorporations enables high accuracy sequencing on a small, scalable platform.**

### **Authors:**

T. Clark; M. Fallahi; S. Paliwal; M. Barmi; A. Nabi; X. Gomes; A. Barua; K. Parizi; M. Jouzi; S. Stern; P. Kenney; H. Rategh; S. Sankar; H. Esfandyarpour

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genapsys, Inc., Redwood City, CA.

---

DNA sequencing technologies have undergone tremendous development over the past decade. Though optical-based sequencing is responsible for the majority of data generated, it requires a large capital investment and aggregation of samples to achieve optimal cost per sample. Accurate DNA sequencing on an accessible platform that is small, scalable, modestly priced, and inexpensive to operate, will allow for a more distributed model where the power of genomics is put back into the hands of individual researchers.

GenapSys has developed a novel electronic-based platform capable of accurately detecting single base incorporations. The detection method, which utilizes CMOS chips, enables the system to be compact, accessible, and affordable. The platform is capable of generating 1.5 Gb of high-quality nucleic acid sequence in a single run, and we routinely generate sequence data that exceeds 99% raw accuracy with read lengths of up to 175 bp.

Here, we demonstrate the functionality of the novel impedance-based GenapSys sequencing technology and highlight the utility of this platform for variant detection in human samples. Performance was evaluated on several well characterized cell lines from the Genome in a Bottle (GIAB) Consortium including NA12878 and the Ashkenazim and Han Chinese Trio samples. Exome libraries were generated using the IDT xGen probe-based capture method and sequenced on the GenapSys GS111 to greater than 50-fold average coverage. We evaluated BCFtools and Google DeepVariant for variant calling. DeepVariant is an analysis pipeline that uses a deep neural network to call genetic variants from next-generation DNA sequencing data. The SNVs detected using the GenapSys platform correlated extremely well with those detected from Illumina sequencing data generated from the same samples and with the high confidence calls from the GIAB consortium.

## PgmNr 1885: Gene expression assessment of different isolations and storage conditions for white blood cells.

### Authors:

Yanru. Xing <sup>1,2</sup>; X.I. Yang <sup>2</sup>; Haixiao. Chen <sup>2</sup>; Sujun. Zhu <sup>3</sup>; Jinjin. Xu <sup>2</sup>; Yuan. Chen <sup>2</sup>; Juan. Zeng <sup>3</sup>; Fang. Chen <sup>2</sup>; Hui. Jiang <sup>1,2</sup>; WenJing. Wang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of the Chinese Academy of Sciences, Beijing, China; 2) BGI-Shenzhen, Shenzhen 518083, China; 3) Obstetrics Department, Shenzhen Maternity and Child Healthcare Hospital, Shenzhen, Guangdong Province, China

---

**BACKGROUND:** Peripheral leukocyte gene expression is a potential biomarker for human health and disease. However, different methods to isolate cells and storage conditions of blood could affect leukocyte transcriptome, and how to reflect human physiological state preferably is largely unknown.

**METHODS:** We used RNA-seq to explore the differences in expression profiles between whole blood (WB) and leukocyte from Buffy Coat (BC), Red Blood Cell Lysis (RBL) and Peripheral Blood Mononuclear Cell (PBMC) and the effects of leukocyte with samples storage under different conditions.

**RESULTS:** RBL and PBMC reduced much more globin ( $1.00\% \pm 1.23\%$ ,  $0.06\% \pm 0.03\%$ ) than BC ( $17.48\% \pm 6.95\%$ ) and WB ( $24.46\% \pm 6.43\%$ ), and with higher transcripts mapping ratio and detected gene number. There were 481 genes observed lower expression in RBL and PBMC than WB and BC, which enriched in erythrocytes related pathway. While 616 genes which enriched in leukocyte related function, expressed slightly higher in RBL than WB and BC, but barely detected in PBMC. Compared to samples treated immediately, the integrity of leukocyte RNA had been little influenced by stored for 24h, while expression of much more genes significantly changed with blood stored at RT (2545) than at 4° (226) in 24h.

**CONCLUSIONS:** RBL reduced much more globin mRNA than BC, and remained more leukocyte information than PBMC, which was more suitable for researches about whole leukocyte function. Samples stored at 4° within 24h had little influence on gene expression, but these differences should also take into account in studies.

**Funding:** National Natural Science Foundation of China (No.81300075), Shenzhen Municipal Government of China (No.JCYJ20170412152854656).

## PgmNr 1886: Single-nuclei RNA-seq on human retina: Improvement in transcriptome profiling of complex tissue.

### Authors:

Q. Liang<sup>1,2</sup>; R. Dharmat<sup>1</sup>; X. Qian<sup>1,2</sup>; X. Cheng<sup>1</sup>; L. Owen<sup>3</sup>; A. Shakoor<sup>3</sup>; Y. Li<sup>1</sup>; S. Kim<sup>1</sup>; A. Vitale<sup>3</sup>; I. Kim<sup>3</sup>; D. Morgan<sup>3</sup>; N. Wu<sup>1</sup>; M. DeAngelis<sup>3,4</sup>; R. Chen<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.; 2) Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, 77030, United States.; 3) Department of Ophthalmology and Visual Sciences, University of Utah, Salt Lake City, UT, 84132, United States.; 4) Department of Pharmacotherapy, the College of Pharmacy, University of Utah, Salt Lake City, UT, 84132, United States

---

The retina is a heterogeneous tissue composed of multiple cell types with each playing a unique role for visual processing. For heterogeneous tissues, bulk RNA-Seq can only provide the average gene expression profile for all cells from the tissue, making the interpretation of the sequencing result challenging. Here we report a single-nuclei RNA-seq transcriptomic study on healthy human neural retinal tissue to identify transcriptome profile for individual cell types. In the pilot experiment, six retina samples from three healthy donors were profiled and RNA-seq data with high quality was obtained for 5873 single nuclei. All major cell types were observed from the dataset and signature genes for each cell type were identified by differential gene expression analysis, largely expanding the known cell type marker list. The gene expression pattern of the macular and peripheral retina was investigated at the cell type level, showing significant improvement from previous bulk RNA-seq studies. Furthermore, our dataset showed improved power in prioritizing genes associated with human retinal diseases compared to both mouse single-cell RNA-seq and human bulk RNA-seq results. In conclusion, we demonstrated that feasibility of obtaining single cell transcriptome from human frozen tissues to provide additional insights that is missed by either the human bulk RNA-seq or the animal models. Building on the success of the pilot project, we are expanding our study to generate 150,000 single-nuclei RNA-Seq from three health donor eye. Progress made in the expanded study will be reported.



## PgmNr 1887: Circulating microRNA profiling in systemic lupus erythematosus.

### Authors:

C. Francis <sup>1</sup>; H.J. Park <sup>1</sup>; S. Manzi <sup>2</sup>; F.Y. Demirci <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Pittsburgh, Pittsburgh, PA; 2) Allegheny Health Network, Pittsburgh, PA

---

Systemic Lupus Erythematosus (SLE) is a chronic multisystem autoimmune disorder, which is mediated by multiple abnormalities in both innate and adaptive immune responses. MicroRNAs (miRNAs) are a class of small noncoding RNA molecules that contribute to epigenetic regulation by modulating gene expression at the post-transcriptional level. MiRNAs are known to play an important role in regulating both innate and adaptive immune responses. Circulating miRNAs have recently emerged as potential biomarker candidates for various human conditions, including autoimmune diseases. The primary objective of this study was to utilize a comprehensive plasma miRNA profiling data obtained from an SLE case-control discovery sample (n=60) to identify a candidate set of highly SLE-relevant circulating miRNAs to be further evaluated in large follow-up studies. MiRNA profiling data were obtained by running the miScript Human miRBase Profiler HC series of PCR arrays (containing 2402 miRNA qPCR assays) on pooled RNA samples (6 pools; 10 RNAs per pool) purified from blood plasmas obtained from 60 Caucasian subjects (30 SLE patients and 30 controls) to compare the circulating miRNA levels between the two groups. Of those plasma miRNAs expressed in >80% of all sample pools profiled, a total of 280 showed >2-fold expression difference between SLE cases and controls, of which 66 showed >25 fold down-regulation in SLE as compared to the control group. The plasma miRNAs that showed significant differential expression in SLE have included some previously reported SLE-associated circulating miRNAs as well as a number of newly implicated ones. *In silico* assessment of the biological functions/pathways regulated by highly SLE-relevant circulating miRNAs using DIANA-mirPath v3.0 has revealed a large number of targeted pathways/categories, including both already known (immune regulation/response, apoptotic processes, and epigenetic regulation) and some potentially new ones. In summary, our discovery study identified several highly SLE-relevant (previously reported + novel) plasma miRNAs that appear to contribute to SLE-associated circulating miRNA signature and warrant further confirmation/validation as potential biomarkers in large follow-up samples/studies. Development of clinically useful biomarkers is essential for better management of this heterogeneous and challenging systemic autoimmune condition that affects relatively young adults.

## PgmNr 1888: Gene expression signature of macrophages infected by *Mycobacterium leprae*.

### Authors:

A.C.P. Latini <sup>1</sup>; B.L. Martins <sup>2</sup>; T.L. Calvo <sup>3</sup>; D.F. Bertoluci <sup>1</sup>; R.M. de Camargo <sup>2</sup>; P.S. Rosa <sup>1</sup>; V.N. Brito de Souza <sup>1</sup>; M.O. Moraes <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Lauro de Souza Lima Institute, Bauru, Brazil; 2) Botucatu Medical School, São Paulo State University, Botucatu, Brazil; 3) Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

---

Leprosy is an infectious disease caused by *Mycobacterium leprae*, a non-cultivable bacillus, making difficult the disease studies. This pathogen is intracellular and resides mainly in macrophages and Schwann cells. The global burden of leprosy was 210,671 new cases in 2017. Leprosy outcomes are importantly determined by human genetic factors. However, the leprosy genetic architecture is not clear. In order to help elucidate the important genes to disease, we decided to determine the gene expression signature for macrophages infected by *M. leprae*. First, we made the transcriptome for macrophages derived from monocytes (MDMs) from 30 healthy individuals (15 males and 15 females), using HumanHT-12 v4 BeadChip (Illumina). Principal component analysis (PCA) was applied and detected 91 differentially expressed genes in infected MDMs. In order to confirm the results, we tested the expression of these genes in 96 healthy individuals (48 males and 48 females), under the same experimental conditions, by using quantitative PCR (Delta Gene Assay, Fluidigm). Among the 91 genes, 17 had differential expression confirmed in MDMs infected: *ADAMDEC1*, *C15orf48*, *C3*, *CCL4*, *CCL4L2*, *G02S*, *PTGES*, *SLC7A11*, *SLC2A6*, *TMEM163*, *GNG2*, *HPSE*, *PRR5L*, *RNASE1*, *SELENOP*, *SERPINF1* and *MS4A6*. These genes are mainly related to cellular migration and metal homeostasis, being important to macrophages functions, but also for microorganism survival. These genes must compose the gene signature of MDMs infected by *M. leprae*, and are important to future strategies to determine new loci associated to leprosy risk.

Finacial support: FAPESP (Grant 2015/01744-9)

## PgmNr 1889: HLA-typing in a large whole-genome sequencing study of atopic dermatitis.

**Authors:**

B. Przychodzen; S. Smieszek; Y. He; D. Xiao; C. Polymeropoulos; G. Birznieks; M. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals, Washington DC, OH.

---

**Background:**

The HLA region, located on chromosome 6, is one of the most diverse regions in the human genome with thousands of alleles present in the human population. Classical class I (HLA-A, -B, and -C) and class II (HLA-DR, -DP, and -DQ) HLA gene products function by presenting foreign antigens to T cells to trigger immune responses. The HLA complex is responsible for the regulation of the human immune system and thus may play an important role in Atopic Dermatitis (AD). AD is a common chronic episodic disorder manifested by itchy red patches. It is likely caused by interactions between an altered skin barrier and immune dysregulation.

**Methods:**

HLA typing was performed on patients of two clinical studies, VP-VLY-686-2102 and VP-VLY-686-3101. Both studies were a randomized, double-blind, placebo-controlled, multi-center study in patients with chronic pruritus associated with AD treated with tradipitant a novel neurokinin 1 (NK1) antagonist or placebo. The dataset consisted of 387 whole genome sequencing (WGS) samples. Human leukocyte antigen (HLA) gene region and its alleles were called from whole genome sequencing using data (average coverage 30X, Illumina NovaSeq).

**Results:**

DRB1\*01:02, amino acid variations at position 9 (pocket 9) was the top variant marginally associated with the prevalence of AD. The odds ratio was 3.42 (CI 95=1.11-10.47; p=0.0284). We have shown that 36 patients carry significantly more of these variants when compared to 11 in a control population. Furthermore, we detect a marginal association with IgE levels, limited by small *n* to be further discerned with the accrual of more samples. Within the 36 patients carrying one of the top three MHC II alleles (Table1), 6/36 patients (16.6%) carried a FLG Loss of Function mutation as compared to 24/181 (13.2%, p=0.5991).

## PgmNr 1890: Identifying candidate genes for hEDS: Initial results of the hEDS gene study.

### Authors:

K. Martinez <sup>1</sup>; O. Erdemir <sup>2</sup>; A. Sanoja <sup>3</sup>; C. Mauss <sup>2,3</sup>; P. Byers <sup>4</sup>; C. Laukaitis <sup>1,2,3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Genetics GIDP, University of Arizona, Tucson, ARIZONA.; 2) Department of Medicine, University of Arizona, Tucson, ARIZONA; 3) College of Medicine, University of Arizona, Tucson, ARIZONA; 4) Departments of Pathology and Medicine, University of Washington, Seattle, WASHINGTON

---

Ehlers-Danlos syndrome (EDS) describes a collection of heritable disorders that have in common joint hypermobility. The presence of additional objective signs and pathogenic variants in more than a dozen genes categorizes a subset of affected individuals into 1 of 13 EDS subtypes. However, the largest group is characterized by hypermobility but without a pathogenic variant in the candidate genes and usually do not have the defining clinical features of the other types. In this group individuals can be formally diagnosed with the hypermobile subtype of EDS (hEDS) or further classified as having generalized hypermobility on the EDS spectrum (G-HSD). The hEDS GENE study is a multi-center, non-interventional study with the goal to identify genes associated with hEDS using a strategy of enrolling families with multiple members who meet the 2017 hEDS criteria. More than 94 people from 27 families have been enrolled and their exomes or genomes sequenced. These include three families with third-degree relatives (first cousins), five with second-degree relatives (aunt/niece or grandparent/grandchild), and 19 families with multiple first-degree relatives (parents/children or siblings). More than forty individuals are also enrolled resulting in a total of more than 134 participants. Initial analysis focused on using a statistical framework called Rare Variant Sharing (RVS) to identify potentially causal variants in our multiplex exome and genome families. With a sharing probability below 0.01, numerous variants have been identified including in collagen and FBN genes, which are associated with EDS. Interesting shared variants have also been identified in other members of the BMP and SERPIN gene families as well as multiple genes encoding proteins with ANKR domains. The potential functional significance of each of these is being evaluated using literature review and pathway analysis. Data from additional families and affected individuals are being interrogated to determine whether variation in these genes can be replicated.

## PgmNr 1891: Variants in the *INADL* gene are associated with worst itch (VAS) in atopic dermatitis samples.

### Authors:

J. Maigrot; S.P. Smieszek; C. Xiao; J. Wang; S.E. Welsh; C. Polymeropoulos; G. Birznieks; M.H. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals Inc., Washington, District of Columbia.

---

The genetic background of atopic dermatitis (AD) is complex, and may vary not only with disease course and intensity, but also between different ethnicities. To investigate this variance, we examined individual outcome parameters of patients in clinical study VP-VLY-686-2102. This was a randomized, double-blind, placebo-controlled, multi-center study of 168 patients with chronic pruritus associated with AD. Study participants were treated with Tradipitant (VLY-686), a novel neurokinin-1 receptor (NK1R) antagonist or placebo. Subjects also consented to participate in pharmacogenetic analysis. Using linear regression, we directly tested the association between 14,322,979 SNPs and Worst Itch Visual Analog Scale (WI-VAS). Among the top loci identified to modify WI-VAS were variants in the *INADL* gene. *INADL* encodes InaD-like protein, which plays a role in tight junction formation. Tight junctions reside immediately below the stratum corneum and regulate the selective permeability of the paracellular pathway. Tight junction defects have been implicated in AD, as they may lead to impaired skin barrier function and immune dysregulation. The region, which contains the variant that is most significantly associated with change in WI-VAS, rs11207834 (p-value = 6.1E-5), has high regulatory potential. We observed a hindering effect of the minor allele on subjects' response in change of WI-VAS. The two cases in this study, which were heterozygous for the minor allele, were non-responders to treatment with Tradipitant. Thus, the minor allele is a potentially predictive biomarker for poor response to treatment with Tradipitant, or it may indicate that these patients require a higher dose for Tradipitant to be efficacious in reducing WI-VAS. Conversely, AD patients whose genotype contains the major allele of this SNP are expected to have a larger positive response to treatment. Pruritus has a large impact on AD patients' quality of life; thus the results of this analysis are relevant as they stratify treatment response in reduction of worst itch and may help personalize therapeutic approaches.

## PgmNr 1892: Whole genome sequencing reveals novel rare loss-of-function variants in the epidermal differentiation complex as predisposing factors to atopic dermatitis.

### Authors:

S.E. Welsh; S. Smieszek; D. Xiao; J. Wang; C. Polymeropoulos; G. Birznieks; M.H. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals, Washington, DC

---

**Background:** The epidermal differentiation complex (EDC) includes over fifty genes encoding proteins involved in keratinocyte development. The proteins encoded are closely related in terms of function, and belong to three gene families: the cornified envelope precursor family, the S100 protein family and the S100 fused type protein (SFTP). Of these fifty genes, filaggrin (FLG) located on chromosome 1 q21, a member of the SFTP family, is the most studied in the context of skin barrier dysfunction.

**Methods:** We investigated the frequency and effect of rare loss-of-function (LOF) variants within the EDC in patients of a clinical study, VP-VLY-686-2102. Study VP-VLY-686-2102 was a randomized, double-blind, placebo-controlled, multi-center study in 168 patients with chronic pruritus associated with AD treated with tradipitant a novel neurokinin 1 (NK1) antagonist or placebo. The dataset consisted of 117 whole genome sequencing (WGS) samples. We were able to replicate the effect in the next batch containing 328 WGS samples part of clinical study VP-VLY-686-3101.

**Results:** We have shown that 45/117 AD patients carry significantly more, rare LOF mutations in the SFTP family of genes as compared to 55/316 in a control population ( $p=0.000004$ ). This group of EDC LOF rare variants (EDC-LR) consists of 20 variants observed in the 45 AD patients resulting in a calculated Odds Ratio of 2.96 and a Relative Risk of 2.38. Among the detected LOF variants, there are 26 cases of FLG LOF mutations as defined by R501X (rs61816761), 2282del4 (rs558269137), other LOFs in FLG as well as LOFs in FLG2, HRNR, LCE4A, LCE5A, TCHH, TCHHL1 and other members of the EDC. We were able to replicate the effect in another batch of 162 samples, OR 2.43,  $p=0.0001$  as compared with controls.

We examine the regional accumulation of rare LOF variants (with SKAT-O analysis), an effect that would be missed in a single marker genome-wide setting. To formalize, we performed two tests: on 1) the FLG region alone and 2) the entire EDC region (LOF set), comparing AD with controls (WGS). For the entire EDC, we obtained a p-value of  $4.7e-20$ , much lower than for FLG alone p-value of  $4.5e-6$ .

**Conclusions:** The two cohorts of whole genome sequencing of AD samples showed enrichment for rare variants in the EDC region in cases compared with controls. These significantly affect risk for AD. The identified LOF variants within the region can serve as biomarkers as well as help delineate the genetic profile in AD patients.

## PgmNr 1893: Evaluation of genetic variants in angiogenic growth factors with susceptibility to psoriasis.

### Authors:

R. Kaur; T. Akhtar; A. Wani; S. Kumar

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Human Genetics, Department of Human geneticsPUNJABI UNIVERSITY, PATIALA, INDIA, India

---

**Background:**Psoriasis is a chronic, immune-mediated inflammatory skin disease. The study was undertaken to analyse the polymorphic status of VEGF, FGF1 and FGF2 in psoriatic patients from North Indian population.

**Methodology:**Genotypes were analysed by PCR, RFLP, ARMS-PCR and tetra primer ARMS-PCR in 200 cases and controls. VEGF serum levels were measured by ELISA. Statistical analysis of the data was done by using SNPstats software, Graph-Pad prism v.5 and vassarstats.

**Results:**The dominant and recessive models showed thathomozygous genotypesof VEGF (rs1570360) play a protective role in psoriasis. However,over dominant model on the basis of gender showed no significant association.AA genotype was significantly associated with reduced risk of type1 psoriasis.On analyzing VEGF (rs2010963), homozygous CC genotype and minor C allele showed significant association with decreased risk of psoriasis. The dominant and overdominant models did not show any association whereas, recessive model showed that genotypes play a protective role. Moreover, the current study showed that no psoriasis risk is associated with VEGF (rs8333061) polymorphism. HT5 ATC haplotype was found to be associated with decreased risk of psoriasis. A significant correlation was found between VEGF serum concentrations measured before and after treatment. Serum VEGF level increased in patients, as compared to controls with significant correlation between serum VEGF levels (before and after treatment) and psoriasis severity. FGF1 Polymorphism of (rs34011) did not show any significant association with the risk of psoriasis whereas FGF1 (rs34010) homozygous AA and heterozygous CA genotypes withdominant and recessive models showed significant association with increased risk for psoriasis.The risk was enhanced in males patients with type1 and type 2 psoriasis.Overdominant model did not showed any association.Polymorphism of FGF2 (rs2922979) was found to be associated with increased risk of psoriasis.The dominant and recessive models also showed association of genotypes as risk factor for psoriasis

**Conclusion:**VEGF (rs1570360 and rs2010963) polymorphisms were found to be associated with decreased risk of psoriasis, whereas FGF1 (rs34010) and FGF2 (rs2922979) polymorphisms have association with increased risk of psoriasis. VEGF (rs8333061) and FGF1 (rs34011) showed no association with psoriasis. A strong association between clinical response to treatment and reduction of VEGF serum levels was also found.



## PgmNr 1894: Gene expression signatures in classical monocytes from African ancestry patients with systemic sclerosis.

### Authors:

P.S. Ramos <sup>1,2</sup>; W. da Silveira <sup>3</sup>; J.R. Wirth <sup>1</sup>; N. Wilson <sup>4</sup>; R.C. Wilson <sup>5</sup>; E.S. Hazard <sup>5</sup>; J.H. Nam <sup>2</sup>; J.C. Oates <sup>1</sup>; M.A. Cunningham <sup>1</sup>; D. Chung <sup>2</sup>; G. Hardiman <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Medical University of South Carolina, Charleston, SC, USA; 2) Department of Public Health Sciences, Medical University of South Carolina, SC, USA; 3) School of Biological Sciences, Queen's University, Belfast, UK; 4) SCTR Institute Research and Coordination Management, Medical University of South Carolina, Charleston, SC, USA; 5) Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, SC, USA

---

Systemic sclerosis (SSc) is a rare fibrotic autoimmune disorder. It disproportionately affects African ancestry (AA) individuals who, despite the higher disease severity, are dramatically underrepresented in research. Monocytes show heightened activation in SSc, and in AA relative to European ancestry (EA) individuals. Monocytes are thus a good target tissue for elucidating disease mechanisms. In this study, we sought to characterize differential gene expression of classical monocytes in SSc patients and unaffected controls of African ancestry. Classical monocytes (CD14++CD16-) were FACS-isolated from 17 female AA SSc cases and 18 female AA controls. Total RNA was used to prepare RNA-Seq libraries using Illumina's TruSeq RNA Exome kit. Upon sequencing on an Illumina HiSeq2500 instrument, data was analyzed by Rosalind, with a HyperScale architecture developed by OnRamp BioInformatics. Briefly, individual sample reads were aligned to the hg19 reference genome using STAR and quantified using HTseq. Differential expression analysis was implemented using DESeq2 and functional enrichment analysis was performed using Advaita iPathway Guide. A total of 743 genes showed differential expression (FDR P-value <0.4). The top differentially expressed genes include the collagen *COL9A2*, the protein phosphatase *PPP1R14B*, the tubulin *TUBB4B*, the kinase-binding *AKAP1*, the ubiquitin ligase *RNF146*, the heparanase *HPSE*, the nuclear factor NF-Kappa-B activator *TRAF3IP2*, and the chromatin regulator *SMARCA4*. The SSc monocyte transcriptome showed an enrichment of genes in the AMPK signaling pathway, genes involved in chromatin organization, transcription factor binding, and glycogen storage diseases. The top upstream regulator is the MAPK11 kinase. Unlike what has been reported in different peripheral blood subsets in EA patients, our results show a weaker upregulation of genes involved in immune and inflammatory processes in monocytes from AA patients. Instead, our study reveals an upregulation of genes involved in cellular processes associated with transcription and energy regulation, consistent with an increased metabolic rate of these myeloid cells in SSc. These results support the increasing awareness that metabolic reprogramming has important roles in mediating immune and vascular responses. Collectively, these results support the need to understand the regulatory architecture of SSc in different cell types and in individuals of different ancestries.



## PgmNr 1895: A burden of rare functional genetic variants associated with Chiari Type I malformation and the co-occurrence with connective tissue disorders.

### Authors:

K. Soldano <sup>1</sup>; M. Garrett <sup>1</sup>; A. Urbizu <sup>1</sup>; D. Loth <sup>2</sup>; A. Macaya <sup>3</sup>; F. Loth <sup>2</sup>; R. Labuda <sup>2</sup>; A. Ashley-Koch <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC; 2) Conquer Chiari Research Center, University of Akron, Akron, OH; 3) Pediatric Neurology Research Group, Vall d'Hebron Research Institute, Universitat Autònoma de Barcelona, Barcelona, Spain

---

Chiari Type I Malformation (CMI) is characterized by herniation of the cerebellar tonsils below the foramen magnum and affects approximately 1% of the American population. Patients with CMI may experience a range of symptoms from headaches, to more significant neurologic morbidity such as blurred vision, muscle numbness and weakness, as well as spasticity. Some forms of CMI co-occur with connective tissue disorders (CTD), such as Ehlers Danlos Syndrome (EDS) (Milhorat et al 2007), and may have a unique patho-mechanism due to cranio-cervical instability. Thus, CMI patients with CTD (CTD+) may have different genetic risk factors than those without a history of CTD (CTD-).

To identify putative genetic risk variants, we performed next generation sequencing (NGS) on samples from 165 adult, Caucasian females with CMI who had been ascertained through the Chiari1000 project at University of Akron or a CMI family study at Duke University Medical Center. Using an Illumina TruSeq Custom Amplicon Library Prep kit, we targeted the coding regions of 21 CMI and EDS candidate genes. Sequencing was performed on an Illumina MiSeq v2 using 150bp paired end reads. Reads were aligned and variants were called using GATK best practices. Using gene-level burden analysis, we compared the frequency of rare, functional variants detected in CMI cases versus publically available ethnically matched controls from gnomAD. A secondary analysis compared the presence of rare variants between CTD+ and CTD- CMI cases.

Six genes (*COL5A2*, *COL6A5*, *COL1A2*, *NRP1*, *VEGFB*, *FLT1*) had significantly more rare variants in CMI cases compared to public controls. Interestingly, 25% of CMI cases possessed rare variants in *COL6A5*. We also identified four genes (*COL7A1*, *CDX1*, *VEGFA*, *DSE*) that were significantly different between CTD+ and CTD- CMI cases. A higher percentage of CTD- patients had variants in *COL7A1*, *VEGFA*, and *DSE* compared to CTD+ patients, while more CTD+ patients had variants in *CDX1* than did CTD- patients.

In summary, we have identified a burden of rare functional variants in CMI cases as compared to controls in several candidate genes. Variants in *COL6A5*, in particular, seem to be associated with CMI. We also have preliminary evidence that rare variants are differentially associated with CTD+ and CTD- forms of CMI. Our findings underscore the contribution of rare genetic variants to CMI, and differences in genetic etiologies of CMI with and without CTD symptoms.

## PgmNr 1896: Differentially expressed miRNAs and mRNAs in psoriatic skin.

### Authors:

M. Loset<sup>1,2</sup>; A.A. Solvin<sup>1</sup>; K. Chawla<sup>1,3,4</sup>; L.C. Olsen<sup>1,3,4</sup>; K. Danielsen<sup>5,6</sup>; M. Jenssen<sup>5</sup>; A.S. Furberg<sup>6,7</sup>; M. Saunes<sup>2,3</sup>; K. Hveem<sup>1,8</sup>; P. Sætrum<sup>1,3,4,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 2) Department of Dermatology, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway; 3) Department of Clinical and Molecular Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 4) Bioinformatics Core Facility, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 5) Department of Dermatology, University Hospital of North Norway, Tromsø, Norway; 6) Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway; 7) Department of Microbiology and Infection Control, University Hospital of North Norway, Tromsø, Norway; 8) HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 9) Department of Computer Science, NTNU, Norwegian University of Science and Technology, Trondheim, Norway

---

Psoriasis is a chronic inflammatory disease of the skin believed to be triggered by environmental factors in genetically susceptible individuals. Gene expression data has increased our understanding of molecular contributors of psoriasis. However, few aberrantly expressed miRNAs in psoriasis have confirmed mRNA targets with biological functions in skin and a limited number of combined miRNA and mRNA sequencing studies have been performed. To increase our understanding of the psoriasis transcriptome and the miRNA-mRNA network, we sequenced small RNAs and ribosomal RNA-depleted total RNA of skin biopsies from patients with psoriasis vulgaris (n=75) and non-psoriatic controls (n=46). Comparing lesional psoriatic and control skin, we identified 270 differentially expressed ( $|\log_2FC| > 0.5$ , FDR < 0.05) miRNAs, of which 130 were upregulated and 140 were downregulated. The number of up- and downregulated ( $|\log_2FC| > 0.5$ , FDR < 0.05) mRNAs were 4,216 and 4,121, respectively. The differentially expressed mRNAs were analyzed by functional enrichment analysis using DAVID v6.8. For the upregulated genes, enriched functional annotation clusters included 'Immunity', 'Cell division' and 'Keratinization', and for the downregulated genes 'Glycoprotein/Cell membrane', 'Disulphide bond' and 'Cell adhesion'. Our results support a model for psoriasis as a result of perturbed interaction between keratinocytes and activated immune cells, leading to an increased proliferation rate of keratinocytes and epidermal thickening. We aim for further systematic analysis using pathway and network tools on miRNA and mRNA pairs with anti-correlated expression patterns. Increased knowledge of the transcriptome of psoriasis has the potential to provide new insight into the pathogenesis of psoriasis in addition to identification of biomarkers and therapeutic targets.

## PgmNr 1897: Comprehensive genomic analysis of data-driven dietary patterns in UK Biobank identifies novel loci and reveals causal associations.

### Authors:

J.B. Cole <sup>1,2,3</sup>; J.C. Florez <sup>1,2,4</sup>; J.N. Hirschhorn <sup>1,3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Programs in Metabolism and Medical & Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts; 2) Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, Massachusetts; 3) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital; 4) Department of Medicine, Harvard Medical School, Boston, Massachusetts; 5) Department of Genetics, Harvard Medical School, Boston, Massachusetts

---

Unhealthy diet is a leading risk factor for global morbidity and mortality. The modest, but real genetic contribution to dietary intake can now be explored in large-scale biobanks, such as UK biobank (UKB). We derived 170 data-driven diet traits from UKB questionnaires, including previously unstudied food intake quantitative traits (FI-QTs) and principal component-derived dietary patterns (PC-DPs). Linear mixed model genome-wide association studies (GWAS) on 143 significantly heritable diet traits in 450K Europeans identified 814 independent genetic loci ( $P < 5 \times 10^{-8}$ ), with 309 also surpassing a study-wide threshold ( $P < 3 \times 10^{-10}$ ). Of these, 205 are novel, and 447 were uniquely associated with either PC-DPs (136) or FI-QTs (311). GWAS of PC1, the most heritable PC-DP ( $h^2_g = 13\%$ ), identified the most loci ( $M = 140$ ) of any diet trait, explaining 6.5% of PC1 phenotypic variance. PC1, similar to previously described prudent vs. Western diet factors, captures positive wholemeal bread consumption and higher amounts of fruit, vegetables, fish, and water and negative white bread, butter/oil spreads, and higher amounts of processed meat and high-fat milk. Though Western and prudent dietary factors have been observationally associated with education and cardiometabolic disease, bidirectional Mendelian randomization (MR) revealed that while higher education shifts eating from Western towards healthier prudent eating (weighted median  $\beta = 0.82$ ,  $P = 6.6 \times 10^{-113}$ ), genetically-predicted prudent eating does not demonstrate causal evidence for reducing type 2 diabetes or coronary artery disease risk. We also identified several olfactory receptor loci associated with specific FI-QTs. SNP rs1453548, explaining >96% of sensitivity to beta-ionone aroma, is associated with increased tea intake. Comprehensive MR of this SNP demonstrates no widespread health impact of tea drinking; apparent causal evidence for decreasing smoking status is likely due to pleiotropy (beta-ionone is also in tobacco). In sum, comprehensive genomic analysis on diet has identified hundreds of associations and highlights the importance of complementary data-driven phenotyping. These results enable tests of causality for previous observational diet-outcome associations, advance our understanding of genetic architecture of diet, and inform public health.

## PgmNr 1898: The magnitude and impact of population differences on genetic risk scores for cardiometabolic traits.

### Authors:

K. Ekoru; A.A. Adeyemo; G. Chen; A.P. Doumatey; J. Zhou; A.R. Bentley; D. Shriner; C.N. Rotimi; the AADM study investigators

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

---

Proposals for using genetic risk scores (GRS) for risk prediction in routine clinical settings have raised concerns about potential health inequalities that could result from limited predictive power of GRS in non-European populations. To provide a context for the implications of routine use of GRS in diverse populations, we assessed the predictive utility of NHGRI-EBI Catalog-derived GRS for cardiometabolic traits in 5,200 sub-Saharan Africans (AF), 9,594 European ancestry (EA) and 9,139 African American (AA) individuals. The predictive utility of the GRS was assessed using the additional phenotypic variance explained and increase in discriminatory ability over traditional risk factors. GRS-trait correlations were found to be strongest in EA. Predictive disparities of GRS between AF and AA were less pronounced but were better in AA, perhaps reflecting European admixture in AA. Among obesity related (OB), blood pressure (BP), lipids and glycemic traits that were evaluated, lipids had the clearest and most consistent GRS-trait gradient and predictive utility across all three ancestry groups. Total cholesterol (TC) showed the strongest GRS-trait association in all groups with GRS effect sizes of 0.226, 0.216 and 0.281 mmol/l per unit increase in GRS (all  $P < 0.0001$ ) among AF, AA and EA, respectively. The percent increase in explained variation attributable to GRS for triglycerides (TG), TC, low-density lipoprotein (LDL), high-density lipoprotein (HDL) and type 2 diabetes was, respectively: 2.4%, 25.1%, 31.0%, 37.5% and 12.4% among AF; 32.0%, 31.3%, 26.6%, 65.8% and 8.1% among AA; and, 47.2%, 95.4%, 127.1%, 10.6% and 25.5% among EA; showing up to a 20-fold and 5-fold greater predictive utility of GRS in EA relative to AF and AA, respectively. The above disparities were recapitulated in discriminatory power, whereby the predictive utility of GRS was 44-, 5- and 4-fold greater in EA relative to AF for raised TG, raised TC and raised LDL respectively; and 4-fold greater for raised TG, LDL and T2D in EA relative to AA. GRS of OB and BP traits showed a similar pattern. While the predictive performance of GRS tends to be lower in non-European populations, we show that it performed more poorly in African ancestry individuals with sub-Saharan Africans faring the worst. Further coordinated and sustained efforts are needed to increase the inclusion of underrepresented populations in genomic discovery to promote equitable benefit from translation of such discovery.

## PgmNr 1899: Discovery of 159 signals for random plasma glucose and dissection of its relationship with type 2 diabetes pathophysiology using GWAS within UK biobank.

### Authors:

M. Kaakinen <sup>1,2</sup>; L. Jiang <sup>2,3</sup>; A. Ulrich <sup>2</sup>; L. Zudina <sup>1,2</sup>; Z. Balkhiyarova <sup>1,2</sup>; Y. Timasheva <sup>2,4</sup>; V. Lagou <sup>5</sup>; I. Prokopenko <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Biosciences and Medicine, Department of Clinical and Experimental Medicine, University of Surrey, Guildford, United Kingdom; 2) Section of Genomics of Common Disease, Department of Medicine, Imperial College London, London, United Kingdom; 3) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia; 4) Department of Medical Genetics and Fundamental Medicine, Bashkir State Medical University, Ufa, Russia; 5) VIB Center for Brain & Disease Research, Leuven, Belgium

---

Fasting plasma glucose (FG) and 2-hour post-prandial plasma glucose (2hGlu) are used as the gold standard tests for diagnosing type 2 diabetes (T2D). High non-fasting i.e. random plasma glucose (RG) levels may also indicate abnormal glucose homeostasis. We sought to characterise the genetic architecture of RG and its genetic relationships with other glycaemic traits and T2D. We performed a genome-wide association study of SNPs with MAF>1% for RG in non-diabetic individuals from the UK Biobank, adjusting for 1) age, sex and population structure (model1, N=401,810) and 2) additionally for time since last meal (model2, N=401,799). The LD score regression was used to evaluate the genetic correlation between RG, FG, 2hGlu, fasting insulin, glycated hemoglobin (HbA1c), homeostasis model assessment of beta cell function/ insulin resistance (HOMA-B/-IR), and T2D. We identified 159 loci associated with RG at genome-wide significance ( $P<5\times10^{-8}$ ), including two common coding variants within *GCKR* and *TET2* and four low-frequency nonsynonymous variants within *EDEM3*, *HERC1*, *NEUROD1* and *GLP1R*. Six loci reached statistical significance only when adjusting for time since last meal (*ADCY9/SRL*, *KCNQ1*, *ERN1*, *HECTD1/HEATR5A*, *SLC38A4*, *PDHX*), and nine without the adjustment (*MTNR1B*, *LOC440040*, *VEGFA/LINC01512*, *TRIM48*, *MIR588/RSPO3*, *RGS17*, *RREB1*, *RFX6/VGLL2*, *RREB1*). Of all the loci, 47 are established for T2D, three for FG, two for fasting insulin, two for fasting proinsulin and one for 2hGlu. Of the 367 established T2D signals with MAF>1%, 45% are nominally significantly associated with RG. We observed strong ( $r>0.6$ ) genetic relationships between RG and FG ( $r[SE]=0.86[0.05]$ ,  $P=1.04\times10^{-65}$ ), modest ( $0.3\leq r\leq 0.6$ ) with T2D ( $r[SE]=0.48[0.06]$ ,  $P=2.19\times10^{-15}$ ), HbA1c ( $r[SE]=0.43[0.06]$ ,  $P=1.04\times10^{-12}$ ) and HOMA-B ( $r[SE]=-0.39[0.07]$ ,  $P=2.35\times10^{-8}$ ), and weaker genetic correlation with HOMA-IR ( $r[SE]=0.28[0.07]$ ,  $P=4.31\times10^{-5}$ ). Our large-scale analysis of RG has detected FG- and T2D-associated variants, and the genetic correlation analysis also shows large genetic overlap between RG, FG, T2D and other glycaemic traits. Replication analysis in over 80,000 non-diabetic Europeans is on-going.

## **PgmNr 1900: LDL-C and type-2 diabetes: An inverse phenotypic association and genetic drivers in the UK Biobank.**

### **Authors:**

Y. Klimentidis <sup>1</sup>; A. Arora <sup>1</sup>; M. Newell <sup>1</sup>; A. Wood <sup>2</sup>; J. Ordovas <sup>3</sup>; J. Zhou <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Epidemiology and Biostatistics, University of Arizona, Tucson, Arizona.; 2) Baylor College of Medicine, Houston, TX.; 3) Tufts University, Boston, MA.

---

Although a low level of low-density lipoprotein cholesterol (LDL-C) is associated with protection against cardiovascular disease, some lines of evidence, including statin trials, suggest that lower LDL-C increases the risk of type-2 diabetes (T2D), and in some cases, obesity. Consistent with this phenotypic evidence, previous studies have also identified complex genetic relationships of lipids with T2D. Using data from the UK Biobank, we examined the prevalence of T2D and related traits across the range of LDL-C levels. We then used a genome-wide approach with several datasets to identify variants with opposing directions of effect on LDL-C and T2D. We find a higher prevalence of T2D, but not higher HbA1c or BMI, among individuals with low LDL-C. We identified known and novel genetic loci simultaneously associated with lower LDL-C and increased T2D (and vice-versa). These loci point to the liver as a principal tissue mediating these effects, and to pathways of triacylglyceride synthesis and glycerolipid metabolism. We thus confirm previous findings regarding the higher T2D risk among those with low LDL-C, and provide: 1) important insights into the mechanisms underlying this association, 2) more detailed phenotypic characterization of previously identified genetic variants, and 3) insight into the mechanisms potentially underlying the diabetogenic effect of LDL-C-lowering medications.

## **PgmNr 1901: Association between the high T2D prevalence and the genetic variants of HDAC4 and GCK interacting with heavy alcohol consumption in Korea.**

**Authors:**

H. Lee; H. Jang; J. Yoon; M. Yoo; S. Park

[View Session](#) [Add to Schedule](#)

**Affiliation:** NIH Korea, Cheongju, South Korea

---

Heavy alcohol consumption is known as an independent risk factor for type 2 diabetes (T2D), which is related with impaired glucose homeostasis and insulin resistance. The identification of genetic variants and the association with diseases may help to explain the genetic etiology of T2D. In this study, we aimed to search for the single nucleotide polymorphisms (SNPs) associated with T2D risk, which interact with heavy alcohol consumption. We conducted genome-wide interaction study using the Korea Association Resource (KARE) data. We found the two SNPs, HDAC4 (rs3791370) and GCK (rs758989) interact with heavy alcohol consumption in prevalence of T2D (odds ratio [OR] = 1.74, and OR=1.86, respectively). We further investigated the association between these SNPs and response to heavy alcohol consumption in risk of T2D prevalence after adjustment with age, BMI, smoking, physical activity, and history of diabetes. In heavy alcohol consumption group, hetero risky allele [TT+C/ C+TT] (OR=1.7) and minor homo risky allele [C+C] (OR=4.4) significantly increased the risk of T2D prevalence compared with non-alcohol consumption group. Moreover, we analyzed the insulinogenic index (IGI60), and Matsuda insulin sensitivity index (ISI) by combined genetic variants in HDAC4 and GCK and heavy alcohol consumption. The presence of risky alleles in heavy alcohol consumption group significantly decreased IGI60 levels. We found out two new genetic variants of HDAC4 and GCK, which interact with heavy alcohol consumption as an environmental factor in association with the prevalence of T2D and may be helpful to better understand the genetic basis of T2D.



## PgmNr 1902: Sex-specific changes in fat accumulation and metabolic parameters in a murine knockout model of *Lyplal1*.

### Authors:

R. Vohnoutka <sup>1</sup>; Y. Hegde <sup>1</sup>; Y. Chen <sup>1</sup>; B.D. Halligan <sup>1</sup>; E.K. Speliotes <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Gastroenterology and Internal Medicine, University of Michigan, Ann Arbor, MI 48109; 2) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109

GWAS studies have associated sex specific SNPs near *LYPLAL1* with central obesity as well as insulin levels and NALFD. The *LYPLAL1* protein is an acyl protein thioesterase that function to alter the location by deacylating proteins and changing their subcellular localization. We used CRISPR/Cas9 to generate a whole body *LYPLAL1* knockout (KO) mouse. We placed wild type (WT) and *Lyplal1* KO mice on a 'Western' diet containing levels of high fat and sucrose (HF/HS diet). During the 23 weeks that the mice were on the HF/HS diet, female KO mice gained less weight than their WT littermates. We carried out body composition measurements by TD-NMR at 14 weeks on the HF/HS diet and there was no change in lean body mass but less fat in female *Lyplal1* KO mice. Similar differences were not observed with male mice. To measure the detailed metabolic phenotype of the mice, after 20 weeks on the HF/HS diet, CLAMS phenotyping showed that was no differences in food intake or activity. Male KO mice showed an increase in fat metabolism but female mice did not. In GTT and ITT tests female *Lyplal1* KO mice were more responsive to glucose and insulin but male mice did not respond differently. After 23 weeks on HF/HS diet, mice were sacrificed, dissected and organs and fat depots weighed. Female KO mice had reduced visceral fat (gonadal white adipose tissue [gWAT], perirenal WAT [prWAT], and subcutaneous fat (inguinal WAT [iWAT]) but not brown adipose tissue [BAT] with gWAT the most significant. Similar changes were not observed in male mice. Blood chemistry values showed an increase in serum triglycerides and a decrease serum cholesterol in KO female mice and levels of Alanine transferase (ALT) and aspartate aminotransferase (AST) were lower in KO female mice compared to WT littermates. There was no difference in serum triglycerides or liver enzymes between the genotypes of male mice, but KO male mice had lower serum cholesterol than male WT mice.

These data collectively indicate a role for *LYPLAL1* in the metabolic response of female mice to a HF/HS 'Western' diet, with KO of *LYPAL1* protein decreasing visceral and subcutaneous adiposity, as well as improving glucose response and decreasing liver damage. Further examination of the mechanism of these novel diet-induced, sex-specific metabolic effects of *Lyplal1* KO will be critical to increasing our understanding of gender effects on human obesity, liver disease and metabolic syndrome.



## PgmNr 1903: Fine mapping of T2D linked 12q24 region in Finnish families.

### Authors:

O. Dwivedi<sup>1</sup>; T. Tuomi<sup>2</sup>; L. Groop<sup>1,3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Finland; 2) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland; 3) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden

---

**Aim:** Previously, a linkage for type 2 diabetes (T2D) was reported at the 12q24 (104-133Mb, b37) region in Finnish families with impaired insulin secretion (Mahtani M *et al.*, Nat Genet., 1996). We here aimed to fine-map the linkage peak in order to identify the putative causal variant/haplotype in the region.

**Material and Methods:** The linkage study included 34 individuals from six families. All family members underwent a dense GWAS and 12 individuals whole genome sequencing (WGS). We first performed family-based association (linear mix model), haplotype sharing analysis and checked the effect on insulin secretion at 30 min (INS30) levels in an independent sample (N>3500). Additionally, we also checked effect of identified variants on expression of nearby genes in human islets (N=191) and performed further validation in a human beta cell line.

**Results:** Family based association revealed most significant association with T2D ( $P=1.4 \times 10^{-4}$ ) for a common haplotype consisting of three missense mutations (rs14259-rs17852561-rs1169081). Association of variants shared by > 5 families with INS30 levels in independent samples revealed most prominent association ( $P<5.9 \times 10^{-5}$ ) for cluster of variants (like rs28629903 and rs1316787) located near the linkage peak and this cluster also influenced the expression of the *MLXIP* gene in colon. A haplotype sharing analysis using >0.5 cM long haplotypes, shared and transmitted among affected members of at least >2 families, revealed two haplotypes shared at regions 124.8 to 125.0 mb (*NCOR2*), and 127.85 to 128.5 mb (near to *TMEM132C*). The shared regions contain low frequency (0.01 to 4%) variants (rs11057587, rs11059329, rs148549814) which were associated with decreased INS30 levels and increased expression only of *TMEM132C* ( $P=0.004$ ) in human islets. *TMEM132C* expression was negatively correlated with insulin secretion in human islets in our previous work (Ottosson-Laakso E *et al.*, Diabetes, 2017) and siRNA mediated knock down in human beta cells leads to increase in insulin secretion ( $P=0.002$ ).

**Conclusion:** The 12q24 linkage families share clusters of common and low frequency variants that affect insulin secretion with the *MLXIP*, *TMEM132C* and *NOCR2* genes as plausible candidates. Of them, *TMEM132C* may be most interesting as functional data support a role in regulation of insulin secretion, *MLXIP* (*MondoA*) would also need further work to explore its potential involvement in impaired insulin secretion influencing the incretin pathway.

## PgmNr 1904: Pathway-based genetic risk scores in T2D risk predict associated phenome and related biomarkers.

### Authors:

A. Faucon; X. Zhong; P. Straub; J. Sealock; L. Davis; N. Cox

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt Genetics Institute, Vanderbilt, Nashville, Tennessee.

Type 2 Diabetes (T2D) is an important, complex and multi-faceted disease. In the United States, the prevalence of T2D has more than tripled in the years between 1998 and 2018. T2D is the number 7 cause of death in the United States and is also associated with increased risk of stroke, heart disease, chronic kidney disease, among other illnesses and with approximately \$327 billion annually for lost work and wages and medical costs. Diabetes is further complicated by comorbidities, which can cost more and further decrease average health-related quality of life (HRQoL) from an optimal 1 to 0.85 in patients with T2D without complications, to 0.64 in patients with T2D and two or more complications. It will be valuable to improve accuracy in predicting who will develop further complications and what types of complications these may be.

In this set of research, we sought to use genetics to elucidate biological mechanisms in phenotypically similar outcomes. We used genetic risk scores (GRSs) built from T2D GWAS variants, and subset them to three biological pathways (Insulin Secretion, Insulin Action, BMI-Dyslipidemia). We utilized a portion of the genotyped Vanderbilt medical center biobank (BioVU) from which we scored each participant in our dataset based on their GRS in each pathway and ran analyses examining associations of each of these scores with median lab values and presence of billing codes. We analyzed the association of each pathway GRS with up to 1814 billing codes and 260 median lab values. Liver disease and renal failure related billing codes were significantly associated only with the BMI-Dyslipidemia risk pathway (OR=1.14, P=8e-8 ) & (OR 1.11 P=1e-7), respectively, not with Insulin Secretion (OR=0.99, P=0.566) & (OR=1.00, P=0.874) nor Insulin Action pathways (OR=1.07, P=0.002) and (OR=1.05, P=0.015), respectively. Use of BMI as a covariate removed signals of statistical significance (Bonferroni p-value= in associations between the BMI-Dyslipidemia GRS and blood glucose (OR=1.043, P = 6.7e-6) and blood urea nitrogen (OR=1.040, P= 1.6e-5), but adjustment for BMI had no impact on signals between the other two pathway GRSs and their related labs. This indicates that BMI may be a more useful measurement to predict downstream risk in participants with T2D risk sourced from the BMI-dyslipidemia pathway.

Moving forward, we aim to leverage BioVU and other EHRs to find other significant biomarkers and comorbidities relating to biological pathways of risk.

## PgmNr 1905: Discovery of kidney function loci and rare variants in multigenerational diabetes pedigrees using variance component linkage analyses and next generation sequencing.

### Authors:

S. Frodsham <sup>1</sup>; B. Moore <sup>2</sup>; T. Dyer <sup>3</sup>; D. Reilly <sup>4</sup>; M. Hansen <sup>4</sup>; J. Lazaro-Guevara <sup>1,2</sup>; M. Pezzolesi <sup>1</sup>; J. Fierro Morales <sup>1</sup>; E. Javier Hernandez <sup>2</sup>; R. Duggirala <sup>3</sup>; J. Blangero <sup>3</sup>; A. Parrado <sup>4</sup>; M. Yandell <sup>2</sup>; M. Pezzolesi <sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Nephrology, University of Utah, Salt Lake City, Utah.; 2) Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah; 3) South Texas Diabetes and Obesity Institute, University of Texas at the Rio Grande Valley, Brownsville, TX; 4) Janssen Research and Development, LLC., Spring House, PA and Boston, MA; 5) Diabetes and Metabolism Research Center, University of Utah School of Medicine, Salt Lake City, Utah

---

Chronic kidney disease (CKD) is a common complex disease with an especially high prevalence in individuals with diabetes. CKD is hallmarked by progressive decline of glomerular filtration rate (GFR), a measure of kidney function, that often culminates in a state of complete renal failure known as end-stage renal disease (ESRD). A recent genome-wide association study (GWAS) meta-analysis of over 1 million individuals identified more than 300 loci associated with kidney function. Variants at these loci explain approximately 20% of the genetic heritability of estimated GFR (eGFR). We hypothesize that rare or lower frequency variants (minor allele frequency <5%) contribute significantly to the residual genetic variance of kidney function and that linkage analysis in large multigenerational pedigrees combined with next generation sequencing methods can not only help validate the findings of large GWASs, but can also identify novel loci of interest in CKD populations. Here, we apply this approach in the context of large multigenerational diabetes pedigrees identified using the Utah Population Database (UPDB), a unique population-based genealogy resource of Utah pioneers and their descendants that is linked to electronic health record data from the University of Utah Health Sciences Center. Using the UPDB, we identified 20 large multigenerational pedigrees, consisting of 1,581 individuals, with a high prevalence of CKD. 230 members of these pedigrees were genotyped using Illumina's Infinium CoreExome microarray. Variance component linkage analysis using SOLAR was conducted to identify eGFR-associated loci. As a complimentary approach, analysis of kidney-related traits and whole-genome sequencing data on a subset of these families was performed using pVAASST. Together, this combined approach highlights several rare missense variants in genes previously associated with CKD and has identified several novel genes that are likely associated with renal function in patients with diabetes.

## PgmNr 1906: A genome-wide CRISPR screen in human beta-cells for regulators of insulin content.

### Authors:

A.K. Grotz <sup>1</sup>; E. Navarro Guerrero <sup>2</sup>; R. Baronio <sup>2</sup>; S.K. Thomsen <sup>1</sup>; A. Wesolowska-Andersen <sup>3</sup>; D. Ebner <sup>2</sup>; A.L. Gloyn <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oxford Centre for Diabetes, Endocrinology & Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK; 2) Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK; 3) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; 4) National Institute of Health Research Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK

---

Genome-wide association studies have identified >400 signals in the genome associated with type 2 diabetes (T2D) risk, a large proportion of which are involved in pancreatic beta-cell dysfunction. We have previously shown how a targeted knockdown screen for insulin secretion can aid prioritisation of effector transcripts at these loci. To build on this work, we now extend this genome-wide using a pooled CRISPR/Cas9 knockout (KO) screen for insulin content in human beta-cells.

We implemented the pooled CRISPR/Cas9 screen in the human pancreatic beta-cell line EndoC-βH1, which is capable of physiologically relevant glucose-responsive insulin secretion. Insulin content was selected as the screening phenotype, as it is compatible with a FACS-based pooled screen. To validate our pipeline, we performed a small-scale screen using previously validated single guide RNAs (sgRNAs) for five genes and assessed their enrichment in cells sorted for low vs high insulin content. sgRNAs for *INS* and *PAM*, two genes with known negative effects on insulin content were enriched in the low insulin population, whereas control sgRNAs showed no enrichment (fold change *INS*: 3.58, *PAM*: 1.32, Ctrl: 1.07).

We performed two independent screens and transduced 650 million cells each with the genome-wide CRISPR library Toronto KnockOut v3 (TKOv3), which targets 18 053 protein-coding genes with 4 sgRNAs/gene. To account for different temporal effects of individual gene KOs and the slow growth rate of EndoC-βH1, cells were collected for FACS at two time points, immediately after puromycin selection and 14 days later. Cells were stained for intracellular insulin and we observed a wider distribution of insulin signal in TKOv3 transduced cells compared to control cells, indicating a wide range of KO-induced effects on insulin content. Low and high insulin content populations were sorted and sgRNA enrichment or depletion was analysed using next-generation sequencing to identify genes that have an effect on insulin content or are essential in beta-cells.

In summary, we have performed a genome-wide pooled CRISPR/Cas9 screen for insulin content in an authentic human beta-cell model. Our hits can now be integrated with genetic and epigenomic data at T2D-risk loci to prioritise effector transcripts for further mechanistic studies. Our unbiased CRISPR screening strategy has also identified unknown regulators of beta-cell function and survival which may aid cell replacement strategies for diabetes.

## PgmNr 1907: Understanding the genetic basis of type 2 diabetes risk: Whole genome sequence analysis of glycemic traits from the NHLBI's TOPMed Program.

### Authors:

A. Manning <sup>1</sup>; J. Wessel <sup>2</sup>; D. DiCorpo <sup>3</sup>; T. Majarian <sup>4</sup>; Y. Chen <sup>1</sup>; S. Gaynor <sup>5</sup>; TOPMed Diabetes Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, Massachusetts.; 2) Department of Epidemiology, Fairbanks School of Public Health, Indiana University, Indianapolis, IN, USA.; 3) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 4) Metabolism Program, Broad Institute, Cambridge, MA; 5) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.

---

Genetic studies have shown that for type 2 diabetes (T2D) risk and diabetes-related glycemic traits, common variant heritability is enriched within active regulatory regions of tissues important to diabetes and insulin resistance. Whole genome sequence (WGS) association analysis allows us to (1) fine-map known and novel loci with active regulatory regions, (2) determine the contribution of rare variants in active regulatory regions to the heritability of T2D risk and glycemic traits, and (3) perform rare variant tests which aggregate protein-coding variants and/or variants in active regulatory regions to increase statistical power. Here, we present a complete WGS analysis of fasting insulin (FI) and fasting glucose (FG) levels in phase 2 Trans-Omics for Precision Medicine (TOPMed) data (fasting glucose [FG] N=26,920 and fasting insulin [FI] N=23,361) with deep (>30x) sequence coverage in fourteen cohorts across five ancestry groups. For our single variant analysis, we restricted analysis to variants with minor allele count > 20 in ancestry-specific and cross-ancestry pooled analysis. We used linear mixed effects models (R/Bioconductor GENESIS in cloud-computing platforms) and adjusted for sex, age and BMI, with empirical kinship for relatedness and population structure. For rare variant analysis, we derived aggregation units from coding variant annotations, regulatory annotations from pancreatic islets, liver, muscle and adipose tissues, and annotation Principal Components (aPCs), a multi-dimensional summary of functional annotation scores. We used sequence kernel association test (SKAT) and the novel variant-Set Test for Association using Annotation infoRmation (STAAR), a general framework that incorporates annotation scores using an omnibus multi-dimensional weighting scheme. In single variant analysis, we observe 8 regions associated ( $P < 5 \times 10^{-8}$ ) with FG and 6 regions associated with log-FI, with most signals from common variants in known regions. In rare variant analysis, one test showed a (Bonferroni) significant association with FG using missense variants in *G6PC2*. At the *G6PC2* locus, multiple distinct associations were observed in the single variant tests, consistent with previous reports, and nominal significant associations were observed when aggregating predicted loss-of-function rare variants. Analysis with larger samples are planned. Our work refines the genetic architecture of glycemic traits using the full spectrum of genetic variation.

## PgmNr 1908: Genomics as a personalized medicine approach in type 2 diabetes risk prediction - P5 FinHealth.

### Authors:

H. Marjonen <sup>1</sup>; M. Marttila <sup>1</sup>; T. Paajanen <sup>1</sup>; H. Kääriäinen <sup>1</sup>; K. Auro <sup>1</sup>; K. Kristiansson <sup>1</sup>; M. Perola <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Institute for Health and Welfare, Helsinki, Finland; 2) Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Finland

---

In P5 FinHealth study we utilize polygenic risk scores (PRS) to provide personalized information on the individual disease risk related to three common diseases (coronary heart disease, type 2 diabetes and venous thromboembolism) for 3.400 volunteering participants. Participants will receive this disease risk information including genetic risk via a secure web portal. We study the value of returning genetic risk information and hypothesize that it would improve prevention, diagnostics and treatment.

We used polygenic risk scores containing up to 7 million genomic regions and validated them in whole genome genotyped population based FINRISK cohorts (N=20.000) using Cox regression models. Follow-up data from national health care registers allowed us to model the impact of genetic and traditional risk factors such as smoking, cholesterol, blood pressure and body mass index (BMI) on a person's risk of disease within the next 10 years.

Altogether 967 T2DM new cases were observed during the 10-year follow-up period. We observed that type 2 diabetes (T2DM) PRS significantly associates with the T2DM disease risk (HR:1.5 per 1 sd PRS, p-value:<2\*10<sup>-16</sup>). Also the top 8% of the FINRISK population who had inherited the highest PRS had fourfold increased risk for T2DM. Moreover, almost 30% of the people with >35 BMI and the highest PRS were diagnosed with T2DM during ten-year follow-up. T2DM incidents in >30 BMI group occurred at a younger age concurrently with higher PRS, and there was a seven-year difference of getting T2DM between the highest and lowest PRS group.

By combining the systemic genetic analyses with more traditional disease risk factors in the FINRISK cohort, we produced estimates on the impact of PRS and selected covariates on risk of T2DM. We use these estimates to assess the future risk of T2DM in P5 FinHealth participants. We start returning personalized health reports to the P5 participants during 2019. We will also provide a guidance tailored for each risk category to assist the health care via the study participant. The reception of the risk information will be monitored by questionnaires and new incidents followed using national health registers for five years. Our approach enables to identify the individuals within highest genetic risk and those with pre-disease symptoms.

## PgmNr 1909: Inactivating mutations in *SLC39A5*: Role in metabolic disease.

### Authors:

H. Nistala<sup>1</sup>; S.M. Chim<sup>1</sup>; J. Dronzek<sup>1</sup>; C. van Hout<sup>1</sup>; S. Brydges<sup>2</sup>; W. Wu<sup>2</sup>; A. Li<sup>1</sup>; K. Neiman<sup>2</sup>; G. Polites<sup>2</sup>; A. Marcketta<sup>1</sup>; L. Mitnaul<sup>1</sup>; D. Carey<sup>3</sup>; C.J. Schoenherr<sup>1,2</sup>; A. Shuldiner<sup>1</sup>; A.N. Economidis<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Regeneron Genetics Center, Tarrytown, NY.; 2) Regeneron Pharmaceuticals, Tarrytown, NY.; 3) Weis Center for Research, Geisinger Medical Center, Danville, PA

---

Recent human and mouse studies have underscored the role of zinc in insulin secretion and systemic glucose homeostasis. Randomized placebo-controlled zinc supplementation trials in humans demonstrated improved glycemic traits in patients with type 2 diabetes (T2D). Moreover, carriers of rare loss-of-function variants in *SLC30A8* (encoding an islet zinc transporter) have been reported to have 65% reduced T2D risk in a large multi-ethnic study. To further investigate the connection between zinc and metabolic traits, we employed a pathway-based target identification approach involving genome-wide association of loss-of-function variation in zinc transporters with glycemic traits and T2D risk. Using exome sequence data from participants of European ancestry in the Regeneron Genetics Center-Geisinger Health System DiscovEHR study, we identified rare loss-of-function variants (MAF<1%) in *SLC39A5* (gene encoding Solute Carrier Family 39, Member 5) that are associated with circulating zinc levels and T2D. Loss-of-function burden tests of *SLC39A5* were associated with increased circulating zinc levels (Beta = 1.1 SD, 95% CI 0.47-1.1,  $p = 5.4 \times 10^{-4}$ ) and decreased T2D risk (OR = 0.5, 95% CI 0.37-0.79,  $p = 1.6 \times 10^{-3}$ ). Furthermore, we report a mouse model of *Slc39a5* loss-of-function that results in significant elevation in circulating zinc levels and concomitant improvement in hepatic steatosis and hyperglycemia resulting from diet-induced obesity or leptin-receptor deficiency. Lastly, ongoing studies demonstrate that the improved liver function in these mice results from hepatic AMPK activation and increased hepatic lipid metabolism. Taken together, these results suggest *SLC39A5* as a potential therapeutic target for fatty liver disease and consequent metabolic derangements including T2D.



## PgmNr 1910: Association of ancient adaptive mitochondrial DNA haplogroups with type 1 diabetes.

### Authors:

J. Qu<sup>1,2</sup>; X. Chang<sup>1</sup>; H. Qu<sup>1</sup>; P. Sleiman<sup>1,3</sup>; C. Polychronakos<sup>4</sup>; H. Hakonarson<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Children's Hospital of Philadelphia, Philadelphia, PA.; 2) Texas A&M University, College Station, TX; 3) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) McGill University Health Center, Montréal, QC

---

The prevalence of type 1 diabetes (T1D) varies significantly across human populations, with the highest prevalence seen in Caucasian populations. Even among European populations, there are conspicuous differences with the highest prevalence in Finnish population (Patterson et al. 2012). To date, more than 70 T1D loci have been identified, however, the population difference of T1D susceptibility is still largely unknown. It has been well recognized that different ancient adaptive mitochondrial DNA (mtDNA) haplogroups have significantly different functional performance, thus have undergone extensive selection in the evolution of human populations (Tranah et al. 2011). In our study, we examined the association of mtDNA haplogroups with T1D in our T1D GWAS datasets. Population substructures of the sample were analyzed by principal component analysis (PCA) using PLINK1.9, and genetic outliers were removed. Altogether, 979 cases and 2009 controls of European ancestry were used for the association test. We determined the mtDNA haplogroups by the phylogenetic tree based software HaploGrep 2 (Weissensteiner et al. 2016), based on 163 mtDNA single nucleotide variants (SNV) genotyped by the Illumina HumanHap550 arrays. We tested 23 mtDNA haplogroups (A, B, C, D, F, G, H, I, J, K, L, M, N, P, Q, R, T, U, V, W, X, Z, and Undetermined) for genetic association by logistic regression adjusted for the first two principal components of the PCA for population stratification, with statistical significance Bonferroni correction level at  $\alpha = 0.05/23 = 2.17 \times 10^{-3}$ . One haplogroup was significant, i.e. Haplogroup V, OR (95%CI) = 3.200 (1.680, 6.097),  $P = 4.05 \times 10^{-4}$ . Another haplogroup demonstrated nominal significance without Bonferroni correction, i.e. Haplogroup K OR (95%CI) = 0.698 (0.528, 0.923),  $P = 0.012$ . Haplogroup V is most commonly seen in northern Scandinavian, which explains its association with the risk of T1D, but not simply marking Scandinavians (before correction OR (95%CI) = 3.627 (1.912, 6.879),  $P = 2.55 \times 10^{-5}$ ). In addition, we identified possible gender effect of Haplogroup K on the age-of-onset of T1D with nominal significance ( $P = 0.048$ ), i.e. later age-of-onset in males [mean of males vs females difference (95%CI) = 3.45 (0.033, 6.86)], which as well as the decreased risk of T1D might be explained by its function as an “uncoupling genome” and less oxidative damage, but the sexual dimorphism (Ventura-Clapier et al. 2017) warrants further investigations.



# PgmNr 1911: Family-based association analysis of type 1 diabetes replicates case-control associations and reveals sequence homology as a source of spurious association in exome chip analysis.

## Authors:

C.C. Robertson; W.M. Chen; S.S. Rich; S. Onengut-Gumuscu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Public Health Genomics, Univ Virginia, Charlottesville, Virginia.

---

**Background:** Type 1 diabetes (T1D) is a complex immune-mediated disease. Twin studies indicate about 50% of T1D risk is due to genetic susceptibility. Genome-wide association studies (GWAS) in unrelated subjects identified >40 regions affecting T1D risk, with fine mapping increasing resolution of putative causal variants. Despite substantial familial clustering, unbiased family-based GWAS have not been performed with large sample sizes. Here, we analyze a large cohort of affected families from the Type 1 Diabetes Genetic Consortium (T1DGC).

**Methods:** We genotyped 13,461 participants from 4,520 affected families using the Illumina HumanCoreExome BeadChip. We confirmed the accuracy of reported relationships with genotype-based relationship inference and tested T1D association using the transmission disequilibrium test. We evaluated novel associations in an independent collection of 2,150 T1D cases and controls with whole genome sequencing.

**Results:** Five known regions reached genome-wide significant ( $p < 5 \times 10^{-8}$ ) association with T1D - 1p13.2 (*PTPN22*), 6p22 (MHC), 11p15.5 (*INS*), 12q24.12 (*SH2B3*), and 12q13.2 (*IKZF4*). We initially identified a significant association with a missense variant in *PRKRA* (rs61999302). This variant has been reported in population genome sequencing projects and previously described as protective against autoimmune disease. Our analyses exhibited strong effects in opposing directions in African ( $OR = 3.4$ ,  $p = 1 \times 10^{-17}$ ) and European ( $OR = 0.4$ ,  $p = 3 \times 10^{-21}$ ) ancestry individuals. The minor allele of rs61999302 results in a local sequence with perfect homology to *HLA-DRB1\*07:01* within MHC haplotype A29-B44-DR7. We and others have shown *HLA-DRB1\*07:01-DQA1\*03:01-DQB1\*02:01* is an African-specific T1D risk haplotype ( $OR = 4.4$ ,  $p = 7 \times 10^{-18}$ ). Thus, association between rs61999302 and autoimmune disease may reflect the known association with HLA class II genes rather than *PRKRA*. Full sequencing of the *PRKRA* region will be required to definitively determine the source of associations with rs61999302.

**Discussion:** In this robust, family-based GWAS of T1D, we replicate known T1D loci and highlight sequence homology as a source of spurious association. This source can be difficult to detect because genotype cluster plots appear normal and alleles segregate appropriately (for rs61999302, all 228 informative trios were consistent with Mendelian inheritance). Caution should be taken when interpreting association results in regions with high sequence homology.

## PgmNr 1912: Transcriptomic and functional enrichment analysis in patients with type 2 diabetes mellitus, dyslipidemia, and chronic periodontitis.

### Authors:

R.M. Scarel-Caminaga <sup>1</sup>; S.C.T. Corbi <sup>1,2</sup>; A.S. Bastos <sup>2</sup>; D.G. Bussaneli <sup>1</sup>; B.R. Silva <sup>1</sup>; R.A. Santos <sup>3</sup>; C.S. Takahashi <sup>4</sup>; C.S. Rocha <sup>5</sup>; B.S. Carvalho <sup>6</sup>; J.F. Vasconcellos <sup>7</sup>; C.V. Maurer-Morelli <sup>5</sup>; S.R.P. Orrico <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Morphology, School of Dentistry at Araraquara, UNESP - São Paulo State University, ARARAQUARA, São Paulo, Brazil;; 2) Department of Diagnosis and Surgery, School of Dentistry at Araraquara, UNESP- São Paulo State University, Araraquara, 14801385, SP, Brazil; 3) Postgraduate Program in Sciences of the University of Franca, Franca, 14404600, SP, Brazil.; 4) Department of Genetics, Faculty of Medicine of Ribeirão Preto, USP - University of São Paulo, Ribeirão Preto, 14049900, SP, Brazil; 5) Department of Medical Genetics, University of Campinas - UNICAMP, Campinas, 13083-887, SP, Brazil; 6) Department of Statistics, Institute of Mathematics, Statistics and Scientific Computing, University of Campinas, 13083-859, São Paulo, Brazil; 7) Department of Surgery, Uniformed Services University of the Health Sciences and Henry Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD, USA.

---

Type 2 diabetes mellitus (T2DM) is a major public health problem that accounts for nearly 90% of all those diagnosed with DM and is often associated with obesity, insulin resistance condition and dyslipidemia, which is described by a metabolic dysfunction resulting from an increased level of lipoproteins in the blood. Chronic periodontitis (CP) is an infectious-inflammatory disease defined by the loss of connective tissue surrounding the teeth and alveolar bone loss. Although bacterial infection is a primary cause, the disease progression depends on the production of host mediators in response to bacterial and metabolic products. In consequence of this microbial challenge, the CP induces local and systemic elevations of pro-inflammatory cytokines. CP can be recognized as the sixth major complication associated with DM, and was detected to a greater extent and severity in individuals with DM compared to normal individual. T2DM, dyslipidemia and CP are intrinsically interrelated pathologies, but to date no study has identified the transcriptomic profile related to these combined diseases. The aim of this study was to characterize the gene expression profile of subjects with poorly or well-controlled T2DM also affected by dyslipidemia and CP. Patients (n=150) were divided into groups: (G1) poorly controlled T2DM with dyslipidemia and CP; (G2) well-controlled T2DM with dyslipidemia and CP; (G3) normoglycemics with dyslipidemia and CP; (G4) systemically healthies with CP; and (G5) systemically healthies without CP. Microarray normalized by the RMA was submitted to the RankProd method to identify candidates for differentially expressed genes (DEGs) considering and thresholds. Functional enrichment analysis was performed using Ingenuity Pathway Analysis - IPA. DEGs were submitted to pairwise comparisons, and selected DEGs were validated by quantitative polymerase chain reactions. Validated DEGs verified from G1 versus G5 were: *TGFB11I*, *VNN1*, *HLADRB4* and *CXCL8*; G2 versus G5: *FN1*, *BPTF* and *PDE3B*; G3 versus G5: *DAB2*, *CD47* and *HLADRB4*; and G4 versus G5: *IGHG3*, *ITGB2* and *HLADRB4*. The molecular profile of individuals simultaneously affected by T2DM, dyslipidemia and CP, demonstrated by the validated DEGs main implications with inflammatory response, immune cell trafficking, and infectious diseases.

## PgmNr 1913: Meta-analysis in 433,540 East Asians identifies 49 new loci associated with type 2 diabetes.

### Authors:

C. Spracklen <sup>1</sup>; M. Horikoshi <sup>2</sup>; Y.J. Kim <sup>3</sup>; K. Lin <sup>4</sup>; on behalf of the AGEN and DIAMANTE consortia

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, Univ North Carolina at Chapel Hill, Chapel Hill, North Carolina.; 2) Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Centre for Integrative Medical Sciences, Yokohama, 230-0045, Japan; 3) Division of Genome Research, Center for Genome Science, National Institute of Health, Cheongju-si, 28159, Korea; 4) Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK

Large-scale meta-analyses of genome-wide association studies (GWAS) have identified >240 loci associated with type 2 diabetes (T2D), primarily in analyses of individuals of European ancestry. With differences in linkage disequilibrium structure and allele frequencies across ancestry groups, large genetic studies in non-European samples may reveal additional loci shared across ancestry groups or ancestry-specific. We conducted the largest East Asian GWA meta-analyses to date in up to 433,540 individuals from across East Asia using genotypes imputed to 1000G Phase 3. Models were adjusted for age, sex, and other study-specific covariates, assuming an additive genetic model. Study-specific association summary statistics were combined using a fixed-effects inverse variance weighted meta-analysis. We identified 178 loci associated with T2D ( $P < 5 \times 10^{-8}$ ), including 129 known and 49 novel loci. While common variants at loci associated with T2D in both East Asians and Europeans exhibited strongly correlated effect sizes ( $r = 0.88$ ), we further identified associations specific to East Asians. Sex-stratified analyses identified one additional male-specific and two additional female-specific loci. *ALDH2* exhibited the strongest difference between males and females ( $P_{\text{het}} = 2.6 \times 10^{-19}$ ) with compelling evidence of association in males ( $P = 5.5 \times 10^{-27}$ ) and no evidence for association in females ( $P = 0.19$ ). At 88 East Asian-identified loci where no European signal has been reported within 500 kb of the East Asian lead variant, candidate genes were identified at 64 loci using colocalized eQTLs, coding variant annotation, PheWAS, and a literature search. Candidate genes include genes that encode key transcription factors in islets (e.g. *SIX3-SIX2*, *NKX6-1*, and *FOXA2*) and other tissue types (e.g. *ZBTB20*, *RANBP3L*, *ZNF703*, and *PPARA*), two miRNA clusters known to play a role in beta-cell function (*MIR17HG* and *DLK/MEG3*/miRNA cluster), and genes responsible for adipocyte and/or skeletal muscle differentiation and metabolism (e.g. *ADRB3*, *CALCR*, *DMRT2*, *JMJD1C*, *NFATC4*, and *GDAP1*). Overall, annotation of loci identified in East Asians suggests a strong role for insulin resistance in T2D pathogenesis through skeletal muscle, adipose, and liver development and function. Taken together, we identified novel T2D loci that provide a foundation for future biological research in T2D pathogenesis and offer potential targets for mechanisms for interventions in disease risk.

## PgmNr 1914: The utility of combining *CREBRF* genotype with BMI to screen Samoans for type 2 diabetes.

### Authors:

J. Tiner<sup>1</sup>; N.L. Hawley<sup>2</sup>; D.E. Weeks<sup>1</sup>; R.L. Minster<sup>1</sup>; G. Sun<sup>3</sup>; H. Cheng<sup>3</sup>; R. Deka<sup>3</sup>; T. Naseri<sup>4</sup>; S.M. Reupena<sup>5</sup>; A. Pomer<sup>2</sup>; S.T. McGarvey<sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America; 2) Yale University, New Haven, Connecticut, United States of America; 3) University of Cincinnati, Cincinnati, Ohio, United States of America; 4) Ministry of Health, Government of Samoa, Apia, Samoa; 5) Lutia i Puava ae Mapu i Fagalele, Apia, Samoa; 6) Brown University, Providence, Rhode Island, United States of America

---

Type 2 diabetes is a major global health concern. This particularly affects Samoans, who have high rates of obesity (66% of women and 45% of men) and, as a result, face high rates of obesity-related morbidities including diabetes. In Samoa in 2013, 23% of women and 21% of men were diabetic. A minor allele of common genetic variant in Samoans, rs373863828 (*CREBRF*:p.R457Q) has been associated with a 1.6-fold lower risk of diabetes. The ADA recommends testing for diabetes in adults when BMI is  $>25 \text{ kg/m}^2$ . Such a threshold might be inappropriate for Samoans, and additionally, more precise thresholds could be determined with the addition of genetic information. Here we determine how the addition of *CREBRF* genotype affects the usefulness of BMI in predicting diabetes status in Samoans. In 2010, as part of the Soifua Manuia Study, we collected BMI, fasting glucose, diabetes diagnosis, age, and *CREBRF* genotype on 2,978 Samoan adults, 25–65 years old. We used receiver operating characteristic (ROC) curves analysis to assess the utility of BMI, via the area under the curve (AUC), in predicting diabetes status. These analyses were subsequently compared to models that included genotype. All models were conducted separately in men and women and included age as a covariate. We also calculated threshold values and specificity at 80% sensitivity, in men and women, both overall and by genotype group. The AUC for diabetes classification was 0.74 in women and 0.75 in men. Including genotype improved the AUCs to 0.75 in women and to 0.77 in men. However, these differences were not found to be statistically significant ( $p=0.61$  and  $p=0.52$ , respectively). In women, the 80% sensitivity thresholds for BMI were  $30.3 \text{ kg/m}^2$  overall (with 27% specificity),  $30.3 \text{ kg/m}^2$  for GG (30%),  $31.4 \text{ kg/m}^2$  for GA (30%), and  $29.7 \text{ kg/m}^2$  for AA (17%). In men, they were  $28.7 \text{ kg/m}^2$  overall (with 40% specificity),  $27.8 \text{ kg/m}^2$  for GG (39%),  $30.6 \text{ kg/m}^2$  for GA (48%), and  $32.2 \text{ kg/m}^2$  for AA (57%). These analyses show that higher Samoan-specific screen thresholds are merited, and that the addition of the *CREBRF* genotype does not usefully alter the ability of BMI to predict type 2 diabetes status as assessed by examining the change in the AUC. However, 80% sensitivity threshold values by genotype suggest that including *CREBRF* genotype may be beneficial over BMI alone, by improving specificity without loss of sensitivity. This benefit needs to be weighed nonetheless against the costs of genotyping for this variant.

## PgmNr 1915: Polygenic risk of obesity in Chilean children: The Santiago Longitudinal Study (SLS).

### Authors:

A.E. Justice <sup>1</sup>; G. Chittoor <sup>1</sup>; M. Graff <sup>2</sup>; M. Preuss <sup>3</sup>; S. Gahagan <sup>4</sup>; E. Blanco <sup>4</sup>; C. Abala <sup>5</sup>; J.L. Santos <sup>6</sup>; B. Angel <sup>5</sup>; B. Lozoff <sup>7</sup>; R.J.F. Loos <sup>4</sup>; V.S. Voruganti <sup>8</sup>; K.E. North <sup>2</sup>; SLS Study and the GIANT BMI Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical and Translational Informatics, Geisinger, Danville, Pennsylvania.; 2) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 3) The Charles Bronfman Institute for Personalized Medicine, Mount Sinai School of Medicine, New York, NY, USA; 4) Division of Academic General Pediatrics, Child Development and Community Health at the Center for Community Health, University of California at San Diego, San Diego, CA, USA; 5) Department of Public Health Nutrition, Institute of Nutrition and Food Technology (INTA), University of Chile, Santiago, Chile; 6) Department of Nutrition, Diabetes and Metabolism, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; 7) Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI, USA; 8) Department of Nutrition and Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis NC 28081, USA

---

GWAS of body mass index (BMI) have identified hundreds of variants, although most findings have been revealed in European adults (EUR), with few GWAS in pediatric populations. The relevance of BMI-associated variants identified in adults has been assessed for childhood BMI using polygenic risk score (PRS), usually limited to tag SNPs reaching genome-wide significance. We aim to ascertain the relevance of PRS derived from >2M variants using EUR adult (PRS\_adult) GWAS of BMI and assess the association with BMI<sub>z</sub> scores (age/sex standardized BMI Z-scores using CDC panel) measured in Chilean children of the Santiago Longitudinal Study (SLS). BMI<sub>z</sub> was calculated for BMI at 5y, 10y, 16y (N=577, 770, 543, respectively). SNPs for inclusion in the PRS\_adult were selected from publicly available BMI GWAS data in more than 322K EUR adults and trained on a sample of 100K UKBiobank participants using PRSice. the best fit PRS\_adult included 583,374 variants and was subsequently used to calculate PRS\_adult in SLS, adjusted for principal components. The PRS\_adult was significantly associated with BMI<sub>z</sub> at 5y (P=2e-7), 10y (P=1e-11), and 16y (P=7e-13), and with a notable increase in variance explained variance (%V) across ages (R<sup>2</sup>=4.6%, 5.8%, 9.1% in 5y, 10y, 16y). Mean BMI<sub>z</sub> was statistically different (P<0.01) between the lowest (Q1) and upper quintiles (Q5) of the PRS across age groups. The greatest difference could be seen BMI<sub>z</sub> measured at 16y for which the mean BMI<sub>z</sub> of those in Q5 of the PRS\_adult was significantly different from all other quintiles. Further, the proportion of obese participants (BMI<sub>z</sub>≥95th%) in Q5 was significant at 16y (P=4.01E-6, 31%-Q5, 5.5%-Q1). This new derived PRS based on adults provides a notable increase in %V for BMI<sub>z</sub> in SLS over a PRS derived from 43 tag variants from a recent GWAS of BMI<sub>z</sub> in 35K EUR children (PRS\_child) (R<sup>2</sup>=1%, 1.9%, 1.3% at 5y, 10y, 16y). Notably, the performance of the PRS\_child peaked at age 10, while PRS\_adult increased from 5-16y. Our study adds to the recent body of literature that demonstrates the importance of PRS derived from GWAS for explaining variation in obesity-related traits and the generalizability across age groups. However, the moderate gain in %V and the difference in performance across ages between PRS\_adult and PRS-child highlight the importance of

selecting age-appropriate SNPs for inclusion in PRS. As such, we will address this and the potential benefit of deriving a PRS from multiethnic GWAS of BMI.

## PgmNr 1916: Tissue-specific transcriptome informed rare variant association analysis in type 2 diabetes.

### Authors:

A. Ndungu <sup>1</sup>; A. Mahajan <sup>1</sup>; J. Flannick <sup>3,4,5</sup>; A.L. Gloyn <sup>1,2</sup>; M.I. McCarthy <sup>1,2</sup>; AMP-T2D-GENES

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology, and Metabolism, University of Oxford, United Kingdom; 3) Boston Children's Hospital; 4) Broad Institute of MIT and Harvard; 5) Harvard Medical School

---

In the analysis of rare variants in sequencing studies, current gene-based association methods are highly sensitive to the variants selected for aggregation. Enrichment for true functional variants is crucial for success. Due to ubiquitous alternative splicing of genes, most coding variants typically have multiple annotated consequences across alternate gene transcripts. Common practice, which assigns the most severe consequence across any transcript, is sub-optimal when the transcript concerned is rare, tissue-specific, or not expressed in disease relevant tissues. We introduce an integrated approach that assimilates tissue-specific transcriptome profiles to prioritize rare variants for gene burden analysis in type 2 diabetes (T2D).

Recently, we reported analysis of exome sequence data from 20,791 T2D cases and 24,440 controls across five ancestries. This analysis used a novel analytical approach, which assigns prior weights to variants based on their estimated probability of haploinsufficiency. Here, we expand this framework by incorporating tissue-level transcriptomes from two T2D relevant tissues. Using PacBio long-read sequencing, we profiled expression and abundance of 31,220 isoforms in 16,520 genes in islets and 34,718 isoforms in 16,199 genes in adipose. We explored whether transcriptome informed variant selection and weighting would improve power to detect rare variants underlying T2D.

We found no additional signals exceeding exome-wide significance (beyond the three reported previously). However, of the 1,232 genes with nominal significance ( $p \leq 0.05$ ) in the initial weighted analysis, 421 and 398 genes in islets and adipose showed stronger association in the isoform aware analysis.

Amongst the top 50 genes from the initial analyses, at *RXRG*, we found a stronger rare variant association with T2D using adipose-specific annotations ( $p=5.5 \times 10^{-4}$  to  $2.7 \times 10^{-5}$ ). *RXRG* had two isoforms in both tissues, but the dominant islet isoform (95%) was rare in adipose (6%) and differed in two terminal exons. The *TMEM216* gene had improved significance in adipose ( $p=1.0 \times 10^{-4}$  to  $9.4 \times 10^{-5}$ ) after excluding variants benign in the adipose isoform.

These findings demonstrate that, alongside continuing efforts to increase available sequence data for complex traits, there is a need to improve tissue-specific variant annotation and prioritization. This is key to improving the power of rare variant association methods and deriving tissue specific insights into disease pathology.



## PgmNr 1917: Insights into the correlation between gluten intolerance and metabolic disorders: A case study.

### Authors:

P. Kaur; R. Kaur

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department Of Human Genetics, Punjabi University, Patiala, Punjab, India

---

Gluten intolerance is increasing at an alarming rate all over the world and becoming major area of concern. Mostly, it is observed in the females who have family history of various metabolic disorders such as diabetes, cancer, hypertension, abnormal cholesterol, obesity, etc. We have a case study elaborating the correlation between gluten intolerance and metabolic disorders. In F-I generation of this family, 1 (paternal grandfather) and 2 (paternal grandmother) are dead, reason not known by the proband; 3 (maternal grandfather) died by stroke; 4 (maternal grandmother) was diabetic, hypertensive, had abnormal cholesterol, suffered from celiac disease in later life and died. In F-II generation, 1 (paternal aunt) and 4 (paternal aunt) were normal; 2 (proband's father) died from prostate cancer and was hypertensive as well; 3 (paternal uncle) is suffering from diabetes and cardiovascular disease; 5 (paternal aunt) from asthma; 6 (maternal uncle) died when he was six months old due to thalassemia major; 7 (proband's mother) is suffering from diabetes, breast cancer and hypertension; 8 (maternal uncle) is hypertensive and dealing with stroke. F-III generation comprises three females, 1 (proband's sister) is suffering from polycystic ovary syndrome after giving birth to two kids; 2 (proband's sister) is unmarried, suffering from gluten intolerance and polycystic ovary syndrome and 3 (proband) is unmarried and suffering from gluten intolerance but have regular menstrual cycles. This pedigree shows a crosslink, between gluten intolerance and various metabolic disorders, which remains to be identified. Efforts are being made to identify the link using bioinformatics approach.



## PgmNr 1918: The genetics of circulating resistin in African ancestry populations.

### Authors:

K.A.C. Meeks; A.P. Doumatey; A.R. Bentley; M.H. Gouveia; G. Chen; J. Zhou; A. Adeyemo; C.N. Rotimi; The AADM Investigators

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.

---

Most genome-wide association studies (GWAS) for circulating proteins do not find genome-wide significant associations for variants within the coding gene. Resistin is one of few proteins for which GWAS consistently report variants within and near the coding gene (*RETN*). The present work sought to provide a nuanced perspective of the genetic regulation of circulating levels of resistin. We performed GWAS and fine-mapping analyses for resistin in 5621 African ancestry individuals, including 3754 continental Africans (AF) from the Africa America Diabetes Mellitus (AADM) Study and 1867 African Americans (AA) from the Howard University Family Study (HUFS). All genetic variants associated with resistin at genome-wide significance ( $p < 5 \times 10^{-8}$ ) in both the AF and AA samples, were within a 130kb region of the *RETN* gene suggesting *cis*-acting effects. The top variant (rs3219175, 75bp upstream of *RETN*) was the same in both samples and had a p-value of  $5.0 \times 10^{-111}$  in AF and  $9.5 \times 10^{-38}$  in AA, explaining 12.1% and 8.5% of variance in circulating resistin, respectively. Previous resistin GWAS reported the same single locus or only a weak secondary locus. In contrast, GWAS for insulin resulted in 13 genome-wide significant loci in AF and 1 in AA, none of which are near the gene coding for insulin (*INS*), consistent with findings of previous insulin GWAS. Pathway analyses mapped *RETN* exclusively to immune pathways and co-expression of *RETN* was primarily observed in immune cells. Bayesian fine-mapping using a stepwise conditional search around the lead SNP, revealed a cluster of 12 causal variants; 9 upstream of *RETN* and 3 downstream. Four of the causal variants are eQTLs for *RETN* expression in GTEx. Regulatory effects were found for 9 of the causal variants with predominant effects in immune cells, including promoter and enhancer effects in T-cells, enhancer effects and histone modifications in B-cells, and active enhancer and transcription start site flanks in neutrophils. Whole Exome Sequencing (WES) on a subset of 680 AA individuals detected only 1 genome-wide significant variant (rs4134852,  $p = 1.02 \times 10^{-12}$ ) upstream of *RETN*, suggesting that coding variants are unlikely to account for the GWAS observation. Our GWAS, fine-mapping, transcriptomic and functional annotation findings indicate that circulating resistin levels are determined primarily by noncoding variants upstream of *RETN*.

## PgmNr 1919: Validation and discovery of BMI-related variants in a Northern Nevadan cohort of the Healthy Nevada Project.

### Authors:

R.W. Read <sup>1,2</sup>; K.A. Schlauch <sup>1,2</sup>; E.T. Cirulli <sup>3</sup>; N.L. Washington <sup>3</sup>; S. White <sup>3</sup>; J. Lu <sup>3</sup>; A. Slonim <sup>1,2</sup>; J.J. Grzymalski <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Health Innovation, Desert Research Institute, Reno, NV, USA; 2) Institute for Health Innovation, Renown Health, Reno, NV, USA; 3) Helix, San Carlos, CA, USA

---

The aggregation of Electronic Health Records (EHRs) and personalized genetics has led to novel discoveries relevant to population health. Here we use the Illumina Human OmniExpress-24 BeadChip and Helix Exome+ sequencing in combination with EHRs from Renown Health in Reno, NV to identify variants associated with BMI and severe class II obesity. Three genome-wide association studies (GWASs) of 500,000 variants on the Illumina platform of 6,645 Healthy Nevada Project participants identified published and novel variants that associated with BMI and obesity. After standard quality control measures, with age and gender as covariates, the first GWAS identified 20 variants associated with BMI ( $p < 1 \times 10^{-5}$ ). When adding Type II Diabetes (DM2) as a comorbidity in the second GWAS, 27 significant variants were identified, with several in common with the BMI-only GWAS. We found 12 variants within the *FTO* gene to be significantly associated with the quantitative BMI trait, an association which has been illustrated in many prior studies. This association was further validated by the Helix Exome+ sequencing of 20,000 additional Healthy Nevada Project (HNP) participants using the same statistical model, in which seven top associations (mean p-value  $7 \times 10^{-13}$ ) are variants in *FTO*. In the original GWAS cohort, *TDH* showed strong associations with BMI (five variants), likely through its secondary association to DM2. *TDH* is a pseudogene, which may be transcribed to an siRNA that regulates gene expression, and is not yet known to have a link to BMI. A third complementary case-control GWAS was performed to examine links with extreme obesity: this identified variants already linked to BMI, with more variants identified in *NEGR1* than in the quantitative BMI GWAS. *NEGR1* likely plays a role in body weight regulation, as identified by knockout mice in prior studies. A novel variant in *PFKFB3*, a gene not yet directly linked to BMI, was strongly associated with obesity in this case-control study and may be protective of developing obesity. Two phenome-wide association studies (PheWASs) examined phenotypic links to BMI and clinical associations to our obesity-associated variants. The first PheWAS identified strong associations between BMI and obesity to DM2, hyperlipidemia, hypertension, sleep apnea, GERD, and asthma. The second showed associations between SNPs in *TDH*, *NEGR1*, and *FAM167A-AS1* to DM2, and SNPs in *CABP5* to IBS. These studies highlight prior and novel links of BMI and obesity.

## PgmNr 1920: Large scale meta-analysis of genome-wide association studies for physical activity highlights shared genetic factors with education and obesity.

### Authors:

Z. Wang <sup>1</sup>; M.C. Cornelis <sup>2</sup>; S. Brage <sup>3</sup>; U. Ekelund <sup>4</sup>; N. Wareham <sup>3</sup>; R.J.F. Loos <sup>1,5</sup>; M. den Hoed <sup>6</sup>; Genome Wide Association of Physical Activity (GWAPA) Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York.; 2) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; 3) MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge, UK; 4) Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway; 5) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; 6) The Beijer Laboratory and Department of Immunology, Genetics and Pathology, Uppsala University and SciLifeLab, Uppsala, Sweden

Low levels of physical activity (PA) and a sedentary lifestyle (SL) are well-established, modifiable risk factors for human health. Despite being moderately heritable ( $h^2 \sim 40\%$ ), our understanding of genetic factors influencing PA and SL are limited. Therefore, we performed the largest meta-analysis of genome-wide association studies (GWAS) to date, including up to 641,692 individuals of European (94%), African (2%), East Asian (1%), South Asian (1%), and Hispanic (2%) ancestries. We studied self-reported moderate-to-vigorously intensity leisure time PA (MVPA); leisure screen time (LST); sedentary behavior at work and commuting. Sex- and ancestry-specific associations were adjusted for age and study-specific covariates. Summary statistics were meta-analyzed using a fixed-effect, inverse-variance weighted approach.

We identified 123 PA and SL associated loci ( $p < 5 \times 10^{-9}$ ); 95 in European ancestry individuals and 28 additional loci in all ancestries combined. In total, 98 loci had not been identified for PA/SL traits before, 87 of which were previously associated with educational attainment (EA,  $n=81$ ) and/or obesity traits ( $n=77$ ). Genetic correlations between MVPA and LST were modest-to-high ( $r=-0.49$ ), as were their correlations with EA ( $r$  [0.43, 0.65]) and obesity traits ( $r$  [-0.41, -0.23]). We also observed genetic correlations with mental health, cancer and aging ( $r$  [-0.48, 0.45]). Using uni- and multivariable Mendelian Randomization (MR), we found that a 1SD genetically higher LST caused a 0.27SD higher BMI ( $p=3.09 \times 10^{-9}$ ). Conversely, a 1SD genetically higher BMI only caused a 0.09SD higher LST ( $p=7.26 \times 10^{-10}$ ). Multivariable MR showed that the bidirectional causal effects between LST and BMI were independent of EA. However, higher LST caused lower EA through an effect on BMI (univariable  $b=-0.48$ ,  $p=3.76 \times 10^{-30}$ ; multivariable  $b=-0.07$ ,  $p=0.24$ ), while higher EA caused lower LST independently of BMI (univariable  $b=-0.29$ ,  $p=5.55 \times 10^{-17}$ ; multivariable  $b=-0.31$ ,  $p=4.52 \times 10^{-17}$ ). Tissue and pathway analyses showed enrichment of gene expression in the central nervous system, especially in reward-related areas, such as basal ganglia, cerebellum, hippocampus and the limbic system. In addition, variants of known relevance for muscle performance in *ACTN3* showed suggestive associations with MVPA and LST ( $p=4.25 \times 10^{-8}$ ).

Genes in loci identified in our study point to the brain as a key player in PA regulation. Furthermore, we confirm that lower LST has a causal effect on higher BMI.

## PgmNr 1921: Genome-wide association study of circulating liver enzymes in the UK Biobank reveals an expanded role for manganese transporter SLC30A10 in liver health.

**Authors:**

L.D. Ward; H. Tu; M.M. Parker; G. Hinkle; P. Nioi

[View Session](#) [Add to Schedule](#)

**Affiliation:** Alnylam Pharmaceuticals, Cambridge, MA.

---

Liver diseases, including non-alcoholic fatty liver disease, cirrhosis, and hepatocellular carcinoma, represent a substantial area of unmet medical need. Susceptibility to liver disease is heritable, and genetic association studies of circulating liver enzymes have provided several potential therapeutic targets for liver disease, such as PNPLA3 and HSD17B13. Motivated by these examples, we studied genetic associations with alanine aminotransferase (ALT) and aspartate aminotransferase (AST) in 337,151 array-typed individuals and 128,382 exome-sequenced individuals from the UK Biobank. We discover 221 loci associated with ALT, 172 of which were previously unreported, and 263 loci associated with AST, 228 of which were previously unreported. The strongest previously-unreported association affecting both ALT and AST is a missense variant in *SLC30A10*, rs188273166 (Thr95Ile) which associates with higher ALT and AST levels (0.3 SD increase in ALT,  $p = 1.1e-21$ ; 0.4 SD increase in AST,  $p = 1.3e-28$ ). Carriers of this variant also have a higher risk of being diagnosed with ICD10 codes beginning with K75 (other inflammatory liver diseases; OR = 1.1;  $p = 0.03$ ), and have higher liver inflammation measured by hepatic MRI (1.6 SD increase in LIF,  $p = 0.03$ ). This SNP is found in approximately 1 in 500 White British individuals. SLC30A10 is a manganese efflux transporter, and homozygous loss-of-function causes a Mendelian syndrome of hypermanganesemia with dystonia, polycythemia, and hepatic dysfunction. Our results demonstrate for the first time that heterozygous missense mutation of *SLC30A10* is sufficient to cause increased susceptibility to liver damage.

## PgmNr 1922: Oceanian reference panels are crucial for understanding the impact of a Polynesian-specific variant on adiposity in Native Hawaiians.

### Authors:

M. Lin <sup>1</sup>; C. Caberto <sup>2</sup>; P. Wan <sup>1</sup>; A. Lum-Jones <sup>2</sup>; M. Tiirikainen <sup>2</sup>; L. Pooler <sup>1</sup>; B. Nakamura <sup>1</sup>; X. Sheng <sup>1</sup>; L. Le Marchand <sup>2</sup>; L. Wilkens <sup>2</sup>; C. Haiman <sup>1</sup>; I. Cheng <sup>3</sup>; C. Chiang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Epidemiology, Department of Preventive Medicine, University of Southern California, Los Angeles, CA.; 2) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 3) 3Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA

---

Statistical imputation applied to genome-wide array data is the most cost-effective approach to improve the catalog of genetic variation in a cohort. However, because of ascertainment bias and a lack of representation in reference individuals, studies in minority populations incur greater inaccuracies in imputed genotypes, further contributing to the obstacles to study these populations. Here we examined the severity of such impacts by genotyping a functionally important Polynesian-specific variant, rs373863828 in *CREBRF* gene, in large number of Native Hawaiians (N=3,188) from the Multiethnic Cohort. Consistent with the initial findings in Samoans and other Oceanian populations, we found the variant is strongly and significantly associated with BMI ( $p = 7.5 \times 10^{-5}$ ). We showed that because of the current absence of Polynesian representation in publicly accessible population sequences, this variant or its proxy could not be studied through imputation. Moreover, the association signal at this Polynesian-specific variant could not be captured by alternative approaches such as admixture mapping. In contrast, highly accurate imputation can be achieved even when only a small number (<200) of internally constructed reference individuals were available, which boosts power for association testing (improved  $p = 1.5 \times 10^{-7}$  for BMI), and revealed strong associations with other adiposity and metabolic traits (e.g.  $p = 3 \times 10^{-6}$ , and  $1.4 \times 10^{-4}$  for hip circumference and T2D, respectively, using imputed data on up to 3,936 individuals). Taken together, our results suggest the alarming reality that the lack of representation would inhibit the discovery of functionally important population-specific loci such as *CREBRF*. Yet, they could be easily detected and prioritized with improved representation of diverse populations in sequencing studies.

## PgmNr 1923: The impact of variants in three genes: *MC4R*, *FTO* and *PPARG* in overweight and obesity in a Brazilian population.

### Authors:

S. Lofgren <sup>1</sup>; G.V. Castro <sup>1,2</sup>; A.L. Andriani <sup>1,2</sup>; A.F.S. Latorre <sup>1</sup>; F.P. Korndorfer <sup>1</sup>; L.K.C Back <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biogenetika Individualized Medicine , Florianopolis, Santa Catarina, Brazil; 2) Federal University of Santa Catarina, Dept of Cell Biology, Embriology and Genetics. Florianopolis, Santa Catarina, Brazil

---

Obesity and overweight are worldwide public health problems with relevant impacts on morbidity and mortality. The genetic predisposition is evident but still poorly understood. The majority of variants implicated so far have shown to have very modest contribution to these traits and great variability has been described across populations. In this work, we analyzed the association of three genes: *PPARG* (rs1801282) *FTO* (rs9939609) and *MC4R* (rs17782313) with the presence of overweight and obesity in a large Brazilian population. The case-control study involved 3641 individuals, comprising of 1622 patients with overweight or different degrees of obesity; and 2019 individuals with normal body mass index (BMI). Genotyping was performed by quantitative PCR and allele and genotype frequencies were determined for association analysis. All genes were associated with increased BMI, obesity or both. *MC4R* rs17782313-C was associated with obesity and overweight in both men and women (OR = 1.27,  $p = 0.038$ ). Interestingly for *FTO* the rs9939609-A allele was associated in both sexes however for women it was a risk factor (OR = 1.60,  $p = 0.013$ ) and for men, on the other hand, a protective factor (OR = 0.66,  $p = 0.015$ ) for both obesity and overweight. *PPARG* was the most associated gene in our sample, associated with overweight and all grades of obesity, and interestingly this association was restricted to women (rs1801282-G OR = 1.34,  $p = 0.004$ ). The combined effect of the three risk alleles on overweight and obesity showed an OR of 1.65 ( $p = 0.008$ ) considering both sexes and the association was even more pronounced in women (OR = 1.97,  $p = 0.002$ ). Our findings indicate that the three genes are robustly associated with overweight and different degrees of obesity in the Brazilian population, reaching together a relatively high impact on these traits. For *FTO* and *PPARG* this association was more pronounced or exclusive to women, suggesting a sex-specific genetic effect of these variants.

# PgmNr 1924: An integrated approach to identify genetic and environmental mediators of insulin resistance.

## Authors:

B. Balliu <sup>1</sup>; I. Carcamo-Orive <sup>2</sup>; M. Gloudemans <sup>2</sup>; M. Durrant <sup>2</sup>; A. Rao <sup>2</sup>; D.A. Knowles <sup>3</sup>; C. Park <sup>2</sup>; J.W. Knowles <sup>2</sup>; E. Ingelsson <sup>2</sup>; S.B. Montgomery <sup>2</sup>; T. Quertermous <sup>2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) UCLA, Los Angeles, California.; 2) Stanford University School of Medicine, Stanford, California; 3) New York Genome Center, New York, New York

---

Insulin resistance (IR) is a common and increasing public health concern that precedes development of type 2 diabetes and cardiovascular disease. While GWAS studies have identified loci associated with IR related traits, environmental risk factors, causal genes, and their mechanisms remain largely unknown. We therefore developed a combined experimental and computational framework to provide mechanistic insights into GWAS loci.

We first model environmental effects on gene expression in liver, fat, and skeletal muscle, three key IR tissues. Specifically, we performed RNA-seq in HepG2 (human hepatocytes), SGBS (human adipocytes), and HMCL-7304 (human skeletal myocytes) cells treated with 21 different environmental perturbations related to IR (234 total samples). We identified thousands of treatment-specific, treatment-by-cell-type-specific, and shared differentially expressed (DE) genes underlying IR response pathways. The largest number of DE genes were observed for glucose (2,853), IGF-1 (1,676), TGFβ1 (1,644), and insulin (1,589) in HepG2 and for dexamethasone (3,706), IL-6 (3,129), TNF-α (2,232), and rosiglitazone (2,096) in SGBS cells.

We then combine these results with GWAS of IR-related traits. We show that genes differentially expressed in multiple treatments and cells were depleted from GWAS results ( $p = 1.41e-02$ ). To distinguish treatments whose association with disease are modifying risk, we compute transcriptional risk scores for each treatment in each cell-type by integrating our DE-by-treatment effects with TWAS effects generated from IR-related GWAS and GTEx eQTLs in IR-relevant tissues. We find several treatments that modify risk of IR-related traits. We further performed co-localization of IR-related trait GWAS and eQTLs from IR-related tissues to identify signals pointing to causal genes in each locus. We identified a single candidate gene for 30% of loci while in 30% of loci multiple genes showed signal. We validate the candidate genes through CRISPR-Cas9 technology in human adipocytes and skeletal muscle cells.

Overall, our results provide a broad resource of dynamics of the transcriptional landscape in IR-related tissues and demonstrate the advantages of large-scale characterization of effects of genetic variation in diversely-stimulated and pathologically-relevant cells.

B. Balliu and I. Carcamo-Orive contributed equally to this work.



## PgmNr 1925: Assessing the causal effects of childhood BMI on 139 genetically correlated adult outcomes using two-sample Mendelian randomization.

### Authors:

S.S. Dong; J.C. Feng; J.M. Ding; S. Yao; K. Zhang; Y. Guo; T.L. Yang

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Life Science and Technology, Xi'an Jiaotong University , Xi'an, Shaanxi 710049, China

---

Investigating the potentially causal effects of childhood obesity on diseases/traits in adulthood could help guide early intervention strategies. Here we aim to assess the causal effects of childhood BMI on adult traits/diseases using 2-sample Mendelian randomization (MR) with current available GWAS summary data. GWAS summary data of childhood BMI were downloaded from the Early Growth Genetics consortium (the sample numbers of the discovery and replication phase were 35,668 and 11,873, respectively). 783 summary datasets for outcomes were collected from public database. For outcomes genetically correlated with childhood BMI, MR estimates using inverse variance weighted analysis and alternate methods (MR-Egger, weighted median, and weighted mode) were performed. Heterogeneity test, MR-Egger, and MR-PRESSO were used to assess horizontal pleiotropy. 139 outcomes (GWAS data derived from at least 46,186 subjects) were genetically correlated with childhood BMI. As it might be expected, MR analyses showed that higher childhood BMI is a risk factor for general health outcomes and socioeconomic status. In addition to the known causal effects of childhood BMI on coronary artery disease, T2D, and osteoarthritis, we also found that higher childhood BMI was associated with increased risk of cholelithiasis (OR = 1.010, CI: 1.006 to 1.014,  $P = 1.62 \times 10^{-7}$ ), increased number of self-reported non-cancer illnesses ( $\beta = 0.1224$ , CI: 0.0830 to 0.1619,  $P = 1.15 \times 10^{-9}$ ), number of treatments taken ( $\beta = 0.1559$ , CI: 0.0990 to 0.2128,  $P = 7.76 \times 10^{-8}$ ) and increased Townsend deprivation index ( $\beta = 0.1201$ , CI: 0.0539 to 0.1863,  $P = 3.79 \times 10^{-4}$ ). For lifestyles, we observed a positive causal association between childhood BMI and adult smoking status ( $\beta = 0.0534$ , CI: 0.0365 to 0.0704,  $P = 6.65 \times 10^{-10}$ ), but childhood BMI was negatively correlated with adulthood alcohol drinking frequency ( $\beta = -0.1591$ , CI: -0.1970 to -0.1212,  $P = 1.90 \times 10^{-16}$ ). In summary, we generated a view of the causal effects of childhood BMI on adulthood traits. Our results suggested that early intervention of childhood BMI may offer help for the improvement of adulthood health status.



## PgmNr 1926: Body mass index and risk of dying from a bloodstream infection: A Mendelian randomization study of 56,000 subjects from the HUNT Study with 23 years follow-up.

### Authors:

T. Rogne<sup>1,2,3</sup>; E. Solligård<sup>1,3</sup>; S. Burgess<sup>4,5</sup>; B.M. Brumpton<sup>6,7,8</sup>; J. Paulsen<sup>9</sup>; H.C. Prescott<sup>10</sup>; R.M. Mohus<sup>1,3</sup>; L.T. Gustad<sup>1,11</sup>; A. Mehl<sup>11</sup>; B.O. Åsvold<sup>12,6</sup>; A.T. DeWan<sup>2,1</sup>; J.K. Damås<sup>1,13,14</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Gemini Center for Sepsis Research, Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, NTNU, Oslo, Norway; 2) Yale School of Public Health, Yale University, New Haven, CT, USA; 3) Clinic of Anaesthesia and Intensive Care, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; 4) MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; 5) Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 6) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway; 7) MRC Integrative Epidemiology Unit, University of Bristol, UK; 8) Clinic of Thoracic and Occupational Medicine, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; 9) Department of Medical Genetics, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; 10) Department of Medicine, University of Michigan and VA Center for Clinical Management Research, both in Ann Arbor, MI, USA; 11) Department of Medicine, Levanger Hospital, Nord-Trøndelag Hospital Trust, Levanger, Norway; 12) Department of Endocrinology, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; 13) Centre of Molecular Inflammation Research, Department of Clinical and Molecular Medicine, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 14) Department of Infectious Diseases, St Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

**Background:** In observational studies, higher body mass index (BMI) has been associated with increased incidence and mortality of bloodstream infection (BSI) or sepsis in the general population, but with reduced mortality among patients with BSI or sepsis. However, traditional observational studies are subject to bias and confounding. We wanted to evaluate the causal association of BMI with risk and mortality of BSI.

**Methods:** We used a population-based cohort in Norway followed from 1995 to 2017 (The HUNT Study). We carried out linear and non-linear Mendelian randomization analyses, including a wide range of sensitivity analyses to evaluate the robustness of the findings.

**Findings:** Among 55,908 participants, the mean age at enrollment was 48.3 years, 26,324 (47.1%) were men, and mean BMI was 26.3 kg/m<sup>2</sup>. During median 21 years follow-up, 2,547 (4.6%) subjects experienced a BSI and 451 (0.8%) died from BSI. Compared with a genetically-predicted BMI of 25 kg/m<sup>2</sup>, a genetically-predicted BMI of 30 kg/m<sup>2</sup> was associated with a hazard ratio (95% confidence interval) for BSI incidence of 1.78 (1.40 – 2.27) and for BSI mortality of 2.56 (1.31 – 4.99) in the general population, and for mortality of 2.34 (1.11 – 4.94) in an inverse-probability weighted analysis of patients with BSI.

**Interpretation:** Supportive of a causal relationship, genetically-predicted BMI was positively

associated with BSI incidence and mortality. Our findings contradict the “obesity paradox” where previous traditional epidemiological studies have found increasing BMI to be protective in terms of mortality for patients with BSI or sepsis.

## **PgmNr 1927: The causal role of central obesity and hyperuricemia on the risk of metabolic syndrome: Mendelian randomization analysis using 10,000 subjects.**

### **Authors:**

M.I. Biradar <sup>1</sup>; K.M. Chiang <sup>2</sup>; H.C. Yang <sup>3</sup>; Y.T. Huang <sup>3</sup>; W.H. Pan <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) National Yang-Ming University and Academia Sinica, Taipei, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 3) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

---

Hyperuricemia, the predecessor of gout, has been found to cluster with multiple components of metabolic syndrome (MetS). It is unclear whether hyperuricemia is a downstream result of MetS or may play an upstream role in MetS development.

Since central obesity is generally considered as the core of MetS; we investigated the causal relationship between uric acid and components of metabolic syndrome, using central obesity as a positive control. Taking advantage of the data from 10K participants of Taiwan Biobank, we carried out Mendelian randomization analyses with uric acid weighted genetic risk score (WGRS) and waist circumference WGRS as instrumental variables and metabolic syndrome components as the outcomes.

Using a generalized linear model, we found that uric acid WGRS was significantly associated with SBP, DBP, mean arterial pressure, triglyceride (positively for the above), and HDL-C (negatively). But it was not associated with fasting serum glucose, HbA1C, waist circumference, waist/hip ratio or BMI. On the other hand, waist circumference WGRS was associated with all the components of MetS including uric acid.

Our MR investigation confirms that central obesity is a causal agent of all MetS components including hyperuricemia and suggests that hyperuricemia does augment the risk of MetS, but not through accumulating fat or hyperglycemia.

## PgmNr 1928: Triglyceride concentration in the Japanese population: A polygenic model analysis.

### Authors:

O. Gervais<sup>1,6</sup>; Y. Kawai<sup>2,3</sup>; S.-S. Khor<sup>2,3</sup>; K. Ueno<sup>2,3</sup>; Y. Hitomi<sup>2,4</sup>; Y.-Y. Wang<sup>5</sup>; K. Tokunaga<sup>2,3</sup>; M. Nagasaki<sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Kyoto, Kyoto, Japan; 2) Department of Human Genetics, The University of Tokyo, Bunkyo-ku, Tokyo, Japan; 3) Genome Medical Science Project, National Center for Global Health and Medicine, Shinjuku-ku, Tokyo, Japan; 4) Department of Microbiology, Hoshi University School of Pharmacy and Pharmaceutical Sciences, Shinagawa-ku, Tokyo, Japan; 5) Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan; 6) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan

---

Metabolic syndrome is one of the main causes of morbidity and mortality, and has become a major public health issue over the past decades. However, although the increase in the prevalence of metabolic syndrome is a worldwide trend and differences in prevalence between populations are known, the bulk of research efforts are concentrated on populations of European descent. In this study, we estimated the genetic parameters of triglyceride concentration in over 13,000 Japanese individuals from the Miyagi and Iwate prefectures, Japan, by applying a polygenic model approach. We calculated the heritability due to common genetic variants at the regional genetic level, by scanning the genome using clusters of 50 SNPs. We identified a number of significant loci, such as *LPL*, *APOA5*, and *SIK3*, and quantified their contribution to phenotypic variation. We also analyzed the chromosome heritability and found that it was proportional to the number of significant loci and the strength of the corresponding signals; in other words, the significant regions identified were located on the chromosomes that displayed a high heritability (e.g. chromosomes 8 and 11). This result was somewhat unexpected, given that we had not found this relationship to hold in a previous study that we conducted on other traits closely related to metabolic syndrome.

## PgmNr 1929: An exome-wide association study identifies novel non-synonymous variants associated with type 2 diabetes-related metabolites in the Korean population.

### Authors:

Y. Kim; T. Park; H. Lee; B. Kim

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Genome Research, Center for Genome Science, Korea National Institute of Health, Chungcheongbuk-do 28159, Republic of Korea

---

Previously, numerous genetic variants associated with T2D-related metabolites have been reported. Despite the efforts, however, most of the discovered variants were located within non-coding regions and further efforts for identifying coding variants are warranted. In this context, an exome array platform, designed for pursuing potentially functional variants, is an efficient approach to comprehensively catalog coding variants associated with various phenotypes. In this study, we performed an association study using 27,140 non-synonymous variants included in the Illumina HumanExome BeadChip and nine T2D-related metabolites identified by a targeted metabolomics approach for 2,338 Korean individuals from the Korea Association REsource (KARE). A linear regression analysis was performed to identify novel non-synonymous variants associated with T2D-related metabolites by adjusting age, sex, and BMI. Six loci were identified ( $P < 1.8 \times 10^{-6}$ , after the Bonferroni correction for multiple testing) including *CPS1* for glycine, *COL4A3*, *DYNC2H1*, *PTPRH*, *CCDC8* for creatinine, and *CYP4F2* for PC ae C36:0. Among them, *CYP4F2* was previously reported while the remaining five loci were newly associated in this study. Taken together, these findings may highlight the role of coding genetic variants for metabolite processing and enhance the understanding of the underlying biology of T2D-related metabolites.

## PgmNr 1930: Biological insights into type 2 diabetes from genetic investigation of plasma metabolites in the METSIM study.

### Authors:

Yin <sup>1,2</sup>; E. Hector <sup>1</sup>; L. Chan <sup>1,2</sup>; A. Jackson <sup>1,2</sup>; H. Stringham <sup>1,2</sup>; J. Vangipurapu <sup>3</sup>; C. Fuchsberger <sup>4</sup>; J. Kang <sup>1</sup>; M. Boehnke <sup>1,2</sup>; M. Laakso <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI 48109, USA; 3) Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 4) Institute for Biomedicine, Eurac Research, Bolzano, Italy

---

Type 2 diabetes (T2D) affects ~10% of the global population and is a major cause of morbidity and mortality. Genetic variation contributes to T2D risk. While genome-wide association studies (GWAS) have identified >240 genetic loci for T2D, the disease mechanisms underlying these findings are mostly poorly understood. Plasma metabolites may contribute to T2D risk or reflect metabolic changes during T2D development. The investigation of the genetic basis for plasma metabolites has the potential to help us understand T2D disease mechanisms.

We profiled 1,562 plasma metabolites in 6,492 Finns in the METabolic Syndrome In Men (METSIM) study using ultra performance liquid chromatography - tandem mass spectrometry. We genotyped (n=10,066) and exome sequenced (n=9,957, ~80x) nearly all METSIM participants. More recently, we performed deep whole genome sequencing (~24x) in 3,074 METSIM participants and generated an integrated Finnish-specific genotype imputation panel. We imputed the remaining METSIM samples using this population-specific reference panel. We investigated phenotype missingness patterns and experimented with phenotype imputation using several methods including minimum-value, k-nearest neighbor, and random forest methods. In an initial GWAS of 20 amino acids, we confirmed multiple previously identified amino acids genomic regions including loci near *CPS1*, *GLDC*, *PKD1L2*, *ALDH1L1*, *GCSH*, *PRODH*, *ASPG*, *SNAI2*, *GLS2*, *SLC7A6*, and *ADAMTS3*, and implicated apparently novel loci at genome-wide significance ( $P < 5 \times 10^{-8}$ ) proximal to *KDM4C* and *WASHC3* that are relevant to amino acid biosynthesis and transport. We also identified amino acid associations with multiple T2D-associated variants, including variants near *GCKR*. We are continuing to carry out GWAS with each of the remaining plasma metabolites using the combined observed and imputed metabolic values. We plan to further explore shared genetic loci and infer pairwise causal relationships between plasma metabolites, T2D risk, and T2D-related quantitative traits. This research has the potential to discover metabolic biomarkers for T2D early intervention and personalized therapy. The findings will help guide functional experiments for T2D genetic associations and reveal T2D biological pathways.

## PgmNr 1931: Low polygenic risk attenuates the obesity-increasing effects of pathogenic mutations in *MC4R*.

### Authors:

N. Chami<sup>1</sup>; M. Preuss<sup>1</sup>; R.W. Walker<sup>1,2</sup>; A. Moscati<sup>1</sup>; R.J.F. Loos<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.; 2) Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 3) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

**Background:** Melanocortin 4 receptor (MC4R) deficiency, due to protein-altering mutations, is the most common cause of severe and early onset obesity. We examine why some carriers of pathogenic mutations that result in MC4R deficiency remain of normal weight, to gain insight into the mechanisms that control body weight.

**Methods:** We identified 69 obesity-increasing mutations (MAF<0.1%) in *MC4R* reported in the human gene mutation database (HGMD) and Clinvar that were available in ~450,000 individuals of the UK Biobank and calculated their penetrance and effect on obesity (BMI  $\geq 30$  kg/m<sup>2</sup>). We then focused on the most penetrant mutations with the largest effect on obesity and examined differences between normal weight and obese carriers. We calculated a genome-wide polygenic risk score for BMI (PRS<sub>BMI</sub>) to assess the polygenic contribution to body weight in normal weight and obese carriers.

**Results:** Of the 69 mutations, only 11 mutations had moderate to high penetrance ( $\geq 30\%$ ) and increased the odds of obesity by more than twofold. Twenty-nine of the 183 carriers of these 11 mutations were of normal weight. Body composition of normal weight carriers was similar to non-carriers, whereas obese carriers had a somewhat higher BMI (P=0.034) than obese non-carriers, due to greater lean mass (P=0.002). Normal weight carriers more often reported that, already at age 10y, they were thinner/average (72%) compared to obese carriers (48%) (P=0.02). The PRS<sub>BMI</sub> of normal weight carriers (PRS<sub>BMI</sub>= $-0.66 \pm 0.18$ ) was significantly lower than of obese carriers ( $0.40 \pm 0.12$ ; P= $8 \times 10^{-6}$ ), and also lower than that of normal weight non-carriers ( $-0.28 \pm 0.003$ ; P=0.04). Further, among *MC4R* mutation carriers, those with a low PRS<sub>BMI</sub> (bottom quartile) have ~5 kg/m<sup>2</sup> lower BMI (~14 kg of body weight for a 1.7m-tall person) than those with a high PRS<sub>BMI</sub> (top quartile).

**Conclusion:** Only 11 (16%) out of 69 previously reported *MC4R* mutations had high penetrance and increased risk of obesity, highlighting the importance of large-scale data to validate the impact of mutations observed in small-scale and case-focused studies. Furthermore, our results show that despite the key role of *MC4R* in obesity, the obesity-increasing effects of pathogenic *MC4R* mutations can be countered, at least in part, by a low polygenic risk, potentially representing other innate mechanisms implicated in body weight regulation.

## PgmNr 1932: Whole exome sequencing to identify genetic risk variants underlying obesity in pediatric patients.

### Authors:

J.B. Cordero <sup>1</sup>; A.H. Shoemaker <sup>2</sup>; C.L. Simpson <sup>1</sup>; J.C. Han <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, TN; 2) Division of Pediatric Endocrinology and Diabetes, Vanderbilt University Medical Center, Nashville, TN; 3) Departments of Pediatrics and Physiology, University of Tennessee Health Science Center, Memphis, TN; 4) Le Bonheur Children's Foundation Research Institute, Memphis, TN

---

### Introduction:

Childhood obesity is a major health problem in many countries. Its repercussions on health in both childhood and later adult life are well-known. Both genetic and environmental factors play a role in the development of obesity. Unraveling the relevant genes remains important since there is abundant evidence indicating that an individual's genetic background is crucial in determining obesity risk. In this study, we performed whole exome sequencing (WES) in 18 children with obesity to identify genetic variants that may contribute to childhood obesity.

### Methodology:

WES analysis was carried out in 11 males and 7 females [age (mean  $\pm$  SD): 12.8  $\pm$  3.3 years, race/ethnicity: Caucasian (67%), African American descent (28%), Hispanic (5%), body mass index (BMI): 39.7  $\pm$  11.0 kg/m<sup>2</sup>, 157  $\pm$  31 % of 95<sup>th</sup> percentile]. Variant annotation was performed using wANNOVAR. Genes related to obesity were identified from GWAS catalog, OMIM, GeneCards, Genetics Home Reference, NCBI Gene, CDC and PubMed. These obesity genes were then cross-referenced with the annotated WES data. Only the functional variants were selected for further analysis. MutationTaster, M-CAP and CADD were used to predict deleteriousness of amino acid changes. Variants predicted to be deleterious using the three algorithms, designated as pathogenic in ClinVar or generate a premature terminator codon were then analyzed individually for association with obesity.

### Results:

After analysis of WES variant data, we identified 14 DNA variants in the genes *KSR2*, *GIPR*, *POMC*, *PCSK1*, *PPARGC1B*, *SDC3*, *VPS13B*, *UCP2*, *UCP3*, *SDCCAG8* and *CEP290* predicted to be deleterious by several in silico prediction tools. We also found two variants within *BBS2* and *UCP3* that were designated as pathogenic in ClinVar. Of interest is a nonsense mutation that lead to a premature stop codon identified in *VPS13B* gene. This gene is implicated in the development of adipocytes, storage and distribution of fats in the body. Further investigation of this variant exploring its role in obesity may be warranted.



## Conclusion:

Whole exome sequencing and application of *in silico* tools significantly enhanced the discovery of genetic variants. Identification of these obesity-related genes can increase our knowledge about the mechanism of obesity development. It can also help in formulating strategies for the early prediction, prevention and management of obesity.

## PgmNr 1933: Investigating gene x environment interaction in BMI: Sex specific and age specific effects.

### Authors:

A. Dawes

[View Session](#) [Add to Schedule](#)

**Affiliation:** Molecular genetics team, LEVEL 3 RILD RD&E , University of Exeter, Exeter, Devon, England

Obesity is a global epidemic, with 26% UK adults obese in 2016. Body mass index (BMI) is used to classify overweight (BMI >25) and obese (BMI>30) patients. Genome wide association studies (GWAS) have discovered over 250 common variants associated with BMI, thought to increase BMI through a polygenic effect. Whilst BMI is highly heritable, heritability being approximately 70%, there is a considerable amount of variation in this heritability, of which environmental and lifestyle factors are thought to contribute towards. Few studies focus on the interaction of genetic and non-genetic factors, likely due to the lack of power in studies to previously detect small effect sizes.

This project aims to identify environmental and lifestyle factors that may influence the genetic associations with BMI. This project uses genotype and phenotype data from population-based studies, including the UK Biobank and GIANT consortium. We studied the following genetic and lifestyle interactions on BMI: age, sex, socio-economic status, smoking and sunscreen use as a negative control. Seventy-six single nucleotide polymorphisms (SNPs) previously associated with BMI were selected, and a 76-SNP genetic risk score (GRS) was generated. Linear regression interaction analysis using the GRS was used to identify the interaction effects.

The GRS sex Interaction effect compares women against men (0.0028, 95%CI: 0.0016, ?0.0039), while the GRS age interaction effect compares older to younger age groups (0.0034, 95%CI: 0.0023, ?0.0046). There is also evidence of interaction of socio-economic status (0.0021, 95%CI: 0.0009, 0.0031), and smokers (0.0048, 95%CI: 0.0028, 0.0052), but not for the negative control sunscreen use (-0.0023, 95%CI: -0.0038, -0.0007). Results for individual seventy-six SNPs follow similar trends as for the GRS results. The strongest sex interaction effects when compared to our negative control are seen in SEC16B (0.0272795, 95%CI: 0.0163921, 0.0381669) and NT5C2 (0.0213557, 95%CI: 0.0048956, 0.0378158).

In summary, there are sex specific and age specific genetic effects on BMI. Additional work to identify the mechanism of the sex specific and age specific interaction effects on BMI by these two variants is being undertaken. This includes a sex specific and age specific phenome wide association analysis investigation of these two genes against phenotypic traits associated with BMI.

## PgmNr 1934: Clinical utility of polygenic risk prediction for obesity across age, sex, and ancestrally diverse population subgroups.

### Authors:

M. Graff<sup>1</sup>; M. Preuss<sup>2</sup>; S. Sakaue<sup>3,4</sup>; I. Ntalla<sup>5</sup>; S.I. Berndt<sup>6</sup>; S. Vedantam<sup>7,8</sup>; A. Locke<sup>9</sup>; E. Marouli<sup>5</sup>; L. Yengo<sup>10</sup>; A.E. Justice<sup>11</sup>; K.L. Young<sup>1</sup>; Y. Lu<sup>12</sup>; M. Kanai<sup>3,13</sup>; T.M. Frayling<sup>14</sup>; Y. Okada<sup>3</sup>; P. Deloukas<sup>5</sup>; K.E. North<sup>1</sup>; J.N. Hirschhorn<sup>7,8</sup>; R.J.F. Loos<sup>2</sup>; on behalf of the GIANT Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, Univ North Carolina, Chapel Hill, NC, USA; 2) The Charles Bronfman Institute for Personalized Medicine at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 3) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka, Japan; 4) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) William Harvey Research Institute, Barts & the London School of Medicine and Dentistry, Queen Mary University of London, London EC1M 6BQ, UK; 6) National Cancer Institute, Division of Cancer Epidemiology & Genetics, Bethesda, MD 20892, USA; 7) Broad Institute of MIT and Harvard, Cambridge, MA, USA; Department of Genetics, Harvard Medical School, Boston, MA, USA; 8) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA; 9) McDonnell Genome Institute and Department of Medicine, Division of Genomics and Bioinformatics, Washington University School of Medicine, St. Louis, MO, USA; 10) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia; 11) Weis Center for Research, Geisinger Health System, Danville, PA, USA; 12) Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA; 13) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA, USA.; 14) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK

Obesity results from gradual accumulation of fat over the life course. Because of its early onset, prevention is most effective early in life. Genome-wide association (GWA)-derived polygenic scores (PGS) provide a new tool to identify high-risk individuals that may benefit from additional preventive measures.

Here, we examine the ability of a PGS to predict adult obesity using the latest GWAS meta-analyses data from GIANT and test its performance across diverse sub-populations. Using PRSice, we built a PGS based on summary statistics (2.1M SNPs) of a BMI-GWAS meta-analysis in >322K individuals of European ancestry. The PGS that explained the most BMI variance (7.1%, LD>0.80, P<0.5) in a subset of 10K unrelated European-ancestry individuals of the UK Biobank (UKBB) was used for analyses in unrelated individuals of European (UKBB<sub>EUR</sub> N~405,646), African (UKBB<sub>AFR</sub> N~8,429), Hispanic (HCHS/SOL N~10,122) and East Asian (Biobank Japan (BBJ) N~154,871) ancestry. We tested association of the PGS with BMI and obesity classes I (BMI≥30kg/m<sup>2</sup>; [≥25kg/m<sup>2</sup> BBJ]), II (BMI≥35kg/m<sup>2</sup>; [≥30kg/m<sup>2</sup> BBJ]) and III (BMI≥40kg/m<sup>2</sup>; [≥35kg/m<sup>2</sup> BBJ]) and calculated its ability to predict obesity (Nagelkerke's R<sup>2</sup>; AUC<sub>ROC</sub>; PPV; sensitivity). The PGS explains the most BMI variance (r<sup>2</sup>=6.9%) in the UKBB<sub>EUR</sub>, ~4 times more than in other ancestries (1.7-2%), and slightly more in younger adults and women (except UKBB<sub>EUR</sub>: men>women). The explained variance decreased with

increasing obesity class (UKBB<sub>EUR</sub>  $R^2$ : 6.4-5.7%; other ancestries: 2-1%). Still, the predictive ability increased with increasing obesity class in UKBB<sub>EUR</sub> ( $AUC_{ROC}$  0.64-0.70), and improved when adding age and sex (0.65-0.73). The  $AUC_{ROC}$  in UKBB<sub>AFR</sub> and HCHS/SOL was low ( $<0.60$ ), but higher in BBJ (0.58-0.62). A predictive test that assumes the top PGS decile to predict obesity has a sensitivity of 17% and PPV of 42% in UKBB<sub>EUR</sub>, which increase to 28% and decrease to 4% in class III, resp. While sensitivity was lower in other ancestries, PPV was of similar magnitude.

Thus, the explained variance and predictive ability of the PGS is greater than any score reported before, but does not yet achieve the clinical utility minimum ( $AUC_{ROC} > 0.80$ ). Portability of PGS<sub>EUR</sub> to other ancestries is poor. We expect improvement as we build PGSs on ~40M variants from GIANT's upcoming meta-analyses ( $N > 3M$ ), develop ancestry-specific PGSs, and test their portability. We will also examine the predictive ability of PGSs while considering environmental factors.

## PgmNr 1935: Integrative analysis of GWAS and tissue-specific gene expression identifies candidate causal genes underlying obesity risk.

### Authors:

D. Hemerich; A. Moscati; M. Preuss; R. J. F. Loos

[View Session](#) [Add to Schedule](#)

**Affiliation:** The Charles Bronfman Institute for Personalized Medicine, The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Obesity, as assessed by body mass index (BMI), is a heritable risk factor for several diseases, affecting >2 billion people worldwide. Genome-wide association studies (GWAS) have identified hundreds of genetic variants influencing obesity risk, mostly in non-coding regions, enriched in the central nervous system. One way to prioritize genes in each locus is to assess whether any of the identified variants are associated with gene expression levels (eQTL). However, the subsequent link from gene expression to disease is rarely made. Summary statistics Mendelian Randomization (MR), or SMR, is an MR approach that integrates GWAS loci with eQTL, and has been used to assess the causal relationship between expression of genes and diseases, such as obesity, using eQTL on whole blood. Tissue-specific information increases power for eQTLs with tissue-specific effects.

Here, we used SMR to prioritize genes within 536 obesity-associated loci by assessing whether observed eQTLs are causally associated with BMI. We integrated brain eQTL from CommonMind Consortium (n=467), ROSMAP (n=494) and GTEx (n=72), and blood eQTL from eQTLGen (n=14,115). We sought further confirmation using other gene-prioritization approaches (DEPICT) and examining the presence of predicted damaging coding variants via Mutation Significance Cutoff (MSC) in each locus.

We identified 119 candidate genes whose expression in blood appear to be causally associated with BMI, and find similar evidence for 163 genes in brain eQTL data (significantly more than in blood, FET  $p < 0.0001$ ), with 20 genes shared in both sets. In addition, 249 genes were prioritized by DEPICT (FDR < 0.05). Finally, 371 genes contain coding variants with predicted damaging effect, 20 of which are rare (MAF < 1%). In total, we identified 756 candidate genes in 351 of the 536 BMI-associated loci, 35 of which were prioritized by more than one line of evidence, e.g. *SLC35E2*, *LUZP1*, *NUCKS1*, *MACF1*, *IPP*, *FAF1*, *NEGR1*, *GPR88* and *ADCY3*. Some of these genes have already been established as obesity genes (*NEGR1*, *ADCY3*, *FAF1*), and others are supported by evidence from animal models (*MACF1*, *GRP88*, *NRXN1*, *NUCKS1*). With additional prioritization tools, we will provide further candidate genes and additional evidence.

The prioritization of likely causal genes is a critical step in the translation of GWAS results into functional follow-up studies to increase our understanding of the underlying biology and, in time, develop better therapeutics for obesity.

## PgmNr 1936: Exome sequencing identifies multiple genes and gene-sets associated with severe childhood obesity.

### Authors:

A.E. Hendricks<sup>1,2</sup>; G. Marenne<sup>2,3</sup>; A. Perdikari<sup>4</sup>; R. Bounds<sup>4</sup>; F. Payne<sup>2</sup>; J. Keogh<sup>4</sup>; C. Lelliot<sup>2</sup>; E. Henning<sup>4</sup>; S. Pathan<sup>4</sup>; S. Ashford<sup>4</sup>; E. Bochukova<sup>4,5</sup>; V. Mistry<sup>4</sup>; A. Daly<sup>2</sup>; N. Wareham<sup>6</sup>; S. O'Rahilly<sup>4</sup>; C. Langenberg<sup>6</sup>; E. Wheeler<sup>2,6</sup>; E. Zeggini<sup>2,7</sup>; I.S. Farooqi<sup>4</sup>; I. Barroso<sup>2,6</sup>; INTERVAL, UK10K

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Mathematical and Statistical Sciences, University of Colorado Denver, Denver, Colorado.; 2) Wellcome Sanger Institute, Cambridge, UK; 3) Inserm, Univ Brest, EFS, UMR 1078, GGB, F-29200 Brest, France; 4) University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Wellcome-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK; 5) Blizard Institute, School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT; 6) University of Cambridge MRC Epidemiology Unit, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK; 7) Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

A substantial proportion of the heritability of obesity remains unexplained. To explore the contribution of rare variants to obesity, we analyzed whole-exome sequence (WES) data of severe childhood obesity cases (BMI standard deviation score > 3 and age of onset < 10 years) and healthy controls.

In stage 1, we used the WES data (927 cases and 4,057 INTERVAL controls) to perform gene-based and gene-set case-control analysis using all cases and cases stratified by developmental disorder (DD) status ( $N_{DD}=226$ ,  $N_{notDD}=701$ ). We used three nested variant filtering strategies: burden of very rare (minor allele frequency,  $MAF<0.025\%$ ) predicted loss-of-function variants (LOF); burden of very rare ( $MAF<0.025\%$ ) predicted deleterious variants by five *in silico* prediction tools (STRICT); SKAT-O analysis of variants with  $MAF<1\%$  and predicted deleterious by at least one tool (BROAD). Gene-based analyses were performed using SKAT in R. Gene-set analyses were performed using PLINK/SEQ gene-region tests, the SMP utility, and 20K permutations for gene-sets related to monogenic obesity, syndromic obesity, DD, loss-of-function intolerant ( $pLI>0.9$ ), or obesity/ BMI-associated loci (GWAS genes). In stage 2, 9 genes from stage 1 ( $p<10^{-4}$  and odds ratio,  $OR>1$ ) were sequenced in 1,810 cases and 2,647 Fenland controls, and meta-analyzed with stage 1.

We identify three genes (*PHIP*,  $P_{meta}=1.23\times10^{-5}$ ; *DGKI*,  $P_{meta}=7.93\times10^{-4}$ ; *ZMYM4*,  $P_{meta}=3.19\times10^{-7}$ ) with an excess burden of very rare variants predicted to affect function in obese cases compared to controls ( $OR>4.8$ ); *DGKI* and *ZMYM4* are novel. Deletions and frameshifts in *PHIP* have been reported in patients with DD and syndromic obesity. We observe for the first-time association with obesity without DD ( $P_{stage1}=7.00\times10^{-4}$ ,  $P_{stage2}=1.72\times10^{-2}$ ). Gene-set enrichment revealed an excess burden of very rare variants predicted to be deleterious in GWAS genes ( $OR=1.39$ ;  $CI_{95}=1.07,1.80$ ) and  $pLI>0.9$  genes ( $OR=1.17$ ;  $CI_{95}=1.07,1.27$ ). Enrichment was increased for GWAS genes that were also  $pLI>0.9$  ( $OR=3.54$ ;  $CI_{95}=1.80,6.95$ ). Results stratified by DD status suggest that  $pLI>0.9$  genes may have mutations underlying obesity with developmental delay while GWAS genes also  $pLI>0.9$  may point to

genes underlying common obesity.

In conclusion, WES of severe childhood obesity and healthy controls identifies novel genes and gene-sets influencing obesity in a non-fully penetrant manner, provides mechanistic insights and has diagnostic and therapeutic implications.

# PgmNr 1937: The association between the gut virome and adiposity traits in women.

## Authors:

Z. Hu <sup>1</sup>; Z. Zhu <sup>2</sup>; X. Lin <sup>3</sup>; K.J. Su <sup>4</sup>; H. Shen <sup>4</sup>; F. Sun <sup>2</sup>; J. Shen <sup>3</sup>; H. Xiao <sup>5</sup>; H.W. Deng <sup>4,5</sup>; Q. Zhao <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN, USA; 2) Quantitative and Computational Biology Program, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA; 3) Department of Endocrinology and Metabolism, The Third Affiliated Hospital of Southern Medical University, Guangzhou, China; 4) Center for Bioinformatics and Genomics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA; 5) School of Basic Medical Sciences, Central South University, Changsha, China

## Aim

The relationship of gut virome with obesity is still not clear. This study systematically investigated the associations between gut virome and multiple adiposity traits in women.

## Methods

This study included 499 Chinese women aged 41-65 years. Total body and regional measures of total mass and fat mass were acquired using the dual-energy X-ray absorptiometry. Fecal samples were collected and analyzed using the Hiseq 4000 platform to obtain viral sequences, which were classified by taxonomic alignment against the NCBI reference genome database using Centrifuge software. The alpha diversity of gut virome was estimated using the Simpson index. The multivariate analysis was used to examine the associations of the alpha diversity and the abundance of individual viruses with multiple adiposity traits [body mass index (BMI), waist circumference (WC), hip circumference (HC), waist/hip ratio (WHR), whole body fat ratio (WBFR), android fat ratio (AFR), gynoid fat ratio (GFR), and android/gynoid fat ratio (AGFR)], with adjustment for age. The false discovery rate method was used for adjusting for multiple testing.

## Results

A total of 393 viruses were identified, 67 of them presented in over 10% of the sample and were subjected to further analysis. The alpha diversity of the gut virome was significantly associated with adiposity traits ( $P = 0.03$ ). Specifically, the Simpson index was significantly associated with BMI ( $\beta = 0.26$ ,  $P = 0.049$ ), WC ( $\beta = 1.12$ ,  $P = 0.002$ ), WHR ( $\beta = 0.008$ ,  $P = 0.009$ ), WBFR ( $\beta = 0.006$ ,  $P = 0.01$ ), AFR ( $\beta = 0.01$ ,  $P = 0.001$ ), and AGFR ( $\beta = 0.01$ ,  $P = 0.03$ ). In the individual virus analysis, three viruses, *Shigella phage SflI*, *Escherichia virus Wphi*, and *Escherichia phage Min27*, were significantly associated with multiple adiposity traits ( $q$  values for the overall tests =  $2.1 \times 10^{-4}$ , 0.008, and 0.036, respectively). Specifically, *Shigella phage SflI* was associated with BMI ( $\beta = 24.07$ ,  $P = 0.03$ ), WC ( $\beta = 89.18$ ,  $P = 0.04$ ), and WHR ( $\beta = 1.24$ ,  $P = 6.1 \times 10^{-6}$ ). *Escherichia virus Wphi* was associated with WC ( $\beta = 29.33$ ,  $P = 0.04$ ), WHR ( $\beta = 0.44$ ,  $P = 5.4 \times 10^{-4}$ ), and AGFR ( $\beta = 0.71$ ,  $P = 0.007$ ). *Escherichia phage Min27* was associated with AFR ( $\beta = -7.81$ ,  $P = 3.9 \times 10^{-4}$ ) and AGFR ( $\beta = -15.33$ ,  $P = 0.002$ ).

## Conclusion

Our study findings, for the first time, suggest the potential contribution of the gut virome in human adiposity traits. These findings may provide further insights into the biological mechanisms linking gut microbiota with the development of obesity in human.



## PgmNr 1938: Association study of *LEP* -2548 G/A polymorphism with obesity in North Indian Punjabi population.

### Authors:

H. Kaur; V. Bains; B. Badaruddoza

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, Guru Nanak Dev University, Amritsar, India

---

Obesity is one of the biggest health epidemic worldwide. It is a well known nutritional disorder among developed countries, and is characterized as a major risk factor for cardiovascular diseases. It is a complex, multifactorial medical condition, affected by a strong interplay among genetic and environmental factors. Obesity is reported to be associated with several genes in addition to known factors, such as diet and lifestyle behavior. Leptin is an adipocyte derived hormone that plays a significant role in homeostatic system, by suppressing food intake and increased energy expenditure by binding to its specific receptor in the hypothalamus. Several mutations in the leptin gene have been shown to cause morbid obesity. Thus, the present study aimed to investigate the association of *LEP* -2548G/A polymorphism with obesity and its related traits in North-Indian Punjabi population. A total of 250 obese subjects and 300 non-obese controls were recruited for the present study from different regions of Punjab. Various anthropometric and physiometric measurements were taken for all the subjects using standardized protocols with written informed consent. Biochemical analysis was performed on the collected blood serum samples to estimate the lipid profile of all the subjects. The *LEP* polymorphism (rs7799039, -2548G/A) was screened using polymerase chain reaction based restriction fragment length polymorphism method and 10% of the samples were confirmed using sanger sequencing method. The frequency of homozygous recessive AA genotype was found to be significantly higher among obese subjects (14%) as compared to non-obese controls (8.3%). The obese subjects possessing the -2548 AA risk genotype had significantly higher BMI, WC, WHR, and TG levels in comparison to GA and GG genotypes. A significant association of -2548 G/A polymorphism was observed in the obese cohort under a recessive model, with the AA genotype providing two fold risk towards the susceptibility of obesity after adjusting for age, gender, diet and lifestyle behavior [OR (95% CI) : 2.06(1.14-3.74), p=0.016]. The -2548 A allele and the selected obesity related covariates accounted for 53%, 26% and 30.2% variation in body mass index (BMI), waist-to-hip ratio (WHR) and triglycerides (TG) respectively. This study provides a evidence of association of *LEP* -2548G/A polymorphism with various anthropometric traits and obesity risk in North Indian Punjabi population.

## PgmNr 1939: Genetic determinants of appetite hormones in adolescence: The Santiago Longitudinal Study (SLS).

### Authors:

D. Kim <sup>1</sup>; A.E. Justice <sup>2</sup>; G. Chittoor <sup>1,2</sup>; E. Blanco <sup>3</sup>; M. Graff <sup>1</sup>; Y. Wang <sup>1</sup>; R. Rohde <sup>1</sup>; C. Albala <sup>4</sup>; J.L. Santos <sup>5</sup>; B. Angel <sup>4</sup>; B. Lozoff <sup>6</sup>; V.S. Voruganti <sup>7</sup>; K.E. North <sup>1</sup>; S. Gahagan <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 2) Biomedical and Translational Informatics, Geisinger, Danville, PA, USA; 3) Division of Academic General Pediatrics, Child Development and Community Health at the Center for Community Health, University of California at San Diego, San Diego, CA, USA; 4) Department of Public Health Nutrition, Institute of Nutrition and Food Technology (INTA), University of Chile, Santiago, Chile; 5) Department of Nutrition, Diabetes and Metabolism, School of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile; 6) Department of Pediatrics, University of Michigan, Ann Arbor, MI, USA; 7) Department of Nutrition and Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis NC 28081, USA

**Introduction:** Adolescence is a critical period for the development of obesity and Hispanic/Latinos (H/L) are at especially high risk. Chronic imbalance of energy intake and expenditure, and further poor regulation of appetite leads to the development of obesity. Despite their importance, there is a paucity of studies on the genetic underpinnings of appetite-related hormone levels, in particular during adolescence. Thus, the current study aims to identify genetic factors associated with appetite hormone levels: leptin, adiponectin, ghrelin, and orexin among H/L adolescents.

**Methods:** Fasting serum appetite hormone levels were measured on a total of 543 H/L adolescents from the Santiago Longitudinal Cohort study aged 16-17 years. Genotypes were generated on the Illumina Multiethnic Genotyping Array with imputation to the 1000 Genomes Phase III AMR reference panel. Quality control included individual call rate >90%, SNP call rate >95%, imputation quality >0.5, Effective Allele Frequency (EAF) >0.05, gender mismatch, and ancestry outliers. We assessed the relationship between genetic variants and appetite biomarkers assuming an additive genetic model and adjusting for sex and population structure using principal components.

**Results:** We identified 39 suggestive loci ( $P < 5 \times 10^{-6}$ ,  $\pm 500$ kb, EAF range: 0.05–0.80) across four appetite biomarkers (9 for leptin, 8 for adiponectin, 9 for orexin, and 13 for ghrelin). Our top finding for orexin was for rs8036016 (EAF:0.17, beta:0.11,  $p: 6.8 \times 10^{-8}$ ). This finding is noteworthy as this intronic SNP is in the *GABRB3*, which has been previously associated with BMI and Prader Willi syndrome. Additionally, we generalized several SNP-trait associations previously observed in Non-Hispanic ancestry adults, including 10 SNPs in four different loci (*ADIPOQ*, *ST13P4*, *LYPLAL1*, *RFC4*) for adiponectin. A novel independent secondary signal was observed for adiponectin on chromosome 12, in *NCOR2*.

**Conclusion:** Our study findings demonstrate the importance of considering ancestry-specific variants, along with generalization of genetic variants previously identified in European adults, for a greater understanding of the role of genetic variation on appetite biomarkers. From a public health standpoint, this genetic interrogation of obesity and appetite biomarkers in adolescence will advance science in developmental pathophysiology, particularly relationships between obesity and appetite. Findings from the study need to be replicated in future research.



## PgmNr 1940: Genetic risk score for adult body mass index is associated with child and adolescent weight gain in an African population.

### Authors:

Z. Lombard <sup>1</sup>; R.J. Munthali <sup>2,3,4</sup>; V. Sahibdeen <sup>1</sup>; J. Kagura <sup>4</sup>; L.M. Hendry <sup>3</sup>; S.A. Norris <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Human Genetics, Faculty of Health Sciences, School of Pathology, University of the Witwatersrand and National Health Laboratory Service, Johannesburg, South Africa; 2) Wits Reproductive Health and HIV Institute, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; 3) Faculty of Science, School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, South Africa; 4) MRC/Wits Developmental Pathways for Health Research Unit (DPHRU), University of the Witwatersrand, Johannesburg, South Africa

---

Ninety-seven independent single nucleotide polymorphisms (SNPs) are robustly associated with adult body mass index (BMI kg/m<sup>2</sup>) in Caucasian populations. The relevance of such variants in African populations at different stages of the life course (such as childhood) is unclear. We tested whether a genetic risk score composed of the aforementioned SNPs was associated with BMI from infancy to early adulthood. We further tested whether this genetic effect was mediated by conditional weight gain at different growth periods. Finally, we compared these results to a genetic risk score composed of SNPs found to be associated with BMI in this cohort to contrast correlation. We used data from the Birth to Twenty Plus Cohort (Bt20+), for 971 urban South African black children from birth to 18 years. DNA was collected at 13 years old and was genotyped using the Illumina MetaboChip array. The weighted genetic risk score (wGRS) for BMI was constructed based on 71 of the 97 previously reported SNPs. The cross-sectional association between the wGRS and BMI strengthened with age from 5 to 18 years. The significant associations were observed from 11 to 18 years, and peak effect sizes were observed at 13 and 14 years of age. Results from the linear mixed effects models showed significant interactions between the wGRS and age on longitudinal BMI but no such interactions were observed in sex and the wGRS. A higher wGRS was associated with an increased relative risk of belonging to the early onset obese longitudinal BMI trajectory (relative risk = 1.88; 95%CI 1.28 to 2.76) compared to belonging to a normal longitudinal BMI trajectory. Adolescent conditional relative weight gain had a suggestive mediation effect of 56% on the association between wGRS and obesity risk at 18 years. The results suggest that genetic susceptibility to higher adult BMI can be tracked from childhood in this African population. This supports the notion that prevention of adult obesity should begin early in life. The genetic risk score combined with other non-genetic risk factors, such as BMI trajectory membership in our case, has the potential to be used to screen for early identification of individuals at increased risk of obesity and other related NCD risk factors in order to reduce the adverse health risk outcomes later.

## PgmNr 1941: Adiposity genetic risk score predicts the development of obesity from childhood to adulthood.

### Authors:

M. Shi <sup>1</sup>; W. Chen <sup>1</sup>; J. He <sup>1</sup>; C. Li <sup>2</sup>; L.A. Bazzano <sup>1</sup>; T.N. Kelly <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, Louisiana.; 2) Department of Epidemiology & Biostatistics, University of Georgia College of Public Health, Athens, Georgia

---

**Background** The global burden of obesity is rising rapidly. Given its well-known adverse cardiometabolic consequences, we investigated whether genetic risk score (GRS) for adiposity may help to predict the development of obesity across the lifespan.

**Methods** The current study was carried out among 1,111 Bogalusa Heart Study (BHS) participants (782 whites and 329 African-Americans) with genome-wide association study (GWAS) data and at least one childhood and one adulthood measure of BMI. An effect-size weighted adiposity GRS was constructed using 97 independent variants that reached genome-wide significance in previous GWAS meta-analyses of BMI, with participants categorized according to quartile of GRS. Associations of both GRS quartile and the continuous GRS with obesity incidence were assessed in race-stratified analyses using two Cox proportional hazards models. Model 1 adjusted for baseline age, sex and ancestry principal components, while model 2 adjusted for covariables in model 1 along with baseline BMI.

**Results** On average, participants attended 7.6 study visits and were followed-up for approximately 29 years. At baseline, the mean age was 9.9 years and BMI was 17.2 kg/m<sup>2</sup>. Results of model 1 showed a positive, graded association between GRS quartile and incidence of obesity among participants of European ancestry (p-value for linear trend=  $2.50 \times 10^{-5}$ ), with a hazard ratio (HR) [95% confidence interval (CI)] of 1.94 (1.43, 2.64) when comparing the highest GRS quartile to the lowest GRS quartile. Although the association was attenuated when baseline BMI was added to the model (Model 2), with a HR (95% CI) of 1.54 (1.09, 2.19), the association remained statistically significant (p-value for linear trend=  $5.00 \times 10^{-3}$ ). Each standard deviation increase in the continuous GRS conferred a 26% and 15% significantly increased risk of obesity in models 1 and 2, respectively ( $P=2.10 \times 10^{-5}$  and  $1.47 \times 10^{-3}$ , respectively). There was no association between GRS and obesity incidence among African-American participants.

**Conclusion** Our results indicated that adiposity GRS may help to predict the development of obesity in individuals of European ancestry, even after accounting for childhood BMI. Given the rapidly decreasing genotype costs, genetic information should be further evaluated as a tool for targeted early prevention of this condition.

## PgmNr 1942: Integration of body fat and lean mass loci reveals genetic clusters with distinct cardiometabolic effects.

### Authors:

L.B.L. Wittemans<sup>1,2</sup>; Y. Lu<sup>3,4,5</sup>; H. Yaghooskar<sup>6,7</sup>; M.H. Preuss<sup>3</sup>; A.M. Erzurumluoglu<sup>2</sup>; D. Hemerich<sup>3</sup>; Y. Ji<sup>6</sup>; T. Karaderi<sup>1,8,9,10</sup>; C. Schurmann<sup>11</sup>; C.N. Spracklen<sup>12</sup>; L.O. Huang<sup>13</sup>; F.R. Day<sup>2</sup>; M.C. Zillikens<sup>14,15</sup>; T. Frayling<sup>6</sup>; Z. Kutalik<sup>6,16,17</sup>; T.O. Kilpeläinen<sup>13,18</sup>; C.M. Lindgren<sup>1,19,20</sup>; K.L. Mohlke<sup>12</sup>; C. Langenberg<sup>2</sup>; R.J.F. Loos<sup>3</sup>; The Genetics of Body Composition Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; 2) MRC Epidemiology Unit, University of Cambridge, Cambridge, UK; 3) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, USA; 4) Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, USA; 5) Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, USA; 6) Genetics of Complex Traits, University of Exeter Medical School, Royal Devon & Exeter Hospital, Exeter, UK; 7) Research Centre for Optimal Health, School of Life Sciences, University of Westminster, London, UK; 8) Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean University, Famagusta, Cyprus; 9) Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 10) DTU Health Technology, Technical University of Denmark, Lyngby, Denmark; 11) Hasso-Plattner-Institute for Digital Engineering, Digital Health Center, Potsdam, Germany; 12) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, USA; 13) Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 14) Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands; 15) Netherlands Genomics Initiative - sponsored Netherlands Consortium for Healthy Aging, Leiden, The Netherlands; 16) University Center for Primary Care and Public Health, University of Lausanne, Lausanne, Switzerland; 17) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 18) Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, USA; 19) Broad Institute of MIT and Harvard, Cambridge, USA; 20) Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK

---

Body fat and lean mass play distinct roles in cardiometabolic risk but their genetic drivers have not been studied at scale. To identify genetic loci for body fat percentage (BF%) and fat-free mass index (fat-free mass/height<sup>2</sup>, FFMI), we performed GWAS on 463,352 UK Biobank participants and exome chip meta-analyses on 112,443 individuals. We identified 1,197 loci for BF% or FFMI, of which 209 reached genome-wide significance for both traits ( $p < 5 \times 10^{-9}$ ). Despite pronounced sex differences in BF% and FFMI, only 3% of loci had sex-dimorphic effects. To prioritise genes affecting body composition, we performed conditional analyses at 65 coding variants (MAF < 5%) that reached genome-wide significance. We identified 20 independently associated coding variants, including variants in *ZNF800*, *RAPGEF3*, *PDE3B* and *ZBTB7B*. The latter (MAF = 2.3%) has been reported for higher BMI and we show that it exerts its effect through increased FFMI with no effect on BF%.

To assess the implications of body composition on health, we created genetic scores following K-means clustering on betas for BF%, FFMI, BMI, height, and WHRadjBMI for common variants reaching genome-wide significance. For each cluster, we conducted tissue enrichment analyses using DEPICT, Mendelian randomization analyses for cardiometabolic traits and association testing of cluster-specific genetic scores to traits in 3 cohorts. We identified 6 clusters, of which 3 represent overall large body size (higher BF%, FFMI and BMI) that show enrichment in the central nervous system and associations with higher cardiometabolic risk. Two additional clusters reflect normal weight adiposity (higher BF%, lower BMI and FFMI) that are enriched in digestive, bone, muscle and adipose tissues. Genetic scores for these 2 clusters are positively associated with leptin, inflammatory and glycemic markers and T2D risk. The last cluster, representing gluteofemoral adiposity (higher BF% and BMI, lower WHRadjBMI) is enriched in adipose tissue and associated with lower glycemic markers, a favorable lipid profile and lower T2D and CHD risk.

Systematic integration of associations of BF% and FFMI loci with multiple anthropometric traits can identify genetic clusters with distinct cardiometabolic effects not adequately captured by BMI. Our findings support the notion that genetic susceptibility to high BF% increases cardiometabolic risk even in the context of a lower BMI, and that higher BF% driven by gluteofemoral fat is metabolically favorable.

## PgmNr 1943: Performance of genetic risk scores in a multi-ethnic study.

### Authors:

S.A. Bien<sup>1</sup>; G.L. Wojcik<sup>2</sup>; J. Haessler<sup>1</sup>; R. Tao<sup>3</sup>; G.M. Belbin<sup>4</sup>; M. Graff<sup>5</sup>; H.M. Highland<sup>5</sup>; Y.M. Patel<sup>6</sup>; C. Hodonsky<sup>5</sup>; K.L. Young<sup>5</sup>; C.L. Avery<sup>5</sup>; S. Buyske<sup>7</sup>; C. Haimon<sup>6</sup>; L. Le Marchand<sup>8</sup>; R.J.F. Loos<sup>4</sup>; T.C. Matisse<sup>7</sup>; C.S. Carlson<sup>1</sup>; U. Peters<sup>1</sup>; E.E. Kenny<sup>4</sup>; K.E. North<sup>5</sup>; C.R. Gignoux<sup>9</sup>; C. Kooperberg<sup>1</sup>; PAGE

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington.; 2) Stanford University, Stanford CA; 3) Vanderbilt University Medical Center, Nashville TN; 4) Icahn School of Medicine at Mount Sinai, New York NY; 5) University of North Carolina at Chapel Hill, Chapel Hill NC; 6) Keck School of Medicine, University of Southern California, Los Angeles CA; 7) Rutgers University, New Brunswick NJ; 8) University of Hawaii, Honolulu HI; 9) University of Colorado Anschutz Medical Campus, Aurora CO

---

Genomic studies have provided critical new insight into the biology of complex traits, drug development, and clinical guidelines. However, these studies have overwhelmingly been performed in populations of European descent, biasing our understanding of complex trait genetics.

In light of differential genetic architecture known to exist between populations, biased representation in research can exacerbate existing disease and health care disparities. Critical variants may be missed if they are at a low frequency or completely absent in European populations. Even more concerning, in previous work we showed strong effect-size heterogeneity across ancestries for published GWAS associations. As such, risk prediction scores derived in one population will be unlikely to accurately rank an individual's risk in other populations.

Herein, we describe the recent work of the Population Architecture using Genomics and Epidemiology (PAGE) study to investigate how current derivations of genetic risk prediction tools would perform for cardiometabolic phenotypes in participants of diverse ancestry. Using the PAGE data comprising over 51,000 individuals with globally representative genetic ancestry, we created polygenic risk scores (PRSs) for harmonized cardiometabolic traits based on published effect sizes from established GWAS consortia of predominantly European-descent participants. Race/ethnicity-specific estimates of correlation,  $R^2$ , between the PRS and the measured trait were computed. For all traits, the  $R^2$  for African Americans is ~40-50% smaller than the  $R^2$  for European ancestry; the  $R^2$  for Hispanics and Asians falls in the middle except for lipid traits, where the  $R^2$  for Hispanics is similar to that for European ancestry. Similar trends were also observed when extending this framework to more recent genome-wide risk prediction scores employing millions of variants weighted for European linkage disequilibrium with LDpred. As expected overall, using European-derived SNP weights to generate multi-ethnic PRS results in substantial reduction of prediction accuracy. We describe implications both for understanding genome architecture and for personalized medicine in diverse, modern datasets such as PAGE.



## PgmNr 1944: Association between genetic admixture estimates for 5 ancestral populations and serum lipid levels in Pacific Islanders.

### Authors:

W.-C. Hsueh <sup>1</sup>; P. Kumar <sup>1</sup>; S. Saied Safabakhsh <sup>2</sup>; L.I. Jones <sup>1</sup>; J.E. Curran <sup>3</sup>; W.C. Knowler <sup>1</sup>; R.G. Nelson <sup>1</sup>; R.L. Hanson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NIDDK, Phoenix, Arizona.; 2) Micronesian Institute for Disease Prevention and Research, Sinajana, Guam; 3) South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, Texas

---

Extensive epidemiological literature has shown differential risks for cardiovascular disorders (CVD) among different ethnic groups across the globe, but whether such differences may be due to genetic or non-genetic factors is hard to dissect due to cultural and environmental confounders. We conducted a community-based study aiming at identifying genetic risk factors for several complex diseases in a high-risk population of Pacific Islanders residing in Guam and Saipan. We analyzed the genetic structure of our study population and individual-level admixture estimates based on unsupervised clustering algorithm implemented in ADMIXTURE software using genome-wide SNP data and a reference panel of diverse global populations. We found that our study population had a high degree of admixture derived from 5 ancestral populations, representing Marianas Islanders (MI) (mean percentage of heritage: 43%), East Asians (EA) (mean: 22%), Micronesians (MC) (mean: 19%), Europeans (EU) (mean: 13%), and Melanesians (mean: 3%). Here, we examined the association of these genetic ancestry estimates with 3 fasting serum lipid measures - total cholesterol (total-C), HDL-C and triglycerides (TG) - in 1,853 subjects. Covariates included age, sex and diabetes. All 3 traits were significantly associated with genetic ancestry ( $p < e-4$  to 0.035). Individuals with MI ancestry tended to have modestly higher total-C than other ancestries. Total-C levels were higher in MI compared to MC ( $\beta = 0.066$  s.d. unit difference between 2 groups,  $p = 0.0076$ ), while the levels were comparable among the other ancestral backgrounds. On the other hand, EA ancestry was associated with the highest HDL-C levels ( $\beta = 0.93$  s.d. unit,  $p < e-4$ , compared to all other ancestries combined), while the levels were comparable among the 4 other ancestral backgrounds. MI ancestry was also associated higher TG levels compared to the other ancestral groups. MI ancestry had significantly higher TG compared to EA ( $\beta = 0.50$  s.d. unit,  $p = 0.0020$ ), EU ( $\beta = 0.70$  s.d. unit,  $p = 0.0079$ ) and MC ( $\beta = 0.86$  s.d. unit,  $p < e-4$ ). In summary, there were significant and large differences in lipid levels among people with different genetic ancestral backgrounds, particularly for HDL-C and TG. Although genetic ancestry estimates may still be confounded by shared environmental factors, the present findings suggest that admixture mapping may be potentially a powerful approach for identifying unique CVD loci in Pacific Islanders.

## PgmNr 1945: Accurate prediction of human height and Body Mass Index (BMI) in Asian population.

### Authors:

N. Kim <sup>1,2</sup>; S. Jeong <sup>1</sup>; J.-Y. Kim <sup>1,2</sup>; K.H. Kim <sup>1,2</sup>; W.-J. Lim <sup>1,2</sup>; H.-Y. Lee <sup>1,2</sup>; Y. Na <sup>1,2</sup>; Y. Cho <sup>1,2</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Genome Editing Research Center, KRIBB, Daejeon, Daejeon, Korea, Republic of; 2) Department of Bioinformatics, Korea University of Science and Technology, Daejeon, Korea, Republic of

---

High-throughput genotyping technologies such as high-density SNP-array and Next-Generation Sequencing gives us opportunity to understand human complex traits and diseases better. One can perform targeted-sequencing that mostly focuses on disease associated (clinically-associated) rare variants. And we can simply compare all derived variants with known databases, HGMD and ClinVar. However, it is well-known that rare variants could only solve less than 15% of all cases for complex traits. We still do not fully understand every part of our genome, how those non-annotated part could do their role. Polygenic risk scoring (PRS) uses previously reported tens and hundreds of alleles with high odd-ratios. But, we cannot guarantee that those variants are common in general population without any symptoms. Moreover, it would not be possible to estimate complex traits and diseases with such a small number of alleles. Another limitation is that one cannot predict if there is very little previous study. There are several argument that polygenic risk scoring does not doing well in Asian and African populations. We have developed a novel algorithms, RRMS, which can give least number of candidate markers for accurate prediction of traits with low heritability by deep learning method. We have successfully demonstrate our methods to predict human BMI and Height. We were able to select 1,725 SNPs to predict human BMI from 8,840 individuals with correlation  $R=0.7$  (estimated vs. observed). 2,216 SNPs to height with  $R=0.8$ . Previous estimation stated that it requires order of 4 (tens of thousands of markers) to predict heritability range between 0.2 ~ 0.3. It requires thousands of SNPs to predict height and BMI due to low heritability. We have compared with UKBB and KNIH data for validation and ethnic differences. We will show what would be the differences between Caucasian vs. Asian in terms of haplotype and marker effects.

## PgmNr 1946: Single nuclei transcriptome analysis of human liver integrated with capture HiC and proteomics data.

### Authors:

M. Cavalli<sup>1</sup>; K. Diamanti<sup>2</sup>; G. Pan<sup>1</sup>; C. Kumar<sup>3</sup>; A.S. Deshmukh<sup>4</sup>; M. Mann<sup>4</sup>; P. Salhen<sup>5</sup>; J. Komorowski<sup>2,6</sup>; C. Wadelius<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Science for Life Laboratory, Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden; 3) AstraZeneca, Gothenburg, Sweden; 4) Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark; 5) Science for Life Laboratory, Division of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden; 6) Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

---

The liver is the largest solid organ and a primary metabolic hub. Recent advances in next generation sequencing allow the study of gene expression profiles of single cells, potentially leading to discovery of new and rare cell sub-populations. In recent years, intact cell nuclei were used to perform single-nuclei RNA-seq (snRNA-seq) for cell types difficult to dissociate and for flash-frozen archived tissue samples. Several studies have reported a high correlation between nuclei and cells relative gene expression. The difference in transcriptomics profiles between defined liver cell populations stems from a precise genetic control of gene expression mostly driven by cell type specific enhancers. HiCap is a targeted chromosome conformation capture technique that allows studying long-range distal interactions between such enhancers and their target promoters. Mass spectrometry-based proteomics analysis are today the first choice to verify the tissue protein levels. In this study, we performed snRNA-seq of a human liver sample to identify sub-populations of cells based on the nuclear transcriptomics. In 4282 single nuclei the average number of active genes were 1377 per nucleus. The major cell types were hepatocytes, sinusoidal endothelial cells, cholangiocytes, Kupffer cells, Nk/T/B cells and interestingly both activated and several groups of inactive hepatic stellate cells. We then integrated long-range HiCap data from the same liver for 23 497 promoters and found 119 233 interactions ( $p < 0.05$ ) to distal elements. We observed reasonable correlation between proteomics and *in silico* snRNA-seq bulk data ( $r = 0.47$ ) using tissue-independent gene-specific protein abundance estimation factors. We specifically looked at genes of medical importance. The *DPYD* gene is involved in the pharmacogenetics of fluoropyrimidines toxicity and some of its variants are analyzed for clinical purposes. We identified a new putative polymorphic regulatory element, which could contribute to variation in toxicity. Hepatocellular carcinoma (HCC) is the most common type of primary liver cancer and we investigated all known risk genes. We found a complex regulatory network for the *SLC2A2* gene with 16 candidate enhancers. Three of them harbor somatic motif breaking and other mutations in HCC in the Pan Cancer Analysis of Whole Genomes dataset and are candidates to contribute to malignancy. Our results highlight the potential of a multi-omics approach in the study of human diseases.

## PgmNr 1947: Genetic architecture of molecular phenotypes and their relationship with the genetics of T2D: A DIRECT study.

### Authors:

A. Viñuela <sup>1</sup>; J. Fernandez <sup>2</sup>; A. Kurbasic <sup>3</sup>; M.G. Hong <sup>4</sup>; S. Sharma <sup>5</sup>; C. Brorsson <sup>6</sup>; J. Adamski <sup>5</sup>; J.M. Schwenk <sup>4</sup>; E.R. Pearson <sup>7</sup>; P.W. Franks <sup>3</sup>; M.I. McCarthy <sup>2</sup>; E.T. Dermitzakis <sup>1</sup>; DIRECT Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Geneva, Switzerland; 2) U. of Oxford, WTCHG, Oxford, United Kingdom; 3) U. of Lund, Clinical Sciences, Malmö, Sweden; 4) SciLifeLab, KTH Royal Institute of Technology, Solna, Sweden; 5) Helmholtz Zentrum, Genome Analysis, Munich, Germany; 6) DTU, Systems Biology, Lyngby, Denmark; 7) U. of Dundee, Molecular & Clinical Medicine, Dundee, United Kingdom

While GWAS studies have produced many associations between genetic variants and complex diseases, we still lack an understanding of the cellular changes that mediate their activity. For this, the DIRECT consortium built a cohort of 3,029 prediabetic and newly diagnosed T2D subjects, genotyped and with gene expression, metabolites and targeted proteins measurements in whole blood. Here, we present work on understanding how genetic variants perturb molecular networks, and their relationship with genetics of diseases.

First, we identified 59,972 independent cis-eQTLs for 15,305 genes, 1,593 independent cis-pQTLs for 397 proteins; as well as 2,320 trans-QTLs for 1,670 genes and 6,033 for 224 proteins (FDR 1%). We observed a substantial allelic heterogeneity, with 83.7% of genes and 84.1% of proteins associated with two or more variants, with many signals shared: 413 pQTLs shared lead SNP or were in high LD ( $R^2 > 0.9$ ) with an eQTL for the same gene (*CXCL1*, *IL18*), or nearby genes. However, we saw discordance in effect size across omics with 67% of the strongest pQTLs associated with weaker eQTLs. Across tissues, we identified in blood the majority of the eQTL signals from the 44 GTEx tissues and the InsPIRE pancreatic islets ( $\pi_1$  from 81% (testis) to 98% (brain)). For pQTLs we observed between 4.1% (heart) and 43.1% (adipose) of eQTLs in other tissues as pQTLs in whole blood. However 20 proteins, such as GDF2 and MBL2, had no expression of the corresponding gene in blood, pointing out to genetic regulation of gene expression in other tissues such as liver were detected as pQTLs in blood.

Next, we looked for evidence of shared genetic effects across QTLs and GWAS variants, identifying 99 co-localizing eQTLs and GWAS variants for T2D, including a co-localizing signal for the *TCF7L2* locus in a secondary signal (rs7918400). Although no eQTL was found in blood to share a signal for the strongest known GWAS association (rs7903146). Finally, to investigate how the molecular network responds to T2D development, we looked for evidence of T2D specific genetic regulation identifying 66 significant genotype-by-T2D interactions affecting gene expression and 113 affecting protein levels (FDR 5%).

Our results show how large multi-omics studies in blood can recapitulate the genetic effects from other tissues, suggesting candidate genes for GWAS variants as a consequence of our ability to replicate most genetic effects found in smaller, disease relevant studies.

## PgmNr 1948: Whole genome sequencing identifies novel early onset asthma susceptibility locus.

### Authors:

P. Sleiman<sup>1,2</sup>; H. Gui<sup>3</sup>; J.J. Connolly<sup>1</sup>; M.A. Behr<sup>1</sup>; A. Mak<sup>4</sup>; M. White<sup>4</sup>; S. Oh<sup>4</sup>; M. Seibold<sup>5</sup>; F.J. Salley<sup>1</sup>; A. Thomas<sup>1</sup>; J. Snyder<sup>1</sup>; F. Mentch<sup>1</sup>; E. Burchard<sup>4</sup>; L.K. Williams<sup>3</sup>; H. Hakonarson<sup>1,2</sup>; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, CHOP, Philadelphia, PA.; 2) Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, PA; 3) Center for Individualized and Genomic Medicine Research (CIGMA), Department of Internal Medicine, Henry Ford Health System, Detroit, MI; 4) Department of Bioengineering, University of California, San Francisco, CA; 5) Department of Pediatrics, National Jewish Health, Denver, CO

---

Asthma is one of the most common chronic conditions in children. However, despite the importance of asthma in pediatrics most genetic studies carried out to date have been performed in adults of European ancestry. As such, the basis of early onset asthma remains poorly understood, particularly in individuals of non-European ancestry. Here, we sought to identify genetic determinants of early onset asthma in a discovery cohort of 2113 African American (AA) asthmatics with recurrent exacerbations vs. 2081 AA controls from the Center for Applied Genomics (CAG) biobank at CHOP. Whole genome sequencing (WGS) data was generated on all 4194 samples through the TOPMed program. The replication cohort consisted of 3922 AA samples from the SAPPHIRE cohort, 2088 cases with asthma onset  $\leq 12$  vs and 1834 controls. Replication WGS data was also generated through TOPMed.

Association testing was performed using generalized linear mixed models (GLMM) in GMMAT. GMMAT fits a GLMM with covariate adjustment and random effects to account for population structure and relatedness. Score tests were generated for each variant followed by Wald tests to estimate effect sizes. For rare variants, burden and SKAT tests were carried out using SMMAT.

Association analysis identified multiple genome wide significant SNPs at a novel locus on chr7p15.3 that has not been previously associated with asthma. The association was replicated in an independent cohort of early onset AA asthmatics from the SAPPHIRE study. The associated variants mapped to an intergenic region upstream of *DNAH11* (rs2529168, disc P  $4.7 \times 10^{-8}$ , OR 1.47; rep P 0.018). *DNAH11* encodes a ciliary dynein protein that is involved in the movement of respiratory cilia. Recessive LOF mutations in *DNAH11* result in primary ciliary dyskinesia which is characterized by bronchiectasis and upper respiratory tract infections. Rare variant analysis in *DNAH11* identified 17 potentially pathogenic alleles in the cohort distributed between 12 rare variants. Burden and SKAT tests were not significant, indicating the mutations were unlikely to be driving the association. We report association of variants at 7p15.3 in WGS data from AA children with early onset asthma and recurrent exacerbations, a locus which has not been previously reported in larger GWAS of asthma in European ancestry adults. This study highlights the benefit and necessity of phenotypically targeted studies and the inclusion of ancestrally diverse individuals in genetic studies.

## PgmNr 1949: Genetics of very early onset inflammatory bowel disease: *IFIH1* and *PLCG2* as causative gene loci and evidence for a gene burden model.

### Authors:

E. Wohler<sup>1</sup>; J. Posey<sup>2</sup>; Z.H.C. Akdemir<sup>2</sup>; S.N. Jhangiani<sup>3</sup>; R. Kellermayer<sup>7,8</sup>; J.R. Lupski<sup>2,3,4,5</sup>; D. Valle<sup>1</sup>; A.L. Guerrierio<sup>6</sup>; N.L.M. Sobreira<sup>1</sup>; Baylor-Hopkins Ctr for Mendelian Genomics (BHCMG)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 5) Texas Children's Hospital, Houston, TX; 6) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD; 7) Section of Pediatric Gastroenterology, Texas Children's Hospital, Baylor College of Medicine, Houston, TX; 8) USDA/ARS Children's Nutrition Research Center, Houston, TX

---

Very early onset inflammatory bowel disease (VEOIBD) manifests before 6 years of age; early-onset and primary location of inflammation (colon instead of small bowel) makes VEOIBD clinically distinct from adult onset IBD. The frequent occurrence of familial cases implies genetic susceptibility for IBD. As part of the Baylor-Hopkins Center for Mendelian Genomics, we performed WES in a cohort of 25 unrelated probands clinically diagnosed with VEOIBD. We found 6 rare variants (<0.01) in the known IBD gene, *NOD2*, in 6 unrelated probands, including one proband with the known risk allele, *NOD2*-p.R311W. We identified rare, likely loss-of-function (LoF), heterozygous (4 probands) or compound heterozygous (1 proband) variants in *IFIH1* in 5 of 25 probands; in 4 of the 5 families the variant was inherited from an unaffected parent(s): including 3 paternally inherited pedigrees and the proband with compound heterozygous alleles. We also identified 4, rare, heterozygous, functional (3 missense and 1 nonframeshifting) variants in *PLCG2* (4 unrelated probands). Segregation analysis of the *PLCG2* variants in 2 of the 4 families showed the variant to be inherited from the father (one affected and one unaffected). Two probands were double heterozygous for *NOD2* and *PLCG2* rare variants; one of these probands also had a heterozygous, rare (not found in the IBD Exomes Browser), predicted to be deleterious (SIFT, MutationTaster and PolyPhen) *IFIH1* variant (p.Val929Ile). Heterozygous, missense, gain-of-function (GoF) variants in *IFIH1* cause Aicardi-Goutieres syndrome and Singleton-Merten syndrome. Heterozygous LoF variants in *PLCG2* cause familial cold autoinflammatory syndrome and autoinflammation, antibody deficiency, and immune dysregulation syndrome. None of the probands in our cohort had features of syndromes previously associated with *IFIH1* GoF or *PLCG2* LoF variants. An *IFIH1* variant (p.Ile923Val) has previously been shown to account for an IBD GWAS signal, and one patient with autoinflammation, antibody deficiency, and immune dysregulation syndrome has been reported to have ulcerative colitis. Together, these findings implicate heterozygous *IFIH1* (LoF) and *PLCG2* rare variants with incomplete penetrance (possibly caused by maternal imprinting) in the VEOIBD trait and suggest that a gene burden model may contribute to some cases of VEOIBD.



## PgmNr 1950: Understanding the contribution of known cardiovascular-related genes to sudden cardiac death in patients undergoing hemodialysis.

### Authors:

T. Schwantes-An<sup>1</sup>; M. Vatta<sup>1</sup>; M. Abreu<sup>1</sup>; L. Wetherill<sup>1</sup>; H.J. Edenberg<sup>1</sup>; T.M. Foroud<sup>1</sup>; G.M. Chertow<sup>2</sup>; S.M. Moe<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Indiana University School of Medicine, Indianapolis, IN.; 2) Stanford School of Medicine, Stanford, CA.

---

Cardiovascular diseases (CVDs) are the leading causes of death in patients with chronic kidney disease (CKD). In patients requiring dialysis, CVD accounts up to 60% of all deaths. In the general population, atherosclerotic disease (e.g. myocardial infarction) is the primary cause of cardiovascular death, and fewer than 5% of patients have sudden cardiac death (SCD). In contrast, among patients on dialysis, SCD accounts for 25-30% of all CV deaths. The increased frequency of SCD as the cause of cardiovascular fatality in CKD begins prior to the onset of dialysis, indicating that the dialysis procedure is not the only risk. We sought to identify associations between SCD and 174 genes associated with inherited cardiac conditions (ICCs) among patients receiving hemodialysis. Samples are selected from the Evaluation Of Cincacalcet HCl Therapy to Lower Cardiovascular Events (EVOLVE) trial. A clinical events committee adjudicated all participants to determine outcomes. Total of 126 SCD cases (37 African American [AA] and 89 European Ancestry [Eur]) and 107 controls (34 AA and 73 Eur) matched for age, sex, race, dialysis vintage, and diabetes were identified. NextGen sequencing was done using the TruSight Cardio kit (Illumina, San Diego CA). Variants were annotated using ANNOVAR and filtered based on function and gnomAD minor allele frequency (MAF) < 0.05. We conducted burden tests using RVTESTS and RAREMETAL to identify enrichment of variants among cases or controls for SCD. Three distinct models of variant collapsing were tested; 1) individual genes, 2) 17 groups of genes each associated with an ICC (Disease), and 3) 11 groups of genes based on gene function defined by GeneCards (gene function). We identified 4,014 single nucleotide variants (SNVs) in 143 genes. After filtering for predicted function and MAF, 2,109 variants (1,624 in Eur and 1,526 in AA) were retained for analysis. After multiple testing correction, statistical significance threshold was  $p=0.0003$  ( $0.05/143$ ). No statistically significant association between SCD and individual genes was identified. Disease-based burden test did not yield significant associations. Lastly, gene function-based burden test also did not yield any significant associations. In conclusion, we found no statistically significant associations between rare variants among genes associated with inherited cardiac conditions and sudden cardiac death in patients undergoing hemodialysis in this modest sized study.



## PgmNr 1951: Spontaneous *DENND1A* mutations among naturally hyperandrogenic female rhesus macaques associate with multiple phenotypic components of human polycystic ovary syndrome (PCOS).

### Authors:

A. Harris<sup>1</sup>; M. Raveendran<sup>1</sup>; J.E. Levine<sup>2,3</sup>; D.H. Abbott<sup>2,4</sup>; J. Rogers<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, and Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 2) Wisconsin National Primate Research Center, University of Wisconsin, Madison, WI; 3) Dept. of Neuroscience, University of Wisconsin, Madison, WI; 4) Dept. Obstetrics and Gynecology, University of Wisconsin, Madison, WI

Polycystic ovary syndrome (PCOS) is a common endocrine disorder affecting about 15% of reproductive aged women. Patients display hyperandrogenism (elevated testosterone (T) levels), disruption of normal ovulation, ovarian release of follicle regulating antimüllerian hormone (AMH), disruption of ovary-stimulating gonadotropin and increased risk for insulin resistance, type 2 diabetes and obesity. Female fetuses exposed to high T levels *in utero* are at increased risk for PCOS. The heritability of PCOS is estimated to be 70% but the particular genes involved are not well understood. GWAS analyses in women have identified several genes of interest, including *ERBB4*, *FSHB* and *DENND1A*. We investigated PCOS-related traits in rhesus macaques (*Macaca mulatta*) in two ways. Experimental exposure of fetal female rhesus macaques to T produces PCOS-like characteristics including adult hyperandrogenism and ovarian dysfunction with hypergonadotropic function, as well as increased visceral fat, insulin resistance and risk for type 2 diabetes. In addition, we identified a separate cohort of naïve female macaques with elevated circulating T (high T). Naturally occurring high T females exhibit infertility, elevated gonadotropin and AMH, insulin resistance and a trend towards higher BMI.

We performed whole genome sequencing on 20 naïve female macaques with high T levels and 21 controls (normal T). We found a missense mutation that alters the highly conserved clathrin binding domain in *DENND1A* that is associated in high T macaques with elevated AMH levels ( $p=2.6 \times 10^{-3}$ ), a risk factor for PCOS in women. This aspartic acid-to-asparagine mutation in *DENND1A* has CADD score of 21, suggesting probable effect on protein function. In addition, the macaque with the highest AMH level is heterozygous for this mutation and a second missense (threonine-to-methionine) change in *DENND1A* protein. The second mutation is also likely to be functional (CADD score 13.7) and is associated with both AMH ( $p < 1.75 \times 10^{-3}$ ) and insulin levels ( $p = 3.08 \times 10^{-6}$ ) in high T female macaques. Meta-analyses of women with PCOS and controls report significant association between a variant adjacent to *DENND1A* and PCOS, with odds ratio 1.22 (Day et al., 2018). These results suggest that naturally occurring genetic variation among rhesus macaques provides a valuable new model of PCOS. Additional studies increasing sample size and extending genetic analyses are in progress.

# PgmNr 1952: Genome-wide meta-analysis revealed multiple novel loci associated with serum uric acid levels in Japanese individuals.

## Authors:

M. Nakatochi <sup>1</sup>; M. Kanai <sup>2</sup>; A. Nakayama <sup>3</sup>; A. Hishida <sup>4</sup>; Y. Kawamura <sup>3</sup>; M. Nakajima <sup>3</sup>; Y. Kamatani <sup>5</sup>; N. Shinomiya <sup>3</sup>; M. Yokota <sup>6</sup>; K. Wakai <sup>4</sup>; Y. Okada <sup>2,5</sup>; H. Matsuo <sup>3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Data Science Section, Department of Nursing, Nagoya University Graduate School of Medicine, Nagoya, Japan; 2) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan; 3) Department of Integrative Physiology and Bio-Nano Medicine, National Defense Medical College, Tokorozawa, Japan; 4) Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Japan; 5) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) Department of Genome Science, School of Dentistry, Aichi Gakuin University, Nagoya, Japan

---

**Aims:** Gout is a common arthritis caused by elevated serum uric acid (SUA) levels. Here we investigated loci influencing SUA in a genome-wide meta-analysis with 121,745 Japanese subjects and have compared our results with those of previous GWASs and combined them.

**Methods:** We performed a genome-wide meta-analysis based on three Japanese cohorts including those of the Japan Multi-institutional Collaborative Cohort (J-MICC) Study, the Kita-Nagoya Genomic Epidemiology (KING) Study, and the BioBank Japan (BBJ). We also performed the trans-ethnic meta-analysis across the present study and the Global Urate Genetics Consortium (GUGC)-based study to carry out fine-mapping analysis and identify further novel loci associated with SUA. We compared the SNP-based heritability ( $h^2$ ) of SUA in our Japanese meta-analysis and the GUGC-based study. The heritability estimates were calculated from summary statistics of SNPs, which were assessed in both studies and have MAF  $\geq 1\%$  in both studies.

**Results:** We identified 8,948 variants at 36 genomic loci ( $P < 5 \times 10^{-8}$ ) including eight novel loci. Of these, missense variants of *SESN2* and *PNPLA3* were predicted to be damaging to the function of these proteins; another five loci—*TMEM18*, *TM4SF4*, *MXD3-LMAN2*, *PSORS1C1-PSORS1C2*, and *HNF4A*—are related to cell metabolism, proliferation, or oxidative stress; and the remaining locus, *LINC01578*, is unknown. We also identified 132 correlated genes whose expression levels are associated with SUA-increasing alleles. These genes are enriched for the UniProt transport term, suggesting the importance of transport-related genes in SUA regulation. Furthermore, trans-ethnic meta-analysis across our own meta-analysis and the GUGC has revealed 15 more novel loci associated with SUA. The  $h^2$  (standard error) estimates were 14.0 % (4.3%) for our Japanese study and 14.4% (3.9%) for the European study

**Conclusions:** Our findings provide insight into the pathogenesis, treatment, and prevention of hyperuricemia/gout, and they provide a potential basis for the development of new treatments for these diseases.

## PgmNr 1953: ATP-binding cassette (ABC) transporter recessive mutations in biliary atresia cases.

### Authors:

W.Y. Lam <sup>1,2</sup>; J.S. Hsu <sup>3</sup>; C.S. Tang <sup>1,2</sup>; M.T. So <sup>1</sup>; P.H.Y. Chung <sup>1</sup>; P.K.H. Tam <sup>1,2</sup>; M.M. Garcia-Barcelo <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 2) Dr Li Dak-Sum Research Centre, The University of Hong Kong – Karolinska Institutet Collaboration in Regenerative Medicine, The University of Hong Kong, Hong Kong; 3) Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

---

Biliary atresia (BA) is an aggressive neonatal disorder characterized by progressive fibrosclerosing obliteration of the extrahepatic biliary tract in a few weeks after birth. Early surgical intervention by Kasai portoenterostomy to establish bile flow is critical to survival, yet disease progression to liver cirrhosis occurs in many patients, leaving liver transplantation the only option. BA has the highest prevalence of 1/5000 live births in the Southeast Asian population, 3 to 4 times higher than the US and European populations. Pathogenesis of BA remains uncertain. Consistent and robust evidence supporting theories of viral infection, immune dysregulation, environmental and multifactorial causes are yet to be established. Multiple genetic studies in different ethnic groups have identified *ADD3* as a susceptibility gene of BA, yet disease-causal genetic risk factor is still in search. We conducted a whole exome sequencing study on 83 trios from the Southeast Asian population, and identified 148 rare non-synonymous *de novo* and inherited recessive mutations in 135 genes that were predicted to be damaging, and expressed in liver or bile duct tissues according to public databases and in-house BA liver organoid gene expression dataset. KEGG pathway analysis showed that ATP-binding cassette (ABC) transporter was significantly enriched (false discovery rate: 0.042), involving 4 genes (*ABCA7*, *ABCA8*, *ABCC6*, *CFTR*) with recessive variants in 4 subjects. In particular, recessive *ABCA7* variants were carried by 2 BA subjects, one of which had 2 *ABCA7* and a *ABCC6* recessive mutations. In mice, *Abcc6* deficiency causes altered gene expression of other bile salt and lipid transporters. *ABCA7* is involved in lipid transport and plays an important role in lipid homeostasis. *ABCA8* regulates lipid efflux, whereas *CFTR* regulates bile secretion at the apical membrane of biliary epithelial cells, and is associated with cystic fibrosis. Since altered bile acid metabolism can induce cholestatic liver injury, our initial findings warrant further functional studies on the specific functions of these ABC transporters in liver and biliary tissues, and their potential link with BA.

## PgmNr 1954: A fifty-year follow-up of familial Hirschsprung disease.

### Authors:

E. Passarge <sup>1</sup>; AmyN. Ziegler <sup>2</sup>; R.J. Hopkin <sup>3</sup>; H.M. Saal <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institut für Humangenetik, Universitätsklinikum Essen, Germany; 2) Harrison, Ohio, USA; 3) Human Genetics Division, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA

---

We describe familial Hirschsprung disease (OMIM 142623) with histologically proven congenital intestinal aganglionosis in the probanda, her two affected brothers, her two affected daughters and in the extended family three first cousins (two males and one female). Originally this family was described in 1967 as part of systematic study of the genetics of Hirschsprung disease (Family 21 in E. Passarge, New Eng J Med 1967; 276:138-143). It presumably involves the highest number of affected members documented. In all patients the non-syndromic long segment disease (HSCR type 2) is present. At the age of 8 months in 1965 the probanda was admitted to Cincinnati Children's Hospital for ileostomy. In part she was then under the care of one of the authors (EP), at that time a resident in pediatrics. During the following 15 years she required 41 surgical procedures of the GI tract or the rectum. In spite of the complicated course she is well adjusted to her disorder. This family was recently re-investigated, including two daughters of the probanda also affected with long segment Hirschsprung disease. In one daughter a 40-cm aganglionic segment of the colon has been removed, in the other a 26-cm long segment. Both daughters also have required multiple additional surgical procedures. A variant S32L has been detected in the *RET* gene (OMIM 164761) in the mother and the two affected daughters. However, in the light of the complex genetics of Hirschsprung disease this is unlikely to be the only cause. Further molecular studies are under way.

Hirschsprung disease, a genetically heterogeneous and clinically variable disorder, results from the absence or malfunction of intestinal ganglion cells. Functionally variants (mutations) in the *RET* gene and three non-coding *RET* variants (rs2435357, rs2506030, rs7069590) and in at least 13 other genes and five predisposing regions contribute in different ways to the etiology.

At least three signal effector pathways are required for normal migration and development of intramural intestinal ganglion cells, (i) the RET tyrosine-kinase receptor and its ligand GDNF (glial-cell-derived neurotrophic factor, OMIM 600837), (ii) endothelin type B receptor (EDNRB, OMIM 600837) and its ligand EDN3 (endothelin 3, OMIM 131244), (iii) SOX10 transcription factor (OMIM 602229). Owing to reduced penetrance a functionally relevant variant in each of the genes involved is neither necessary nor sufficient to cause intestinal aganglionosis.

## PgmNr 1955: Genome-environment interactions for inflammatory bowel disease in the UK Biobank.

### Authors:

A. Yang; L. Jostins-Dean

[View Session](#) [Add to Schedule](#)

**Affiliation:** Kennedy Institute of Rheumatology, University of Oxford, Oxford, United Kingdom

**Introduction:** Inflammatory bowel disease (IBD) is a relapsing, debilitating condition with rising global prevalence that comprises Crohn's disease (CD) and ulcerative colitis (UC). It results from an inappropriate immune response to gut bacteria, but the causes of this dysbiosis are not well understood. Previous studies have identified 260 genetic variants and ~25 environmental predictors for IBD. While a combination of genetic susceptibility and environmental exposure is thought to lead to IBD, few large gene-environment interaction studies have been performed, so this combined genetic-environmental burden remains poorly characterized. We therefore embarked on a large-scale study of genome-environment (specifically, polygenic risk score (PRS)-environment) interactions that predict IBD.

**Methods:** For this study we used the UK Biobank, a prospective cohort study collecting genome-wide genotyping and phenotype data for ~500,000 participants. Using self-report and electronic health records, we identified 1948 CD and 3720 UC patients after quality control. We used 232 known genetic variants to construct PRSs for CD, UC, and overall IBD, which all associated with their respective diseases ( $p < 1e-16$ ). Based on prior literature and biological plausibility, we tested 27 candidate environmental variables and their interactions with the PRSs for association with CD, UC, or IBD in a logistic regression model.

**Results:** We found 15 environmental variables that associate with IBD, some of which confirm previous studies, others which differ from previous reports or are novel. For example, we replicated associations for red meat consumption (OR for CD=1.05,  $p=7e-4$ ), appendectomy (OR for IBD=0.37,  $p=1e-6$ ), years on hormone replacement therapy (HRT) (OR for CD=1.09 per SD years,  $p=3e-5$ ), dietary fiber (OR for IBD per SD fiber=0.94,  $p=4e-3$ ), and years of cigarette smoking (OR for CD=1.10 per SD years,  $p=7e-3$ ; OR for UC=0.86,  $p=4e-7$ ). Meanwhile, we found a novel association for maternal smoking around birth (OR for IBD=1.10,  $p=9e-4$ ). In addition, we uncovered a new PRS-dietary fiber interaction (OR interaction for IBD=0.94 per SD fiber per SD PRS,  $p=9e-3$ ) and new PRS-HRT interactions (OR interaction for UC=0.94 per SD HRT per SD PRS,  $p=7e-4$ ).

We will also present estimates of gene-environment interactions for individual IBD risk variants, and an estimate of the total additional heritability of IBD risk explainable by gene-environment interactions.

## PgmNr 1956: Exome sequencing analysis of a Mayer-Rokitansky-Kuster-Hauser syndrome cohort.

### Authors:

A. Jolly<sup>1</sup>; Z. Coban Akdemir<sup>1</sup>; S.N. Jhangiani<sup>2</sup>; D.M. Muzny<sup>2</sup>; A. Koch<sup>3</sup>; J.E. Dietrich<sup>4,5,6</sup>; I.B. Van den Veyver<sup>1,4,6</sup>; S. Brucker<sup>3</sup>; R.A. Gibbs<sup>1,2</sup>; J.R. Lupski<sup>1,2,5,6</sup>; N. Wu<sup>1,7,8,9</sup>; J.E. Posey<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Department of Obstetrics and Gynecology, University of Tübingen, Tübingen, GERMANY; 4) Department of Obstetrics and Gynecology, Texas Children's Hospital, Houston, TX; 5) Department of Pediatrics, Texas Children's Hospital, Houston, TX; 6) Texas Children's Hospital, Houston, TX; 7) Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, CHINA; 8) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Beijing, CHINA; 9) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences, Beijing, CHINA

---

Mayer Rokitansky Küster Hauser (MRKH) syndrome represents a disorder of Mullerian development with an incidence of 1/5000 that can include renal aplasia and cervical somite anomalies. While functional models have elucidated numerous candidate genes, only *WNT4* (OMIM 158330) variants have been firmly associated with a phenotype of MRKH with hyperandrogenism. Familial clustering of cases further suggests MRKH has a genetic etiology, and human genetics studies have identified candidate genes and genomic regions that contribute to a proportion of cases. DNA extracted from 71 probands with phenotypic features consistent with MRKH was submitted for exome sequencing (ES). ES data were analyzed for rare, likely deleterious SNVs and CNVs.

Cohort wide analysis for a burden of rare variants within a single gene identified likely damaging variants in *GREB1L* (OMIM 617805), a known disease gene for renal hypoplasia and uterine agenesis, in 5 probands. An additional case including deletion of *GREB1L* was found by CNV analysis. In screening for rare, predicted deleterious variants in 19 candidate genes, potential molecular diagnoses in 6 cases were found involving *PAX8*, *LHX1*, *GATA3*, *HOXA13*, and *TBX6*. CNV analysis revealed an approximately 500kb deletion in 16p11.2, involving *TBX6*, in 3 additional cases. In two cases, likely damaging variants in known disease genes for syndromes with overlapping phenotypic features were found (*TBX3*, *MNX1*).

MRKH has been suggested to be a genetically heterogeneous condition with studies describing candidate gene variants in one or two probands. Copy-number variation of genomic regions, such as 16p11.2, have been described in multiple cases suggesting that these contribute to a substantial portion of heritability in MRKH. In our study, SNV and CNV involving *TBX6* and *GREB1L* are proposed molecular diagnoses for 11/71 (15%) of cases, supporting them as candidate disease genes in MRKH. Furthermore, the finding of 3 known disease genes associated with syndromic features that overlap with MRKH (*GREB1L*, *MNX1*, and *TBX3*) in 7 probands suggests that candidate gene approaches should include genes associated with component phenotypic features of interest (e.g. abnormalities of uterine development) rather than limiting candidate genes to those associated with the exact

manifestation of the studied disease (e.g. MRKH). This approach will enable further elucidation of the full spectrum of clinical presentation associated with variants in known disease genes.



# PgmNr 1957: Meta-analysis of genome-wide association studies reveals 6 loci influencing irritable bowel syndrome.

## Authors:

C. Eijsbouts<sup>1,2</sup>; T. Zheng<sup>3</sup>; N.A. Kennedy<sup>4</sup>; F. Bonfiglio<sup>3</sup>; C.A. Anderson<sup>5</sup>; L. Moutsianas<sup>5,6</sup>; J. Holliday<sup>7</sup>; N. Allen<sup>7</sup>; P.J. Whorwell<sup>8</sup>; R. Spiller<sup>9</sup>; G. McVean<sup>1</sup>; M. D'Amato<sup>10,3</sup>; L. Jostins<sup>11,12</sup>; M. Parkes<sup>13</sup>; The Bellygenes Initiative

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; 2) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Karolinska Institutet, Department of Medicine Solna, Stockholm, Sweden; 4) IBD Pharmacogenetics, College of Medicine and Health, University of Exeter, Exeter, UK; 5) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK; 6) William Harvey Research Institute, Barts & The London School of Medicine & Dentistry, Queen Mary University of London, Charterhouse Square, London, UK; 7) Nuffield Department of Population Health, University of Oxford, Oxford, UK; 8) Neurogastroenterology Unit, Wythenshawe Hospital, Centre for Gastrointestinal Sciences, University of Manchester, UK; 9) Nottingham Digestive Diseases Centre, NIHR Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust and the University of Nottingham, UK; 10) Monash University, School of Biological Sciences, Clayton VIC, Australia; 11) Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK; 12) Christ Church, University of Oxford, Oxford, UK; 13) Division of Gastroenterology, Department of Medicine, University of Cambridge, UK

## Introduction

Despite its high prevalence, the genetic basis of irritable bowel syndrome (IBS) is poorly understood, with only a single known locus passing the genome-wide significance threshold in previous genome-wide association studies (GWAS). The wealth of phenotypical and genomic data contained in the UK Biobank (UKBB) typically allows for more powerful analyses. However, the IBS diagnoses needed for this purpose are recorded using various definitions for different subsets of the population. We combined data covering multiple definitions of IBS in a GWAS meta-analysis to identify novel genetic risk loci.

## Methods

We developed a digestive health questionnaire for the UKBB, asking about a previous diagnosis of IBS and symptoms relevant for a diagnosis via Rome III Criteria. We combined the 174,778 responses with existing UKBB data in the form of self-reported illnesses (at UKBB enrolment) and hospital in-patient data supplied by clinicians (ICD-10 codes). We compared these diagnostic sources by conducting a GWAS for each definition using linear mixed models as implemented in BOLT-LMM, before pooling UKBB cases and controls with independent cohorts to maximize power. Genome-wide significant signals were then included in a replication study by 23andMe.

## Results



IBS diagnoses arrived at by different means showed poor overlap (3.8 - 26.8% pairwise, among fully-examined patients), but nonetheless produced highly consistent, yet not identical, associations (min. pairwise  $r_g=0.70$ ,  $SE=0.06$ ) in individual GWAS. By combining IBS diagnoses across these diagnostic sources in the UKBB (40,548 cases, 360,845 controls) and several other cohorts from the Bellygenes initiative (12,852 cases, 139,981 controls), for a total of 53,400 cases and 433,201 controls, we identified 6 novel associations. These were provisionally mapped to genes involved in colon cell cycle control (*FAM120A*, *PHF2*) and bipolar disorder (*DOCK9*). All signals were subsequently replicated in an independent 23andMe cohort (205,252 self-reported cases and 1,384,055 controls). Genome-wide results (observed  $h^2=0.0247$ , 95% CI [0.022, 0.028], LDSC intercept=0.9951,  $SE=0.007$ ) show considerable overlap with those previously reported for insomnia ( $r_g=0.42$ ,  $SE=0.05$ ), depression ( $r_g=0.53$ ,  $SE=0.05$ ) and neuroticism ( $r_g=0.63$ ,  $SE=0.09$ ).

## Conclusion

We have uncovered novel genetic risk loci for IBS susceptibility that emphasize the potential role of the brain-gut axis in IBS development.

## PgmNr 1958: The genetic architecture of liver cirrhosis: A multi-trait genome-wide association analysis.

### Authors:

C. Emdin<sup>1</sup>; M. Haas<sup>1</sup>; D. Klarin<sup>1</sup>; K. Aragam<sup>1</sup>; M. Chaffin<sup>1</sup>; L. Jiang<sup>2</sup>; W.Q. Wei<sup>2</sup>; Q. Feng<sup>2</sup>; J. Denny<sup>2</sup>; S. Kathiresan<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Brookline Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, 02142, MA.;  
2) Departments of Biomedical Informatics, Vanderbilt University, Vanderbilt, TN 3724

---

Liver cirrhosis is a leading cause of death worldwide and plasma levels of alanine transaminase levels are a marker of liver injury. The genetic architecture of liver cirrhosis is largely unknown. We undertook a multi-trait genome-wide association analysis including alanine transaminase levels (n=382895, UK Biobank), 5770 cases of cirrhosis and 572850 controls from seven cohorts to identify novel genetic determinants of cirrhosis. We identified 12 novel loci at genome-wide significance ( $p < 5 \times 10^{-8}$ ). A variant downstream of *EFNA1*, encoding the ephrin A1 receptor expressed on hepatocytes, associated with reduced cirrhosis risk (OR 0.90,  $1.7 \times 10^{-37}$ ). A missense variant in *APOE* (rs429358, C1310R) associated with protection from cirrhosis (OR 0.84,  $p = 9.2 \times 10^{-38}$ ) but higher risk of coronary artery disease (OR 1.10,  $p = 2 \times 10^{-9}$ ). A common missense variant in *MARC1* (rs2642438, A165T), an enzyme expressed in hepatocytes, associated with protection from cirrhosis (OR 0.91,  $p = 1.2 \times 10^{-44}$ ) and lower total cholesterol levels ( $-0.03$  SD,  $p = 1.9 \times 10^{-36}$ ). An independent rare predicted damaging missense variant in *MARC1* (M187K, 0.9% frequency) also associated with lower total cholesterol levels ( $-0.05$  SD,  $p = 2 \times 10^{-8}$ ) and protection from cirrhosis (OR 0.75,  $p = 0.01$ ). We developed a polygenic score for cirrhosis using 16 independent genome-wide significant variants. Individuals in the top 1% of the polygenic score were at similar risk of cirrhosis (OR 2.80,  $p = 6 \times 10^{-14}$ ) as homozygous carriers of known variants causing hereditary hemochromatosis (OR 3.24,  $p < 2 \times 10^{-18}$ ) or alpha-1-antitrypsin disease (OR 3.36,  $p = 0.06$ ) and had a five-fold increased risk of death from liver disease (OR 5.02,  $p = 4.5 \times 10^{-9}$ ). These findings identify novel therapeutic targets for the prevention and treatment of cirrhosis and suggest that a polygenic score could be used to identify individuals predisposed to liver cirrhosis.

## PgmNr 1959: Large-scale genome-wide association study identifies candidate genes for childhood steroid-sensitive nephrotic syndrome.

### Authors:

Jia <sup>1</sup>; T. Yamamura <sup>2</sup>; Y. Hitomi <sup>1</sup>; Y. Kawai <sup>1</sup>; S. Khor <sup>1</sup>; K. Ueno <sup>1</sup>; T. Horinouchi <sup>2</sup>; K. Nozu <sup>2</sup>; C. Nagano <sup>2</sup>; M. Nagasaki <sup>3</sup>; K. Tokunaga <sup>1</sup>; K. Iijima <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan.; 2) Department of Pediatrics, Graduate School of Medicine, Kobe University, Kobe, Japan.; 3) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan.

---

Idiopathic nephrotic syndrome (INS) is the most common cause of kidney disease in children, about 80% of pediatric patients respond to steroid treatment and are classified as steroid-sensitive NS (SSNS). Previous genome-wide association studies (GWASs) have identified *HLA-DR/DQ* as the predominant genetic risk factors of childhood SSNS in multiple populations. Our group reported the first GWAS for childhood SSNS in Japanese, genome-wide significant associations were identified in *HLA-DR/DQ* region (*J Am Soc Nephrol*, 2018).

To identify more loci contributing to susceptibility to childhood SSNS, we performed an extended GWAS with the largest sample size to date. In the discovery stage, 1,018 patients with childhood-onset SSNS and 3,331 adult healthy controls with Japanese ancestry were genotyped using Affymetrix 'Japonica Array'. Whole-genome imputation was conducted using a phased reference panel of 2,036 healthy Japanese individuals (2KJPN panel). Quality control (QC) was performed to exclude the samples with low calling rate (<97%) and the variants with low genotyping rate (<97%), minor allele frequency (MAF) <0.5% and Hardy-Weinberg equilibrium (HWE) test P-value <1×10<sup>-5</sup>. Association analysis was conducted using logistic regression with the adjustment of gender and principal components. In discovery stage, the most significant association was detected in *HLA-DR/DQ* region (P=2.80×10<sup>-33</sup>, odds ratio (OR)=2.49). HLA fine-mapping was done by imputing classical HLA genes (*HLA-A*, *-C*, *-B*, *-DRB1*, *-DQB1*, *-DPA1* and *-DPB1*), 870 cases and 2,903 controls passed the post-imputation QC (call threshold>0.4). *HLA-DRB1\*08:02-DQB1\*03:02* was identified as the most significantly susceptible haplotype (P-value after multiple corrections [Pc]=1.16×10<sup>-22</sup>, OR=3.38), while *HLA-DRB1\*13:02-DQB1\*06:04* was the most significant protective haplotype (Pc=1.63×10<sup>-16</sup>, OR=0.18). Heterozygous patients with *HLA-DRB1\*08:02-DQB1\*03:02* and *HLA-DRB1\*13:02-DQB1\*06:04* present at lower frequency (1/870=0.11%) than in healthy individuals (15/2903=0.52%), although the difference is not significant (P=0.14), suggesting a dominant effect of the protective haplotype over the susceptible haplotype. Furthermore, two regions on chromosome 19 and 9 achieved genome-wide significance (P=4.94×10<sup>-20</sup>, OR=1.90; P=2.54×10<sup>-8</sup>, OR=0.72). Replication studies were successfully performed in other Asian sample sets.

# PgmNr 1960: Genetic variants associated with serum levels of alanine aminotransferase among hepatitis C virus-infected patients: A genome-wide association study.

## Authors:

P.C. Liu <sup>1,2</sup>; Y.H. Huang <sup>1</sup>; Y.J. Lin <sup>1</sup>; C.L. Jen <sup>3</sup>; S.N. Lu <sup>4</sup>; C.J. Chen <sup>5</sup>; H.I. Yang <sup>3</sup>; M.H. Lee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan, Taipei, Taiwan; 2) Faculty of Medicine, National Yang-Ming University, Taipei, Taiwan; 3) Genomics Research Center, Academia Sinica, Taipei, Taiwan; 4) Department of Gastroenterology, Chang-Gung Memorial Hospital, Kaohsiung, Taiwan; 5) Academia Sinica, Taipei, Taiwan

## Background

Serum alanine aminotransferase (ALT) level is an inflammatory marker for patients with hepatitis C virus (HCV) infection. Patients with elevated ALT levels had increased risk for hepatocellular carcinoma. We performed a genome-wide association study to explore single nucleotide polymorphisms (SNPs) associated with serum ALT levels and validated the findings in an external set.

## Methods

In genome-wide discovery phase we included 892 anti-HCV seropositives seronegative for hepatitis B surface antigen. The serum ALT levels were measured at baseline and during the follow-up period from 1991 to 2005. All of the study participants were free of hepatocellular carcinoma. They were examined by the Axiom™ Genome-Wide CHB SNP array. SNPs with call-rate < 97%, minor allele frequency < 5%, and deviating from Hardy-Weinberg equilibrium ( $p < 10^{-6}$ ) were excluded. Baseline serum ALT level was treated as a quantitative trait in the analyses. Additive models were used to estimate the effect of the SNPs. Internal validation was performed on repeated serum ALT measurements of the same study participants to evaluate whether these SNPs were associated with the long-term changing patterns of ALT levels. In addition, external validation was performed on 496 study participants with the same criteria in Taiwan Biobank. Logistic regressions were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) of the SNPs associated with serum ALT levels ( $> 45$  vs.  $\leq 45$  U/L).

## Results

Among 806 study participants passing quality control, 357 (44.3%) were with serum ALT levels  $\leq 15$  U/L, 319 (39.6%) (15, 45] U/L, and 130 (16.1%)  $> 45$  U/L. Among 564,464 SNPs included in the analyses, ten candidate SNPs had  $p$  value  $< 10^{-5}$ . Based on serial measurements of serum ALT levels, there were 158 (19.6%) had ALT persistently  $\leq 15$  U/L, 327 (40.6%) ever  $> 15$  U/L but never  $> 45$ , and 321 (39.8%) ever  $> 45$  U/L. There were six SNPs located in chromosome 9, 11, 13, 14 and 15 were significantly associated with serum ALT levels ever  $> 45$  U/L, showing consistent associations with elevated serum ALT levels. Finally, we found rs12376677 [OR: 1.75 (1.22-2.51)] and rs568800 [OR: 1.32 (1.06-1.64)] were associated with elevated serum ALT levels among individuals with HCV infection in the external validation set.

## Conclusions

Genetic variants were found to be associated with elevated serum ALT levels. The potential mechanisms for these SNPs need further investigations.

## PgmNr 1961: Identification of susceptibility loci for reproductive traits in Samoan women.

### Authors:

Z. Erdogan-Yildirim <sup>1</sup>; J.C. Carlson <sup>2</sup>; A. Pomer <sup>3</sup>; N.L. Hawley <sup>3</sup>; H. Cheng <sup>4</sup>; T. Naseri <sup>5</sup>; S.M. Reupena <sup>6</sup>; R. Deka <sup>4</sup>; G. Lambert-Messerlian <sup>7</sup>; S.T. McGarvey <sup>8</sup>; D.E. Weeks <sup>1</sup>; R.L. Minster <sup>1</sup>; TOPMed Reproductive Health Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.; 2) Dept. of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.; 3) Dept. of Chronic Disease Epidemiology, School of Public Health, Yale University, New Haven, CT; 4) College of Medicine, University of Cincinnati, Cincinnati, OH; 5) Ministry of Health, Government of Samoa, Apia, Samoa; 6) Lutia i Puava ae Mapu i Fagalele, Apia, Samoa; 7) Dept. of Pathology and Laboratory Medicine and Dept. of Obstetrics and Gynecology, Women & Infants Hospital of Rhode Island, Brown University, Providence, RI; 8) Dept. of Epidemiology and International Health Inst., Public Health, Brown University, Providence, RI

---

Obesity is a global public health problem with severe effects on not only cardiovascular health but also reproductive health in women. Menstrual irregularities and a wide spectrum of reproductive endocrine disorders are associated with higher body mass index (BMI). The high prevalence of female obesity (66.3%) in Samoa and the population's genetic isolation make it an ideal population for discovery of susceptibility genes involved in female reproductive phenotypes.

To detect genes and variants associated with reproductive traits in Samoan women in the Soifua Manuia Study, we performed GWAS with genotyped and imputed variants (via a Samoan-specific reference panel) for levels of biomarkers related to ovarian health and a type of infertility called polycystic ovary syndrome (PCOS) (1) anti-Müllerian hormone (AMH,  $n=831$ ), (2) total testosterone (TT,  $n=790$ ) and (3) sex hormone binding globulin (SHBG,  $n=995$ ). Serum TT and SHBG levels were assayed with a chemiluminescent automated immunoassay, and AMH with an ultrasensitive ELISA assay. Association of SNPs with  $MAF \geq 0.01$  was assessed using linear mixed model adjusted for age, age<sup>2</sup>, principal components of ancestry, and empirical kinship.

For testosterone, we observe genome-wide significant ( $p < 5 \times 10^{-8}$ ) SNPs on 9q34.2 in the intergenic region between *LCN1P1* and *ABO* with the highest peak at rs9411365 ( $p = 3.75 \times 10^{-9}$ ). *LCN1P1*, lipocalin 1 pseudogene 1, is expressed in ovaries and may bind testosterone as does lipocalin 1. For SHBG, the top 2 SNPs are in 6q23.3 (rs4270801,  $p = 1.95 \times 10^{-7}$ ) and in 12q23.3 downstream of *CASC18* (rs17037733,  $p = 5.18 \times 10^{-7}$ ). rs4270801 is located downstream of *MYB* 5kb distal of *MIR548A2* and has potential regulatory function (RegulomeDB score=4). For AMH, the 2 top SNPs are in 14q24.3 (rs7156476,  $p = 1.75 \times 10^{-6}$ ) and in 9q31.1 (rs1008300304,  $p = 1.93 \times 10^{-6}$ ). rs7156476 (Samoan MAF=0.13, non-Samoan MAF=0.48) is located proximally to *FOS* which acts as a gonadotropin-releasing hormone repressor. rs1008300304 (Samoan MAF=0.017, non-Samoan MAF=0.00016) is in intron 3 of *SLC44A1* and is 1Mbp proximal to variants associated with age at menarche in an earlier candidate gene study. In total, 10 loci for TT, 18 loci for SHBG and 9 loci for

AMH were suggestively significant ( $p < 10^{-5}$ ).

Additional models that adjust for covariates such as BMI, smoking and menstrual status are necessary to refine these findings. Subsequent replication in Samoans and other Polynesians is needed to validate these findings.

## PgmNr 1962: Whole exome sequencing of congenital isolated hypogonadotropic hypogonadism reveals known and novel candidate genes.

### Authors:

S. Tang<sup>1,2,3</sup>; F. Zhang<sup>1,2,3</sup>; J. Zhang<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Obstetrics and Gynecology Hospital, State Key Laboratory of Genetic Engineering at School of Life Sciences, Institute of Reproduction and Development, Fudan University, Shanghai 200011, China; 2) Key Laboratory of Reproduction Regulation of NPFPC, Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200032, China; 3) Shanghai Key Laboratory of Female Reproductive Endocrine Related Diseases, Shanghai 200011, China

---

Congenital isolated hypogonadotropic hypogonadism (IHH) is a highly heterogeneous disorder, both clinically and genetically, which leads to delayed puberty and infertility due to developmental abnormalities of the reproductive system. Dozens of genes were found responsible for the pathogenesis of IHH to date. However, due to its incomplete penetrance and complexity, the molecular basis of IHH and its contribution remain elusive, and genome-wide systematic analysis of patients with IHH is still lacking. Here, using whole exome sequencing (WES), 18 unrelated males with IHH and their normal parents were enrolled for screening known genes related to IHH and analyzing novel candidate genes with 2500 in-house healthy control individuals. Ten coding variants of 7 known IHH genes (*FGFR1*, *PROKR2*, *CHD7*, *ANOS1*, *PNPLA6*, *POLR3A*) and eighteen variants of 18 novel candidate genes are identified in 17/18 (94.4%) patients. All of variants we identified are heterozygous. The novel candidate genes we found in this study mainly participate in development of olfactory bulb, migration of GnRH neuron and regulation of key signal pathway in forebrain. In addition, half of cases (9/17, 52.9%) harbor at least two mutations in distinct genes. Interestingly, the heterozygous mutation, *PROKR2* W178S, previously reported several times, is recurrent in three patients in our cohort along with other mutations of different genes. In all, the results of our study provided a mutational landscape for patients with IHH and strongly suggested that IHH is in favor of polygenic model.

## PgmNr 1963: Haptoglobin-2 variant increases susceptibility to acute lung injury during sepsis.

### Authors:

V. Kerchberger<sup>1,2</sup>; J.A. Bastarache<sup>1,3</sup>; C.M. Shaver<sup>1</sup>; H. Nagata<sup>4</sup>; J.B. McNeil<sup>1</sup>; S.R. Landstreet<sup>1</sup>; N.D. Putz<sup>1</sup>; W.K. Yu<sup>1,5</sup>; J. Jordan<sup>1</sup>; N.E. Wickersham<sup>1</sup>; D.R. Janz<sup>6</sup>; C.R. Parikh<sup>7</sup>; E.D. Siew<sup>8</sup>; L.B. Ware<sup>1,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Allergy, Pulmonary, and Critical Care, Vanderbilt University Medical Center, Nashville, TN.; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA; 3) Department of Cell and Developmental Biology, Vanderbilt University, Nashville, Tennessee, USA; 4) Department of Anesthesiology, Keio University School of Medicine, Tokyo, Japan; 5) Department of Chest Medicine, Taipei Veterans General Hospital, Taipei, Taiwan, School of Medicine, National Yang-Ming University, Taipei, Taiwan; 6) Section of Pulmonary and Critical Care Medicine, Louisiana State University School of Medicine, New Orleans, Louisiana, USA; 7) Division of Nephrology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA.; 8) Division of Nephrology and Hypertension, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA; 9) Department of Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville Tennessee, USA

### Background

Haptoglobin (*HP*) is the primary endogenous scavenger for cell free hemoglobin (CFH) released during hemolysis. CFH is present in 80% of septic ICU patients. Higher CFH levels correlate with worse organ dysfunction and mortality. A common *HP* genetic variant *HP2* cannot scavenge CFH as effectively as wild type *HP1*. Using a murine sepsis model and a cohort of septic ICU patients, we tested the hypothesis that individuals with *HP2* would have increased lung inflammation and injury, and increased risk for the acute respiratory distress syndrome (ARDS).

### Methods

In transgenic littermate mice homozygous for murine homologues of human *HP1* and *HP2*, we modeled polymicrobial sepsis with elevated CFH by injection of intraperitoneal cecal slurry and intravenous CFH. We measured survival at 72 hours, and lung inflammatory markers myeloperoxidase (MPO) activity and CXCL1 levels from bronchoalveolar lavage (BAL) fluid. We quantified lung microvascular permeability in excised whole lungs by Angiosense<sup>680</sup> NIR imaging and lung endothelial apoptosis by TUNEL assay. In a cohort of septic critically ill adults, we genotyped *HP* in 344 patients by real-time PCR and imputed *HP* genotype from Illumina microarray data in an additional 152 patients. We defined ARDS cases by the Berlin Criteria for at least one day during the first four ICU days. We used logistic regression to measure *HP* genotype's association with ARDS, controlling for clinical risk factors.

### Results

In the murine sepsis model, *HP2* homozygotes (*HP2-2*) had decreased survival at 72 hours ( $P = 0.03$ ) and increased lung inflammatory markers: both MPO activity ( $P=0.014$ ) and BAL CXCL1 ( $P=0.011$ ). *HP2-2* mouse lungs had worse vascular permeability ( $P=0.037$ ) and increased endothelial apoptosis



( $P=0.004$ ). In 496 septic ICU patients, 76 (15.3%) were *HP1-1*, 196 (39.5%) were *HP2-2*, and 224 (45.2%) were heterozygotes. 185 (35.8%) developed ARDS. Each *HP2* allele increased ARDS risk by  $OR=1.41$  (95% CI: 1.06-1.88,  $P=0.018$ ) when controlling for common clinical risk factors. Notably, *HP2* affected ARDS risk only in patients with elevated CFH levels ( $P=0.026$ ,  $N=411$  for elevated CFH vs  $P=0.46$ ,  $N=81$  for no CFH).

### *Conclusions*

Septic *HP2-2* mice develop more severe acute lung injury compared to wild type *HP1-1*, and humans with *HP2* have increased ARDS risk during sepsis when CFH levels are elevated. Genotyping the *HP2* variant may have use in guiding risk-stratification and development of novel therapies targeting CFH as an injury mediator in ARDS.

## PgmNr 1964: Whole genome sequence analysis and follow-up of pulmonary function and COPD in >19,000 multi-ethnic participants of the NHLBI Trans-Omics for Precision Medicine (TOPMed) program.

### Authors:

W. Kim <sup>1,19</sup>; X. Zhao <sup>2,20</sup>; D. Qiao <sup>1</sup>; C. Yang <sup>3</sup>; S. Kasela <sup>4,5</sup>; N. Shrine <sup>6</sup>; T. Sofer <sup>7,8</sup>; J. N Nguyen <sup>3</sup>; E. C Oelsner <sup>9</sup>; P. Balte <sup>9</sup>; B. Yu <sup>10</sup>; L. Lange <sup>11</sup>; S. J. London <sup>12</sup>; T. Lappalainen <sup>4,5</sup>; J. Dupuis <sup>13</sup>; G. O'Connor <sup>14</sup>; R. Reed <sup>15</sup>; B. Cade <sup>7,8</sup>; S. A. Gharib <sup>16</sup>; M. Daya <sup>11</sup>; S. Redline <sup>7,8,17</sup>; B. M. Psaty <sup>18</sup>; K. Barnes <sup>11</sup>; S. S Rich <sup>3</sup>; G. R. Abecasis <sup>2</sup>; E. K Silverman <sup>1</sup>; R. Graham Barr <sup>8</sup>; M. H. Cho <sup>1,21</sup>; A. Manichaikul <sup>3,21</sup>; On behalf of the TOPMed PFT, COPD and Lung Working Groups

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Channing division of network medicine, Brigham and Women's Hospital, Boston, Massachusetts.; 2) Center for Statistical Genetics, and Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA; 3) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; 4) New York Genome Center, New York, NY, USA; 5) Department of Systems Biology, Columbia University, New York, NY, USA; 6) Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, United Kingdom; 7) Department of Medicine, Harvard Medical School, Boston, Massachusetts. USA; 8) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, Massachusetts, USA; 9) Department of Medicine and Epidemiology, Columbia University Medical Center, New York, New York, USA; 10) Department of Epidemiology, Human Genetics & Environmental Sciences, UTHealth School of Public Health, Houston, Texas, USA; 11) Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, University of Colorado School of Medicine Anschutz Medical Campus, Aurora, Colorado, USA; 12) Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Durham, North Carolina, USA; 13) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; 14) Boston University School of Medicine, Pulmonary Center, Boston, Massachusetts, USA; 15) Division of Pulmonary and Critical Care Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA; 16) Division of Pulmonary, Critical Care and Sleep Medicine, University of Washington, Seattle, Washington, USA; 17) Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA; 18) Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, Washington, USA; Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA; 19) This author is the presenting author; 20) This author is the first author; 21) These authors jointly supervised this work

---

Chronic obstructive pulmonary disease (COPD), diagnosed by reduced lung function, is a leading cause of morbidity and mortality. We performed whole genome sequence (WGS) analysis of lung function and COPD in a multi-ethnic sample of 10,678 participants from population- and family-based studies, and 8,499 individuals from COPD-ascertained studies in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. We recapitulated 11 known GWAS loci and identified 21 distinct novel associations, including four common variants from stratified analysis of African Americans. Two novel

common variants showed evidence of replication in the UK Biobank consisting of 321,047 White British samples (rs74469188: nearest gene = *CMIP*,  $P_{\text{discovery}}=2.3\times10^{-8}$ ,  $P_{\text{replicate}}=1.4\times10^{-5}$ ; rs7188378: nearest gene = *FTO*,  $P_{\text{discovery}}=3.4\times10^{-8}$ ,  $P_{\text{replicate}}=2.0\times10^{-4}$ ). For rs5953026 identified for FVC in TOPMed White samples, it had a significant association with increased peak expiratory flow in UK Biobank (nearest gene = *ZNF157*,  $P_{\text{discovery}}=3.3\times10^{-8}$ ,  $P_{\text{replicate}}=0.0018$ ). In the gene-based burden test using rare putative loss of function (pLoF) variant, *MMP25* and *ARHGEF17* were significantly associated with FVC and FEV1/FVC, respectively (*MMP25*: 15 rare pLoF variants,  $P_{\text{burden}}=9.2\times10^{-7}$ ; *ARHGEF17*: 5 rare pLoF variants,  $P_{\text{burden}}=7.7\times10^{-5}$ ). Phenome-wide analysis in the UK Biobank demonstrated associations with respiratory failure and pulmonary hypertension for two low frequency variants (rs371740347: nearest gene = *CFHR5*; rs114353081: nearest gene = *ERCC8*), while FVC associated variants in Xp11.3 were colocalized with cis-eQTL from GTEx. Our study demonstrates the advantages of deep sequencing in heterogeneous, multi-ethnic samples for identification of genetic associations with common, infrequent and X chromosome variants.

## PgmNr 1965: Predicted gene expression analyses reveal sex-specific pathways for COPD.

### Authors:

V.L. Martucci <sup>1,2</sup>; S. Zhao <sup>3</sup>; J. Hellwege <sup>1,4</sup>; D.R. Velez Edwards <sup>1,5,6</sup>; M.C. Aldrich <sup>1,2,5,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN; 4) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 5) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 6) Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN; 7) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN

---

**Introduction.** Men and women with chronic obstructive pulmonary disease (COPD) present differently with regards to symptom severity, disease outcome, and comorbidities. Women typically develop the disease with lower smoking exposure and experience more airway obstruction and less alveolar destruction than men. The mechanism underlying these differences is unknown, but may be related to differences in genetic susceptibility or nicotine metabolism. Predicted gene expression approaches provide an approach for linking genetic variants to biologic mechanism.

**Objectives.** Our goal was to systematically investigate the effect of genetic variation on predicted gene expression in males and females with COPD.

**Methods.** We used S-PrediXcan, a novel tool for imputing gene expression based on information gathered by the Genotype-Tissue Expression (GTEx) project. We ran S-PrediXcan in 44 tissues using previously published COPD GWAS results for men (N=167,020) and women (N=194,174) with COPD diagnostic codes in the UK Biobank. We then used WebGestaltR to identify functional categories with gene enrichment in both Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO).

**Results.** We identified 138 genes significantly associated with COPD in men and 130 in women ( $p \leq 0.001$ ). Of these genes, only 12 overlapped between sexes. No genes were significantly associated after multiple testing correction. To understand the biological process affected by these genes, we ran enrichment analysis using the top 1,000 genes by p-value in men and women. Enrichment analyses identified three GO categories and 24 KEGG pathways in men and four GO categories and zero KEGG pathways in women that were significantly enriched. The most significantly enriched GO categories in men were all associated with immune response signaling, while in women, all were associated with glycoprotein metabolism and glycosylation.

**Conclusion.** These findings suggest that the biological mechanisms underlying COPD may differ between men and women, which could have important implications for COPD prevention, risk stratification, and treatment and advance personalized medicine for individuals with COPD.

## PgmNr 1966: RNA sequencing analysis of lymphoblastoid cell lines in an African-American Crohn's disease cohort.

### Authors:

R.Y. Cordero <sup>1</sup>; L.W. Datta <sup>2</sup>; S.R. Brant <sup>3</sup>; C.L. Simpson <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, Genomics, and Informatics, University of Tennessee Health Science Center, Memphis, Tennessee.; 2) Department of Medicine, Meyerhoff Inflammatory Bowel Disease Center Johns Hopkins University School of Medicine Baltimore, MD; 3) Department of Medicine, Robert Wood Johnson Medical School, Rutgers University, New Brunswick , NJ; 4) Department of Ophthalmology, University of Tennessee Health Science Center, Memphis, TN

---

### Introduction

Crohn's disease, a subtype of inflammatory bowel disease (IBD), is an immune-mediated disorder that involves any part of the gastrointestinal tract. In recent years, there has been a steady increase in reported cases of Crohn's disease in African-Americans. In this study, we have carried out RNA-Seq analysis to compare the gene expression profiles of lymphoblastoid cell lines (LCL) from patients with Crohn's disease and healthy controls from African American ancestry.

### Methods

RNA sequencing was carried out on LCLs from 60 Crohn's disease patients and 60 healthy controls with a median read length of 100bp. RNA-Seq library preparation was performed with the Illumina® TruSeq Stranded mRNA Library Prep kit. Paired-end sequencing was performed on the Illumina® HiSeq 2500 instrument with a 2x100 bp setup in Rapid mode. Quality control checks on raw sequence data were done using FastQC. Low-quality reads and adapters were trimmed using Trimmomatic. Reads were mapped to the human genome (GRCh37) using STAR aligner. Differential expression analysis was done using Cuffdiff and edgeR.

### Results

The top genes that were upregulated included immunoglobulin genes, *IGHV3-66*, *IGKV3D-7*, *IGKV1D-13*, *IGKV1-33*. The downregulated genes that were identified are *TRBV- T Cell Receptor Beta Variable 27* - which has a role in antigen recognition, *IGHV3-73* which is an Immunoglobulin gene and *MT3*, part of the Metallothioneins that have been shown to have altered expression in inflammatory bowel disease in certain populations. Expression quantitative trait loci (eQTL) studies, comprehensive pathway and alternative splicing analysis are ongoing and will be presented.

### Conclusion

In this study, we harnessed the power of transcriptomic profiling to detect genes of interest in Crohn's

disease. Identifying the dysregulated genes will contribute to the understanding of the molecular pathways involved in this chronic condition. This could potentially lead to novel therapeutic targets, biomarkers of disease severity and prognosis in this disease.

## PgmNr 1967: Association of protein tyrosine phosphatase non-receptor type 22 gene functional variant (C1858T) with genetic susceptibility of psoriatic arthritis in Kuwaiti Arabs.

### Authors:

M. Haider <sup>1</sup>; A.M. Al-Awadhi <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Kuwait University, Faculty of Medicine, Safat, Jabriya, Kuwait; 2) Department of Medicine, Faculty of Medicine, Kuwait University, Jabriya, Kuwait; 3) Rheumatic Diseases Unit, Amiri Hospital, Dasman, Kuwait

---

Psoriatic arthritis (PsA) is a chronic, systemic inflammatory arthritic disease characterized by joint inflammation that is associated with cutaneous psoriasis, and can lead to pain, swelling or stiffness in one or more joints. Epidemiological evidence has shown a higher heritability for psoriatic arthritis compared with psoriasis vulgaris. It has been considered to result from a complex interplay between genetic, immunologic and environmental factors. A single nucleotide polymorphism [C1858T] in the protein tyrosine phosphatase (*PTPN22*) gene which encoded Arg620Trp in the lymphoid protein tyrosine phosphatase (LYP) has shown to be a negative regulator of T-cell activation. Recent evidence suggests that it shows an association with different autoimmune disorders. This study investigated an association between *PTPN22* gene [C1858T] functional variant genotypes and susceptibility of psoriatic arthritis in Kuwaiti patients. We have investigated the association of *PTPN22* gene functional variant [C1858T] in 105 Kuwaiti patients with psoriatic arthritis and compared it to that in 214 healthy controls. Genotypes for *PTPN22* gene [C1858T] variant were determined by using a PCR-RFLP method. The frequency of homozygous variant genotype (TT) was found to be significantly higher in PsA patients compared to that in the controls (OR 19.7,  $p < 0.0001$ ). Collectively, the variant genotype was detected in homozygous and heterozygous combinations in 30% patients ( $p < 0.0001$ ) compared to 16% in the controls. The variant genotype was associated more strongly with PsA patients of age (25-34y). No correlation was detected between the variant genotype and gender of the PsA patients. Our data shows an association of *PTPN22* gene functional variant [C1858T] with psoriatic arthritis in Kuwaiti patients and highlights its role in determining the genetic susceptibility along with other factors.

## PgmNr 1968: A machine learning classifier trained on zebrafish periderm enhancers prioritizes orofacial-cleft-associated variants near *KRT18* for functional tests.

### Authors:

R. Cornell <sup>1</sup>; H. Liu <sup>2</sup>; K. Duncan <sup>1</sup>; A. Helverson <sup>1,5</sup>; P. Kumari <sup>1</sup>; E. Leslie <sup>3</sup>; J. Carlson <sup>4</sup>; P. Breheny <sup>5</sup>; A. Erives <sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Anatomy and Cell Biology, Univ Iowa, Iowa City, Iowa.; 2) Department of Periodontology, School of Stomatology, Wuhan University, Wuhan, China; 3) Department of Human Genetics, Emory University School of Medicine; 4) Department of Biostatistics, University of Pittsburgh; 5) Department of Biostatistics, University of Iowa, Iowa City, Iowa; 6) Department of Biology, University of Iowa, Iowa City, Iowa

---

Genome wide association studies have identified 40 loci associated with risk for orofacial cleft. A current challenge in the field is to distinguish functional single nucleotide polymorphisms (SNPs) from those merely in linkage disequilibrium with them. In many of the loci mentioned above the candidate risk gene is expressed in palate epithelium. Identifying the functional subset of SNPs at these loci may be facilitated by an improved understanding of the sequence constraints of enhancers active in oral epithelium, including in the superficial or periderm layer of it. As models for this cell type, we deployed zebrafish periderm, mouse embryonic palate epithelium, and a human oral keratinocyte cell line. We carried out ATAC-seq on each of these epithelial cell types and on a complementary mesenchymal tissue; ATAC-seq peaks present in the former and not the latter, and positive for the H3K27Ac histone mark in relevant published datasets, were candidates for epithelium-specific enhancers. We trained machine-learning classifiers on such candidates, and used the classifiers to rank orofacial cleft-associated SNPs near the *KRT18* gene, which is expressed in oral periderm. All three classifiers identified rs2070875 to be the SNP with the strongest effect on the score of the element harboring it; the significance of this finding was strongest with the classifier trained on zebrafish periderm enhancers. Allele-specific reporter assays in an oral epithelium cell line confirm that rs2070875 alters an epithelial enhancer. This study provides tools for prioritizing orofacial cleft associated SNPs near genes expressed in oral epithelium for functional tests.



## PgmNr 1969: Genetics of osteoarthritis in the Million Veteran Program.

### Authors:

M.N. McDonald<sup>1,2</sup>; P. Lakshman Kumar<sup>1,2</sup>; V. Srinivasasainagendra<sup>1,2</sup>; R.A. Dennis<sup>3</sup>; V.S. Jagadale<sup>3</sup>; C.J. Brown<sup>1,2</sup>; H.K. Tiwari<sup>1,2</sup>; J.A. Singh<sup>1,2</sup>; J.S. Richman<sup>1,2</sup>; M. Bamman<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Alabama at Birmingham, Birmingham, AL, USA; 2) Birmingham Veterans Affairs Medical Center, Birmingham, AL, USA; 3) Central Arkansas Veterans Healthcare System, Little Rock, AR, USA

---

The societal and patient-centered impacts of end-stage osteoarthritis (OA) among United States Veterans are profound. OA has a strong genetic component with heritability estimates >30%. Previous genome-wide association studies (GWAS) have been performed primarily in populations of European descent. We sought to leverage the rich resource of the Million Veteran Program (MVP) to test the hypothesis that genetic variants explain a meaningful proportion of OA prevalence in a large, ethnically diverse cohort. Genotype variants were typed using the Affymetrix Axiom Biobank Array (MVP Release 3) and imputed results were generated using the 1000 Genomes reference panel. In an effort to replicate known findings and identify new genetic variants and regions associated with OA, we analyzed a total of 338,368 MVP participants between 40-80 years of age, of whom 185,169 carry an OA diagnosis (ICD9 715.x or ICD0 M15-M16). Analyses were stratified by major ancestry group: European (n=251,665), African American (n=57,645), Hispanic (n=19,956) and Asian (n=2,354). After quality control steps, each SNP was tested for association with OA under an additive model adjusted for age, sex and population substructure using PLINK2. Results from the stratified analyses were combined in a meta-analysis using METAL. Similar to previously published OA GWAS among our top significant findings are known BMI variants. Future directions include evaluating whether significant findings replicate in the UK Biobank as well as investigating association of genetic variants with end-stage joint replacement.

Funding was provided from the Department of Veterans Affairs Office of Research and Development, Million Veteran Program Grant I01 RX002745-01.

## PgmNr 1970: Gut virome and bone mineral density in postmenopausal women.

### Authors:

Xinrui. Wu <sup>1</sup>; Zunsong. Hu <sup>2</sup>; X.U. Lin <sup>3</sup>; Zifan. Zhu <sup>4</sup>; J. Greenbaum <sup>5</sup>; Fengzhu. Sun <sup>4</sup>; Hui. Shen <sup>5</sup>; Q.I. Zhao <sup>2</sup>; Jie. Shen <sup>3</sup>; Hongmei. Xiao <sup>6</sup>; Hongwen. Deng <sup>5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology and Health Statistics, Xiangya School of Public Health, Central South University, Changsha, China; 2) Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN, USA; 3) Department of Endocrinology and Metabolism, The Third Affiliated Hospital of Southern Medical University, Guangzhou, China; 4) Quantitative and Computational Biology Program, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA; 5) Center for Bioinformatics and Genomics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA; 6) School of Basic Medical Sciences, Central South University, Changsha, China

---

Osteoporosis is a metabolic bone disease characterized by reduced bone mineral density (BMD), increased risk of fractures which can cause long term disability and mortality. Animal studies have highlighted the important role of the gut virome in human health. However, no studies have investigated the potential effect of gut virome on bone health. In this study, we included 499 Chinese postmenopausal women aged 41-65 years. The dual-energy X-ray absorptiometry was used to measure BMD in different body sites. Stool samples were collected and analyzed using the HiSeq 4000 platform. Viral sequences were classified by taxonomic alignment against NCBI reference genome database using Centrifuge software. Partitioning around medoids algorithm based on Jensen-Shannon distance was used to cluster the viral enterotypes. We estimated the alpha-diversity by using Shannon index. Partial spearman correlation was performed to investigate the association between viral composition and BMD phenotypes. Constrained linear regression was applied to detect BMD-related viruses. Linear discriminate analysis was then used to identify viral biomarkers for differential osteoporosis status. The samples were clustered into three viral enterotypes, where the most enriched bacteriophages were *Staphylococcus phage SPbeta-like*, *Shigella phage SflV* and *uncultured crAssphage*. Alpha-diversity of virus was negatively correlated with lumbar 1 to lumbar 4 BMD ( $\beta=-0.100$ ,  $P=0.027$ ) and four species including *Shigella phage SflI* ( $\beta=3.23E-05$ ,  $P=0.030$ ), *Staphylococcus phage SPbeta-like* ( $\beta=1.06E-04$ ,  $P=0.005$ ), *Salmonella virus HK620* ( $\beta=-3.23E-05$ ,  $P=0.039$ ) and *Enterobacteria phage mEp460* ( $\beta=-3.83E-05$ ,  $P=0.006$ ) were found to be related to ulna-radius-ultra distance BMD. In total, 21 viruses were identified as biomarkers for distinguishing between normal, osteopenia and osteoporosis groups. Our study, for the first time, provided evidence that human gut virome may influence the regulation of bone metabolism. These findings may contribute to a better understanding of the biological mechanisms underlying osteoporosis.

## PgmNr 1971: Genetic studies of height-associated protein expression levels.

### Authors:

E. Bartell<sup>1,\*</sup>; M. Fujimoto<sup>2,\*</sup>; J. Hirschhorn<sup>1,3</sup>; A. Dauber<sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Harvard Medical School, Boston, MA; 2) Cincinnati Center for Growth Disorders, Division of Endocrinology, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, OH; 3) Division of Endocrinology, Boston Children's Hospital, Harvard Medical School, Boston, MA; 4) Division of Endocrinology, Children's National Medical Center, Washington, DC

---

Genome-wide association studies (GWAS) have identified thousands of common genetic variants associated with human height, implicating hundreds of genes and loci. However, the mechanisms by which many of these genes affect human height are still poorly understood. Integrating knowledge of phenotypic interaction with protein expression levels can help disambiguate the functional underpinnings of these associations.

Objective: To interrogate biological function at height-associated loci in the GH/IGF signaling pathway.

Design/Methods: We used data from the Cincinnati Genomic Control Cohort (CGCC), a community-based cohort comprised of 1,020 children surrounding Cincinnati. Protein expression levels for proteins (free, total *IGF-1*, intact, total *IGFBP-3*, *PAPP-A2*, *IGF-2*, *IGFBP-5*) were measured by ELISA and corrected for age- and sex-effects. In this abstract, we associated protein-level phenotypes using plink qassoc, corrected for 20 principal components, and stratified by sex and population, in ~870 European- and African-descent individuals. Meta-analyses were performed using the METAL fixed-effects model. GWAS of anthropometric traits were performed in the UK Biobank of ~400,000 individuals using Bolt-LMM, or curated from publically available summary statistics.

Results; We identified 17 independent genome-wide significant protein-level-associated loci ( $p < 5 \times 10^{-8}$ ). The two most robust associations were, or were in tight LD with, previously identified expression quantitative trait loci (eQTLs). The *IGFBP-3* locus is associated with serum total *IGFBP-3* and *IGF-2* levels. Despite falling within a height locus, conditional analyses show the effect on protein levels is independent of the height signal ( $p = 2.8 \times 10^{-31}$  post conditioning). However, conditional analyses show the protein level signal colocalizes with sitting height ratio (SHR, sitting height divided by standing height). The *IGFBP-5* locus is associated with *IGFBP-5* levels and is also independent of height signal identified in the region ( $p = 3.3 \times 10^{-32}$  post conditioning). Preliminary analyses show this signal to be colocalized with birth weight.

Conclusion: We have identified novel pQTLs for *IGF-2*, *IGFBP-3*, and *IGFBP-5* that act independently from human height but may interact with related anthropometric traits including SHR. Additionally, this suggests that SNPs affecting adult height in these loci do not work via increasing serum levels of these proteins but rather through a different and undetermined mechanism.

## PgmNr 1972: Multi-locus variation of Notch and Hedgehog signaling pathways potentially contributing to cases of limb anomalies.

### Authors:

H. Hijazi <sup>1</sup>; E. Yilmaz <sup>2</sup>; D. Romeo Bertola <sup>3,4</sup>; G. Lopes Yamamoto <sup>3</sup>; S.S. da Costa <sup>4</sup>; H. Kocak Eker <sup>5</sup>; N. Elcioglu <sup>6</sup>; A. Gezdirici <sup>2</sup>; S. Isikay <sup>7</sup>; Z.H. Coban Akdemir <sup>1</sup>; Y. Bayram <sup>8</sup>; C. Rosenberg <sup>4</sup>; C.M.B. Carvalho <sup>1</sup>; J.E. Posey <sup>1</sup>; S.N. Jhangiani <sup>9</sup>; D.M. Muzny <sup>9</sup>; D. Pehlivan <sup>1,10</sup>; R.A. Gibbs <sup>9</sup>; J.R. Lupski <sup>1,9,11,12</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, Texas.; 2) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, 34303, Turkey; 3) Instituto da Criança do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo, Department of Pediatrics, São Paulo, Brasil; 4) Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brasil; 5) Department of Medical Genetics, Konya Education and Research Hospital, Konya, Turkey; 6) Department of Pediatric Genetics, Marmara University School of Medicine, Istanbul 34899, Turkey; Eastern Mediterranean University School of Medicine, Cyprus, Mersin 10 99628, Turkey; 7) Department of Physiotherapy and Rehabilitation, Hasan Kalyoncu University, School of Health Sciences, Gaziantep, 27410, Turkey; 8) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 9) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA; 10) Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA; 11) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA; 12) Texas Children's Hospital, Houston, Texas, USA

**Background:** Spatial and temporal regulation of Notch and Hedgehog (Hh) signaling during limb bud development is critical for normal growth and patterning of vertebrate limbs. Rare variant mutant alleles in genes encoding members of these pathways are associated with a spectrum of limb developmental phenotypes. The majority of disease associations involve a single gene with monoallelic variation cosegregating in an autosomal dominant fashion with reduced penetrance or as a sporadic birth defect in association with a *de novo* mutational event. The molecular etiology of reduced penetrance, frequently observed amongst families with limb malformations, remains mostly unresolved. Recent studies have shown that penetrance of congenital scoliosis due to variation at the *TBX6* locus can result from a null allele and a common variant non-coding allele that reduces expression to a level below haploinsufficiency. Moreover, a two-loci mutational burden model in the *SMAD6*-*BMP2* and *TBX4*-*FGF10* pathways can explain penetrance of craniosynostosis and lethal lung developmental disorders, respectively.

**Methods:** We hypothesized that multi-locus pathogenic variation in a signaling pathway important to limb development might underlie the penetrance of some limb birth defects. We applied exome sequencing (ES) and array comparative genomic hybridization (aCGH) to affected individuals from seven unrelated families with morphologically heterogeneous limb malformations.

**Results:** In each of the seven studied families, likely damaging multi-locus pathogenic variation was found in genes encoding proteins of the Notch and Hh signaling pathways. Affected individuals harbored rare variants ( $MAF \leq 0.001$ ) in at least two genes of the implicated pathways. This included

single-nucleotide and copy-number variants in previously disease-associated genes such as *GLI2*, *HOXD13*, *DYNC2H1*, *PTCH1*, *SUFU*, *BHLHA9*, and *FGFR1* as well as genes not previously known to be associated with limb deformities in humans such as *NLK*, *HAND2*, *DNER* and *EIF3A*. Additionally, two probands were found to harbor common single nucleotide variants in *HES3*, a transcription factor that plays a role in Notch signaling.

**Conclusions:** Our data suggest a possible epistatic interaction of rare variants in factors of the Notch and Hh signaling pathways in individuals with limb malformations. Further study of additional family members can elucidate the relationship between intrafamilial genotypic variability and disease trait expression.

## PgmNr 1973: A large-scale whole genome sequencing to identify less common and rare variants associated with osteoporosis and fracture: The NHLBI TopMed Study.

### Authors:

Y.H. Hsu<sup>1,2,3</sup>; H. Xu<sup>4</sup>; M. Montasser<sup>5</sup>; C. Crandall<sup>6</sup>; C. Delaney<sup>7</sup>; M. Fu<sup>5</sup>; A. Justice<sup>8</sup>; D. Karasik<sup>1</sup>; J. O'Connell<sup>5</sup>; C. Farber<sup>11</sup>; J. Perry<sup>5</sup>; S. Salimi<sup>5</sup>; E. Streeten<sup>5</sup>; X. Zhang<sup>4</sup>; R. Wallace<sup>9</sup>; D. Evans<sup>10</sup>; J. Tobias<sup>12</sup>; B. Richards<sup>13</sup>; J. Blangero<sup>14</sup>; B. Psaty<sup>7</sup>; K. Wiggins<sup>7</sup>; C.T. Liu<sup>4</sup>; R. Jackson<sup>15</sup>; D. Kiel<sup>1,3</sup>; B. Mitchell<sup>5</sup>; NHLBI TopMed

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Harvard Medical School, Boston, Massachusetts. Marcus Institute for Aging Research and Harvard Medical School, Boston, MA; 2) Harvard School of Public Health, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Boston University, School of Public Health, Boston, MA; 5) University of Maryland, Baltimore, MD; 6) University of California, Los Angeles, CA; 7) University of Washington, Seattle, WA; 8) University of North Carolina, Chapel Hill, NC; 9) University of Iowa, Iowa City, IA; 10) University of Queensland Diamantina Institute, Woolloongabba, QLD, Australia; 11) University of Virginia, Charlottesville, VA; 12) Department of Clinical Science at North Bristol, University of Bristol, Bristol, UK; 13) Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, QC, Canada; 14) University of Texas Rio Grande Valley School of Medicine, Edinburg, TX; 15) Ohio State University Wexner Medical Center, Columbus, OH

---

Although common-variant based genome-wide association studies (GWAS) have identified loci associated with osteoporosis and fractures, such approaches have inherent limitations and it is unlikely to impute less common, rare and structural variations that have larger effect sizes and might reveal unique pathways influencing bone metabolism. Bone mineral density (BMD) at lumbar spine (LS) and femur neck (FN) has been used to diagnose osteoporosis. Previously, we reported the largest common-variant GWAS analysis (N=~400k) on estimated BMD from heel bones. However, the identified loci only explained ~20% variance of BMD, comparing to the 85% BMD heritability estimated from twins. Thus, current common-variant based GWAS approach only detects a small fraction of genetic heritability. To explain the missing heritability, whole-genome sequencing (WGS) is rapidly being incorporated into mainstream biomedical researches to provide comprehensive enumeration of sequence variation. To identify novel less common and rare variants that are associated with BMD, we utilized the deep-coverage WGS (average 30X coverage) in ~10K Caucasian participants obtained from Amish, CHS, FHS, SAFOS and WHI studies (the NHLBI TOPMED Program) as well as ALSPAC and TWINSUK studies. We first performed gene-based collapsing association tests (allele-count and SKAT tests) to identify rare functional coding variants (MAF  $\leq$  0.5%) of BMD. Only loss-of-function, missense with deleterious effects and protein-altering short insertion/deletion coding variants were included. Covariates adjusted in the models included age, age<sup>2</sup>, sex, weight, height, ancestral genetic background, cohort studies and menopause status (women only). To consider potential familiar relatedness among subjects, a linear mixed-effect model with kinship/genetic relationship matrix was applied. Genome-wide significance was  $p < 4.27 \times 10^{-6}$  after Bonferroni correction. The most significant association was found in *IGHE* gene ( $p=7.18 \times 10^{-8}$ ) with lower LS BMD. The other novel findings included *SLC26A11*, *ERGIC3* and *STMN1* genes associated with LSBMD. No

genome-wide association was found for FN BMD, but strong association ( $p < 10^{-5}$ ) was found with *TFAP2E*, *CYP2B6*, *GDF10* and *IL6* genes. Additional analyses on BMD and fracture are underway to include functional non-coding variants (e.g. variants affecting gene regulation in bone cells), which will provide a comprehensive ascertainment on sequence variants and their impact to bone health.

## PgmNr 1974: Genome-wide association study identifies three loci associated with chronic widespread musculoskeletal pain.

### Authors:

M.S. Rahman; M.B. Freidin; FMK. Williams

[View Session](#) [Add to Schedule](#)

**Affiliation:** Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom

---

Chronic widespread musculoskeletal pain (CWP) is a complex trait with poorly understood pathogenesis. Heritability estimates of CWP range between 48-54% indicating a substantial genetic contribution. Candidate gene approaches to study genetic factors contributing to CWP have yielded inconsistent findings. Genome-wide association studies (GWAS) of CWP are limited. We performed the largest so far GWAS for CWP using 249,843 participants of Northern European descent from questionnaire responses within UK Biobank, aged 40-72 years. CWP cases (n=6914) were identified by combining self-reported diagnosis of pain all over the body, simultaneous pain in knee, shoulder, hip and back lasting for 3+ months, and those with a doctor's diagnosis of fibromyalgia. We identified 3 genomic loci (2 in chromosome 3, 1 in chromosome 22) having genome wide significance ( $p < 5E-08$ ). The loci include genes *RNF123*, *ATP2C1* and *COMT*. Four additional loci were identified at suggestive at  $p < 5E-07$ . Gene-based analysis identified additional 20 associated genes (MAGMA  $p < 2.631E-6$ ), of which only *ARVCF*, adjacent gene to *COMT*, had been implicated to pain. This is the first study to identify *COMT* as associated with CWP using an agnostic study approach, but it has been widely investigated as a candidate gene (e.g. CWP/Fibromyalgia, sciatic pain, neuropathic pain, shoulder pain, headache, cancer pain). A significant positive genetic correlation was observed with body mass index ( $rg=0.3071$ ,  $p=8.81E-18$ ), depressive symptoms ( $rg=0.65$ ,  $p=2.06E-36$ ) and neuroticism ( $rg=0.3931$ ,  $p=1.68E-19$ ). Our study suggests new genes underlying the pathophysiology of CWP/fibromyalgia. Gene expression and tissue enrichment analysis as well as replication effort are underway



## PgmNr 1975: Meta-analysis of >1.4 million individuals for human height.

### Authors:

S. Vedantam<sup>1,2</sup>; L. Yengo<sup>3</sup>; E. Marouli<sup>4</sup>; A. Locke<sup>5</sup>; S. Berndt<sup>6</sup>; A.R. Wood<sup>7</sup>; S. Raghavan<sup>8,9</sup>; T. Assimes<sup>10,11</sup>; P. Deloukas<sup>4</sup>; G. Lettre<sup>12</sup>; Y. Okada<sup>13</sup>; T. Frayling<sup>7</sup>; P.M. Visscher<sup>3</sup>; J.N. Hirschhorn<sup>1,2,14</sup>; GIANT consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Endocrinology, Boston Children's Hospital, Boston, Massachusetts, USA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Program in Complex Trait Genomics, Institute for Molecular Bioscience, University of Queensland, St Lucia, Brisbane, Australia; 4) William Harvey Research Institute, Barts & the London Medical School, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ UK; 5) McDonnell Genome Institute, Division of Genomics and Bioinformatics, and Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; 6) National Cancer Institute, Bethesda, MD USA; 7) Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, EX2 5DW, UK; 8) VA Eastern Colorado Healthcare System, Aurora, CO, USA; 9) Department of Medicine, University of Colorado School of Medicine, Aurora, CO, USA; 10) VA Palo Alto Health Care System, Palo Alto, CA, USA; 11) Department of Medicine, Stanford Univ. School of Medicine, Stanford, CA, USA; 12) Montreal Heart Institute and Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal, Québec, Canada; 13) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan; 14) Departments of Pediatrics and Genetics, Harvard Medical School, Boston, MA, USA

---

Human height is a highly heritable and classic polygenic trait. The GIANT consortium has previously identified hundreds of variants associated with adult height; a HapMap-imputed meta-analysis with UK Biobank of ~700,000 individuals yielded 3290 near-independent SNPs with  $p < 10^{-8}$ . These provided a prediction accuracy of ~19.5% in an independent European-ancestry sample. We are performing a much larger GWAS for height, thus far including >1.4 million people from 278 studies covering six major ancestries, imputed to the 1KG Phase 3 and (for European-ancestry samples) HRC reference panels. Linear mixed models and four principal components were used to account for relatedness and population stratification. After QC filters, we identified 2295 autosomal loci >1Mb apart that have signals at genome-wide significance. The strongest-associated (lead) variants, in each locus included 168 with frequency <5%, and 6 insertion/deletion variants. Because of the large sample size apparently distinct loci could reflect weak LD with very strong associations >1Mb away. We used GCTA-COJO to perform conditional joint analysis from summary statistics to address this issue and to identify independent signals within loci. This analysis identified 3,925 likely independent association signals with  $p < 10^{-8}$ . Of these, 233 signals had lead variants with frequency <5%, and the 3,925 variants clustered in many fewer (725) genomic loci >1 Mb apart. These variants had a prediction accuracy of ~23.3% in 8,552 independent unrelated European-ancestry samples. Using S-BayesR we find that prediction based on common HapMap 3 SNPs yields an accuracy of ~32.3%. We are further increasing sample size to ~3 million and performing trans-ethnic meta analyses, fine mapping using results from multiple ancestries, pathway gene set analyses, prediction in multiple ancestries, and X chromosome analyses. The large number of signals and loci will more completely define the genetic

and biologic basis of human height and provide general insights into contribution of common and low frequency variants to polygenic traits.

## **PgmNr 1976: Genomic prediction of osteoporotic fracture risk using machine learning techniques on 1,103 SNPs of 5,133 individuals in the cohort study of osteoporotic fractures in men.**

### **Authors:**

Q. Wu <sup>1,2</sup>; B. Bhattarai <sup>3</sup>; F. Nasoz <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, Nevada.; 2) Department of Environmental & Occupational Health, School of Public Health, University of Nevada Las Vegas, Las Vegas, Nevada.; 3) Department of Computer Science, University of Nevada Las Vegas, Las Vegas Nevada; 4) The Lincy Institute, University of Nevada Las Vegas, Las Vegas Nevada

---

Predicting an individual's disease risk from genomic variants remains a major challenge due to complex gene interactions and regulations. Osteoporosis is a common, silent, and costly disease without symptoms. Correctly identifying individuals who will sustain osteoporotic fracture is critical for preventing this devastating outcome in aging population worldwide. Numerous Single Nucleotide Polymorphisms (SNPs) associated with osteoporotic fracture were discovered by GWAS and GWAS meta-analysis. However, how to best utilize these genetic variants to effectively predict an individual's fracture risk remains unclear. Traditional statistical approaches lack the flexibility or the adequacy to model complex gene interactions and regulations. Thus, we developed several machine learning models of genomic data and ran experiments to identify the best machine learning model for fracture prediction. We used genomic data of Osteoporotic Fractures in Men cohort Study (N=5,133) as the data source. Genotype imputation was conducted at the Sanger Imputation Server. 1,103 fracture-associated SNPs were identified and corresponding genetic risk scores were calculated for each participant. Traditional fracture risk factors and genomic variants were included for analysis. Data were normalized and split into a training set (80%) and a validation set (20%). Synthetic Minority Over-sampling technique was used to account for low fracture samples rate in the data. Fracture prediction models were built using random forest, gradient boosting, and multi-layer perceptron (MLP) with backpropagation algorithms separately and logistic regression was used as a reference model. The fracture prediction performance of each model was assessed by area under the ROC curve (AUC), recall, and precision in the validation set. We found that the performance of MLP in predicting fracture was the best among the four models, with AUC of 0.98, recall of 0.70, and precision of 0.84. The performance of random forest and gradient boosting were also better than that of logistic regression. Random forest and gradient boosting had the AUC of 0.94 and 0.93, recall of 0.64 and 0.62, and precision of 0.59 and 0.77, respectively. Logistic regression had the worst performance with AUC of 0.90, recall of 0.55 and precision of 0.30. Thus, we concluded that all three machine learning techniques had better fracture prediction performance than logistic regression, and MLP performed best for fracture prediction in men.

## PgmNr 1977: Exome sequencing of a multiplex family with idiopathic scoliosis implicates *KIF7* in IS pathogenesis.

### Authors:

M. Cuevas<sup>1</sup>; M. Cattell<sup>1</sup>; X. Chen<sup>1</sup>; E. Terhune<sup>1</sup>; C. Wetthey<sup>1</sup>; J. Casey<sup>1</sup>; B. Sutphin<sup>1</sup>; S. Pradhan<sup>1</sup>; R. Baschal<sup>1,2</sup>; A. Monley<sup>1,3</sup>; K. Jones<sup>2</sup>; E. Baschal<sup>1,4</sup>; B. Appel<sup>2</sup>; N. Hadley Miller<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Orthopedics, University of Colorado Anschutz Medical Campus, Aurora, CO.; 2) Department of Pediatrics, University of Colorado Anschutz Medical Campus, Aurora, CO; 3) Musculoskeletal Research Center, Children's Hospital Colorado, Aurora, CO; 4) Department of Otolaryngology, University of Colorado Anschutz Medical Campus, Aurora, CO

---

**Background:** Idiopathic scoliosis (IS) is an axial developmental disorder of the spine with a strong familial genetic component. Given the significant variability in presentation and progression of this condition in otherwise normal individuals, the genetic architecture of IS is believed to be complex in nature. One strategy for gene identification is the study of a single multiplex family with IS that potentially harbors a rare causal personal variant that can be related to disease expression.

**Methods:** Peripheral blood samples were collected from a large multiplex family affected with IS. We extracted DNA and performed exome sequencing for three affected individuals. Bioinformatic filtering identified unique, low frequency variants (MAF <5%) that were shared by all members of the family and were predicted to be damaging to the protein. This gene list then underwent gene ontology enrichment analyses (GO, PANTHER, BiNGO) to identify enriched gene categories. Genes within top GO terms were prioritized for further study in zebrafish through CRISPR-Cas9 knockouts. Zebrafish phenotypes were studied through direct visualization, histological staining, and micro-CT.

**Results:** Within this family, 19 variants were found in 19 genes that passed bioinformatic filtering. Of the 19 variants, 16 are nonsynonymous SNPs, 1 is an in-frame insertion, 1 is in a splice site, and 1 is a stop-gain variant. Gene prioritization resulted in study of *KIF7*, which encodes a motor protein in the ciliary axoneme. Sanger sequencing of this variant in additional members of the family confirmed that this variant segregated with the phenotype of affected vs unaffected individuals. Generation of *Kif7* mutant zebrafish resulted in homozygous mutants that exhibit spinal curvatures in early larval stages through adulthood.

**Conclusion:** Study of a single multiplex family with IS demonstrated that a variant in *KIF7*, a cilia-related gene, is potentially related to the scoliotic phenotype. Functional studies within a knockout zebrafish model support this observation. Ciliopathy related genes have been reported to be related to IS potentially through a dysregulation of signaling pathways driving segmentation and somite production of the axial skeleton. Further study of this gene family in additional IS families and IS cohorts are needed.

## PgmNr 1978: Targeted sequencing in genetic screening and gene discovery in sporadic polydactylies.

### Authors:

G. You <sup>1</sup>; Q. Fu <sup>1</sup>; Z. Wang <sup>2</sup>; Y. Xu <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory Medicine, Shanghai Children's Medical Center, Shanghai, Shanghai, China; 2) Pediatric Orthopedics, Shanghai Children's Medical Center, Shanghai, Shanghai, China

---

### Background

Polydactylies are a highly heterogeneous group of skeletal deformities in clinical and genetical. The purpose of this study was to gain insight into the frequencies of genetic variants in genes implicated in polydactyly by targeted sequencing.

### Methods

A total of 181 patients diagnosed with polydactylies were recruited over a period of 3 years. Targeted sequencing was performed for 721 genes associated with the pathogenesis of skeletal dysplasia, and candidate variants were selected using bioinformatics analysis.

### Results

After bioinformatics analysis and filtering process, a total of 890 rare variations identified as “loss of function” or “deleterious” variants in 98.9% (179 of 181) of cases with polydactyly by targeted sequencing. For each sample, an average of 4.9 variants were identified. Among the six non-syndromic polydactyly associated genes, five mutations were identified in *GLI3* and one novel mutation was identified in Zone of polarizing activity (ZPA) regulatory sequence (ZRS). Among the known syndromic polydactyly associated genes inherited in dominant manner, fifteen mutations were identified in seven genes in fifteen polydactyly cases.

### Conclusions

This study demonstrates the feasibility of targeted sequencing for the molecular diagnosis of polydactyly. However, this panel still display genes with low or no diagnostic yield.

## PgmNr 1979: A genome-wide association study implicates the *BMP7* locus as a risk factor for nonsyndromic metopic craniosynostosis.

### Authors:

C.M. Justice<sup>1</sup>; A. Cuellar<sup>2</sup>; K. Bala<sup>2</sup>; J.A. Sabourin<sup>1</sup>; M.L. Cunningham<sup>3</sup>; J.M. Phipps<sup>4,5</sup>; E. Simeonov<sup>7</sup>; R. Kaneva<sup>8</sup>; C. Senders<sup>9</sup>; M. Zwienenberg<sup>10</sup>; T. Roscioli<sup>11</sup>; M. Barba<sup>12,13,14</sup>; K. Conway<sup>15</sup>; V.C. Sheffield<sup>16</sup>; L. Brody<sup>17</sup>; J.L. Mills<sup>18</sup>; D. Kay<sup>19</sup>; R.J. Sicko<sup>19</sup>; P.H. Langlois<sup>20</sup>; R.K. Tittle<sup>21</sup>; L. Botto<sup>22</sup>; M.M. Jenkins<sup>23</sup>; J.M. LaSalle<sup>24</sup>; W. Lattanzi<sup>13,14</sup>; A.O.M. Wilkie<sup>4,5,6</sup>; A.F. Wilson<sup>1</sup>; P.A. Romitti<sup>15</sup>; S.A. Boyadjiev<sup>2</sup>; National Birth Defects Prevention Study

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genometrics Section, Computational and Statistical Genomics Branch, Division of Intramural Research, NHGRI, NIH, Baltimore, MD, USA; 2) Department of Pediatrics, University of California Davis, Sacramento, CA, USA; 3) University of Washington, Department of Pediatrics, Division of Craniofacial Medicine, Seattle Children's Craniofacial Center and Seattle Children's Research Institute, Seattle, WA, USA; 4) MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, UK; 5) Oxford Centre for Genomic Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford, UK; 6) Craniofacial Unit, Oxford University Hospitals NHS Foundation Trust, Oxford, UK; 7) National Institute of Pediatrics, Sofia Medical University, Sofia, Bulgaria; 8) Molecular Medicine Center, Department of Medical Chemistry and Biochemistry, Medical Faculty, Medical University of Sofia, Sofia, Bulgaria; 9) Department of Otolaryngology, Head and Neck Surgery, University of California Davis, Sacramento, CA, USA; 10) Department of Neurosurgery, University of California Davis, Sacramento, CA, USA; 11) Neuroscience Research Australia, University of New South Wales, Sydney, Australia; 12) Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy; 13) Istituto di Neurochirurgia, Università Cattolica del Sacro Cuore, Rome, Italy; 14) Istituto di Anatomia Umana e Biologia Cellulare, Università Cattolica del Sacro Cuore, Rome, Italy; 15) Department of Epidemiology, College of Public Health, The University of Iowa, Iowa City, IA, USA; 16) Department of Pediatrics, Division of Medical Genetics, Carver College of Medicine, The University of Iowa, Iowa City, IA, USA; 17) Gene and Environment Interaction Section, NHGRI, NIH, Bethesda, MD, USA; 18) Epidemiology Branch, Eunice Kennedy Shriver NICHD, NIH, Bethesda, MD, USA; 19) Division of Genetics, Wadsworth Center, NYS Department of Health, Albany, NY, USA; 20) Birth Defects Epidemiology and Surveillance Branch, Texas Department of State Health Services, Austin, TX, USA; 21) Department of Nutritional Sciences, University of Texas at Austin, Austin, TX, USA; 22) Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT, USA; 23) National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, GA, USA; 24) Department of Medical Microbiology and Immunology, Genome Center, and MIND Institute, University of California Davis, Davis, CA, USA

---

Craniosynostosis (CS) arises from premature closure of one or more cranial vault sutures in infants. Following up the independent associations we identified for *BMP2* (rs1884302) and *BBS9* (rs10262453) in our genome-wide association study (GWAS) for sagittal nonsyndromic CS (sNCS), we conducted a GWAS for another midline defect, metopic NCS (mNCS). Our analysis of 228 non-Hispanic white (NHW) case-parent trios from an international sample identified three variants that exceeded

genome-wide significance ( $P \leq 5 \times 10^{-8}$ ): rs781716 ( $P = 4.79 \times 10^{-9}$ ; odds ratio [OR]=2.41) intronic to *SPRY3*; rs6127972 ( $P = 5.61 \times 10^{-9}$ ; OR=2.26) intronic to *BMP7*; and rs62590971 ( $P = 4.05 \times 10^{-8}$ ; OR=0.38) located ~155 kb upstream from *TGIF2LX*. Only the association for rs6127972 ( $P = 0.004$ , OR=1.45) was replicated in an independent NHW sample (194 unrelated mNCS cases, 333 controls) with a meta-analysis  $P = 3.11 \times 10^{-9}$  (OR=1.45). Both rs1884302 and rs10262453, identified in our sNCS GWAS, were genotyped in our mNCS discovery and replication samples. The variant rs10262453 was borderline significant in our discovery sample ( $P = 9.22 \times 10^{-5}$ , OR=1.72), replicated ( $P = 0.0004$ , OR=1.61), and had a meta-analysis ( $P = 1.33 \times 10^{-7}$ , OR=1.62). Notably, the C allele for this variant was over-transmitted in mNCS probands, whereas the A allele was over-transmitted in sNCS probands. We sequenced all seven exons of *BMP7* in 183 mNCS cases, of which 118 were included in the 228 mNCS discovery sample case-parent trios, and did not identify any variants with obvious deleterious effects. *BMP7* expression and BMP7 protein secretion were measured in mesenchymal stem cells derived from matched fused metopic and open sutures from the same probands, and no significant difference was observed, regardless of the rs6127972 genotype. Functional assessment of rs6127972 using western blot and ELISA produced no difference in BMP7 expression or protein levels in synostotic metopic versus open sutures. Data from luciferase studies suggested that the locus may act as a repressor element with the risk allele exerting stronger repression. Zebrafish transgenic analysis produced inconclusive results. In summary, our GWAS identified *BMP7* (rs6127972) to be strongly associated with mNCS, providing support for the role of *BMP7* in skeletal development. Our meta-analysis of sNCS and mNCS GWAS data identified *BBS9* (rs10262453) as a risk variant for both subtypes. These findings provide additional important insights into genetic antecedents for midline CS.

## PgmNr 1980: *LRP5* gain of function mutation and osteopetrosis: An illustrative case and review of the literature.

### Authors:

S.M. Jafri <sup>1</sup>; E. Burke <sup>1</sup>; C. Wahl <sup>1</sup>; G. Vézina <sup>4</sup>; D. Bulas <sup>1</sup>; D.R. Adams <sup>1,3</sup>; M.T. Collins <sup>2</sup>; K.S. Pan <sup>2</sup>; W.A. Gahl <sup>1,3</sup>; T.C. Markello <sup>1</sup>; U.D.N. Undiagnosed Diseases Network <sup>5</sup>; C. Toro <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Undiagnosed Diseases Program, National Human Genome Research Institute, NIH, Bethesda, Maryland.; 2) National Institute of Dental and Craniofacial Research, NIH, Bethesda, MD; 3) Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 4) Neuroradiology, Children's National Medical Center, Washington, D.C.; 5) Common Fund, Office of the Director, National Institutes of Health

---

**INTRO:** Autosomal dominant osteopetrosis type 1 (ADO1) is a rare disease caused by heterozygous, gain-of-function mutation in the low-density lipoprotein receptor-related protein 5 (*LRP5*) gene. *LRP5* codes for a coreceptor in the canonical Wnt signaling pathway, which is known to be involved in the development of some bone disorders and plays a crucial role in bone mass regulation. Here we describe the case of a 51-year-old female who presented with severely elevated bone density, atypical skeletal findings, and an acquired Chiari malformation.

**METHODS:** The study participant was seen as part of an IRB-approved protocol associated with the NIH Undiagnosed Diseases Program (UDP). The UDP evaluates individuals with severe medical conditions that remain refractory to diagnosis despite extensive clinical workup. *LRP5* variants were identified by clinical exome analysis. Targeted clinical studies were used for phenotyping. Intracranial volume was measured by CAT scan-based 3D reconstruction.

**RESULTS:** The subject showed many typical features of *LRP5* associated osteopetrosis including increased bone mineralization, a palatal torus and a history of an inability to float in water. Less common characteristics included a Chiari malformation. Molecular analysis detected a known osteopetrosis-associated *LRP5* mutation and also mutations in two potentially modifying genes *DIO2* and *CYP12A1*. Intracranial volume stabilization was seen to occur after institution of sex hormone suppression therapy. Features of our case showed overlap with previously reported cases, notably on the severe end of the phenotypic spectrum.

**DISCUSSION:** We present a case of severe *LRP5*-associated osteopetrosis along with detailed phenotyping data. Chiari malformations are a known but unusual complication of osteopetrosis likely resulting from osseous changes of the skull. Management of osteopetrosis can be challenging. In this case, treatment to suppress sex hormones caused an apparent stabilization supporting the consideration of that therapy as a treatment in this condition. Finally, we consider a hypothesis that changes in the *DIO2* and *CYP12A1* genes may modify the *LRP5* phenotype.



## PgmNr 1981: Deciphering disorders involving scoliosis and comorbidities.

### Authors:

S. Zhao <sup>1,2</sup>; Y. Zhang <sup>1,2</sup>; L. Wang <sup>1,2</sup>; Z. Wu <sup>1,2,4</sup>; J. Lupski <sup>3,5,6</sup>; J. Zhang <sup>1,2</sup>; G. Qiu <sup>1,2</sup>; N. Wu <sup>1,2,3</sup>;  
Deciphering disorders Involving Scoliosis and COMorbidities (DISCO) study

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Peking Union Medical College Hospital, Beijing, Beijing, China; 2) Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China.; 3) Department of Molecular and Human Genetics, Baylor College of Medicine; Houston, TX 77030, USA.; 4) Medical Research Center, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing, China; 5) Departments of Pediatrics, Texas Children's Hospital and Baylor College of Medicine, Houston, TX 77030, USA.; 6) Texas Children's Hospital, Houston, TX 77030, USA.

### Objective

The Deciphering disorders Involving Scoliosis and COMorbidities (DISCO) international study group is committed to interpret the genomic data produced from subjects with scoliotic phenotype, to increase the diagnostic yield and to decipher the novel genetic etiology and the disease mechanisms. Our ultimate goal is to provide precise clinical practice guidelines for molecular diagnosis and disease management, genetic counseling, and even targeted therapy to the patients.

### Method

Through the past eight years, we have been collecting DNA sample and clinical information from patients with deformities involving scoliosis and their available family members. Exome sequencing (ES) or whole genome sequencing (WGS) were performed on these patients and families to explore the genetic etiology underlying scoliosis and comorbidities. Candidate genetic/genomic variants were studied using *in vivo* and *in vitro* models.

### Results

In total, we have sequenced 3723 patients and family members using either ES (N=2923) or WGS (N=800). Their clinical diagnoses included congenital scoliosis (N=1300), adolescent idiopathic scoliosis (N =300), Mayer-Rokitansky-Kuster-Hauser (MRKH) syndrome (N=710), short stature (N =730) and SAPHO syndrome (N =120). These genetic data have enabled publication of 50 scientific papers and 53 abstracts in international. Novel genes such as *MYH3* were identified to be associated isolated scoliosis, while *HOXA10* and *BMP4* have been linked to scoliosis involving multi-system malformations.

### Conclusion

Our investigations into the pathogenesis of scoliosis advanced our understanding of somitogenesis during development and the course of scoliosis conditions. We unraveled the risk factors and potential mechanisms of scoliosis and other disease, and develop more effective clinical diagnostic

strategy and intervention methods based on these discoveries. By translating our research findings into clinical application, DISCO study group increased the diagnostic yield and improved disease management for those patients suffering scoliosis.

## PgmNr 1982: A cartilage exome combined with genome-wide association study (GWAS) meta-analysis identifies association of *COL11A1* with adolescent idiopathic scoliosis (AIS).

### Authors:

A.M. Khanshour<sup>1</sup>; Y.H. Kidane<sup>1</sup>; I. Kou<sup>2</sup>; L. Antunes<sup>3</sup>; J. Zhao<sup>4,5</sup>; N. Makki<sup>6</sup>; R.S. Gray<sup>7</sup>; J.J. Rios<sup>1,8,9</sup>; N. Ahituv<sup>4,5</sup>; C.A. Gurnett<sup>3</sup>; Y.Q. Song<sup>10</sup>; S. Ikegawa<sup>2</sup>; C.A. Wise<sup>1,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sarah M. and Charles E. Seay Center for Musculoskeletal Research, Texas Scottish Rite Hospital for Children, Dallas, TX 75219, USA.; 2) Laboratory of Bone and Joint Diseases, RIKEN Center for Integrative Medical Sciences, 108-8639 Tokyo, Japan.; 3) Department of Neurology, School of Medicine, Washington University, St. Louis, MO 63130, USA.; 4) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94131, USA.; 5) Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94131, USA.; 6) Department of Anatomy and Cell Biology, College of Medicine, University of Florida, Gainesville, FL, USA.; 7) Department of Pediatrics, Dell Medical School, The University of Texas, Austin, TX, USA.; 8) McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.; 9) Departments of Orthopaedic Surgery, Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA.; 10) School of Biomedical Sciences, The University of Hong Kong, Hong Kong, China.

AIS is a rotational deformity of the spine that affects ~3% of children worldwide. Genetic contributions to AIS are complex and largely undiscovered. Our group and other researchers previously identified AIS common risk loci by GWAS. More recently we reported a multi-ethnic meta-analysis that identified new common variant associations in inter-and intra-genic non-coding regions. Gene-functional enrichment analysis found significant perturbation of pathways involving cartilage development. To discover new AIS disease genes, we have performed a new GWAS of 11,469 non-Hispanic white subjects (1,358 cases/10,111 controls) in which restricted our search space to common (MAF>0.01) missense variants in the cartilage exome (GO:0051216). In the discovery stage, three variants surpassed a nominal significance threshold after multiple testing correction ( $P<5.3E-4$ ): P1335L in *COL11A1* [ $P=5.6E-05$ ]; K286E in *IMPAD1* [ $P=2.4E-4$ ] and T263I in *TGFB1* [ $P=3.5E-4$ ]. To validate our findings, we performed a meta-analysis (N=96,693) that combined the discovery summary statistics for the three variants with data from independent cohorts from Japan (n=79,211), Hong Kong (n=3,062), and Missouri (n=2,951). We obtained the strongest signal in P1335L in the *COL11A1* ( $P=4.3E-9$ , OR=1.11) gene that encodes the alpha1 chain of type XI collagen. Separately, whole exome sequencing in 73 families identified a novel P738L mutation (CADD=24) that arose *de novo* and a rare (gnomad\_AF=0.0001) mutation (G1420S, CADD=28) that displayed dominant inheritance in another family. Log2 Counts Per Million (log2CPM) from two independent RNA-Seq studies confirmed high *COL11A1* expression in mouse chondrocytic cell line (log2CPM=6.3, mean=2.0, sd=4.5) and mouse intervertebral disc (IVD) (log2CPM=11.1, mean=3.3, sd=2.8). Type XI collagen, a fibrillar collagen composed of alpha 1,2, and 3 peptide chains, is important for fibril-forming and structural organization of cartilages including the IVD. In humans, heterozygous disruptive *COL11A1*

mutations cause Marshal and Stickler syndromes, and a common haplotype at this locus has been associated with degenerative disc disease. Moreover, enrichment of rare *COL11A2* mutations has been described in cohorts with severe AIS. From these results we hypothesize that missense changes in the collagen XI triple helix, particularly at conserved proline or glycine residues, may alter the integrity of the IVD and predispose to AIS during rapid growth.

## PgmNr 1983: Transcriptome-wide association study identifies novel candidate genes associated with osteoporosis.

### Authors:

M. Zhu; P. Yin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Biomedical Information Technology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China

---

Osteoporosis (OP) is a polygenetic disease which is usually characterized by low bone mineral density (BMD). GWAS studies have identified hundreds of genetic loci associated with BMD. However, the causality of these loci remains elusive. To identify causal genes of the associated loci, we detected trait-gene expression associations by Transcriptome-Wide Association Study (TWAS) method, which directly imputes gene expression effects from GWAS summary data, using a statistical prediction model trained on GTEx reference transcriptome data, restricting at OP, blood and muscles-skeletal tissues data. Then we performed a colocalization analysis using the COLOC software, which evaluated the posterior probability with different patterns: association characterized by a single shared causal variant or two distinct causal variants, termed *PP4* and *PP3* in the COLOC notation respectively. The ultimate analysis identified 276 TWAS-significant genes at a Bonferroni-corrected threshold of  $P < 3.7E-8$ , including 142 genes showing strong evidence of a single shared causal variant with *PP4*  $> 0.8$  and 134 exhibiting evidence of joint causal variants with *PP3*  $> 0.9$ . Among them, 210 potential causal genes were validated by VarElect tool, which indicates them of directly or indirectly involved in OP with literature evidences. Further several candidate genes were enriched for differential expression genes in osteoblasts cells expression profiles (GSE35956 and GSE35959) from GEO database, including *IBSP*, affecting calcium and hydroxyapatite binding, and *CD44*, regulating alternative splicing of gene transcription, and *SPTBN1*, interacting with calmodulin in a calcium-dependent manner and candidate. Protein-protein interaction and pathway enrichment analysis detected several OP-associated pathways, including the MAPK signaling pathway and Osteoclast differentiation, B cell receptor signaling pathway. Transcriptome fine-mapping identifies more disease-related genes and provide additional insight into the development of novel targeted therapeutics to treat OP and reduce the risk of fracture.

## PgmNr 1984: Genetic fine mapping identifies novel *de novo* mutations in chromosome 22 potentially associated with the risk of non-syndromic cleft palate only (nsCPO).

### Authors:

S. Beiraghi <sup>1</sup>; S. Mazumder <sup>2</sup>; U. Radhakrishna <sup>3</sup>; W.S. Looman <sup>4</sup>; B.G. Van Ness <sup>5</sup>; A.K. Mitra <sup>2,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatric Dept, 6-150 Moos, Univ Minnesota, Minneapolis, Minnesota.; 2) Harrison School of Pharmacy, Auburn University, Auburn, AL, USA.; 3) Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, Michigan, USA.; 4) School of Nursing, University of Minnesota Minneapolis, MN, USA.; 5) Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, MN, USA.; 6) Experimental Therapeutics Division, O'Neal Comprehensive Cancer Center, University of Alabama at Birmingham School of Medicine, Birmingham, AL, USA.

---

Nonsyndromic cleft palate only (nsCPO), a type of orofacial cleft, is a common congenital craniofacial developmental malformation in humans. Recent studies have revealed a number of genetic variants in several candidate genes and chromosomal regions associated with the risk of nsCPO. Our overarching aim is to perform chromosome wise genetic fine mapping studies to identify additional novel mutations associated with nsCPO risk.

We have earlier shown presence of novel functional variants within chromosome 1 as risk factors for nsCPO. In the presented study, we used exome sequencing data from 4 subjects (proband, affected brother, mom and dad) to discover *de novo* variants within chromosome 22 that may be involved in the complex genetic etiology of nsCPO.

Peripheral blood was obtained following informed consent, genomic DNA was isolated and pair-end exome sequencing was performed at 100X depth of coverage on Illumina's HiSeq2500 next-generation high-throughput sequencing system. Exome sequencing data was processed using Galaxy and variants were called using SAMtools and GATK Unified Genotyper. Common overlapping variants were filtered out using Western European (CEU) genomes in the 1000 Genomes Project. Variants were annotated using human reference database (GRCh37.75), snpEFF (likely effect) and ClinVar (clinical annotations).

Consequently, 72,580 variants were found in Chromosome 22 with a QUAL score >30. Since our focus was to look for *de novo* mutations, variants with at least one mutant allele in the proband and the affected sibling (homozygous mutant or heterozygous) and none in either parent (homozygous wild type) were considered. Eighty (80) *de novo* variants were discovered. These included genes implicated in inborn genetic diseases including CDC45 (chr22:19483898:T/C), PI4KA (chr22:21067806:C/G), MCM5 (chr22:35815296:C/T), TMPRSS6 (chr22:37475013:aa/aAGa), and TRIOBP (chr22:38104705:A/T). Interestingly, ClinVar clinical annotations revealed CDC45 and MCM5 are associated with Meier-Gorlin syndrome (OMIM# 224690), a genetic disease characterized by cleft palate as one of the symptoms.

Thus, our study identified novel *de novo* genetic mutations within chromosome 22 as potential risk factors for nsCPO.

## PgmNr 1985: The role of *FZD6* and related genes in craniofacial morphogenesis.

### Authors:

L. Maili<sup>1,2</sup>; C. Urbina<sup>2</sup>; B.T. Chiquet<sup>2,3</sup>; A. Letra<sup>2,3</sup>; G.T. Eisenhoffer<sup>1,4</sup>; J.T. Hecht<sup>1,2,3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Graduate School of Biomedical Sciences, University of Texas Health Science Center, Houston, TX;  
 2) Pediatrics Department, McGovern Medical School, University of Texas Health Science Center, Houston TX; 3) Center for Craniofacial Research, University of Texas School of Dentistry at Houston, Houston TX; 4) Department of Genetics, University of Texas MD Anderson Cancer Center, Houston, TX

---

Craniofacial birth defects are the most common structural abnormalities at birth. They include nonsyndromic cleft lip and palate (NSCLP), which has a prevalence of 1 in 700 and affects approximately 4000 newborns in the United States annually. Although many genes have been implicated in NSCLP etiology, only a portion of the heritability has been explained. We have previously identified *FZD6*, a receptor in the Wnt pathway, as a NSCLP gene in a large multigenerational family and further shown that knockdown or overexpression of *fzd6* causes cartilage and bone abnormalities in zebrafish. Additionally, we and others have shown that loss of co-receptor *lrp6* causes bilateral cleft lip, while knockdown and mutant forms of *lrp5* cause abnormal craniofacial cartilage phenotypes in zebrafish. Furthermore, the facial region completely fails to form in mice deficient for *dkk1*, a *Wnt* antagonist. These genes collectively act at the receptor level of the  $\beta$ -catenin mediated Wnt signaling, which is known to play an important role in craniofacial morphology.

The goal of this study was to evaluate the effects of gene dosage perturbations for *fzd6* and related genes in zebrafish using craniofacial morphometry. First, we evaluated  $\beta$ -catenin activation during craniofacial development utilizing a stable zebrafish reporter line. Reporter expression was analyzed from 1-7 days post fertilization (dpf) and was highly active in the oral cavity and other facial regions between 3-5 dpf. As development progressed,  $\beta$ -catenin expression became restricted to cells around the peri-oral region. Upon *fzd6* knockdown, morphometric analysis revealed that head width ( $p=0.0002$ ), olfactory distance ( $p=0.0002$ ), mouth perimeter ( $<0.0001$ ) and mouth width ( $p=0.0005$ ) were significantly reduced while mouth height was increased ( $0.006$ ) in morphant embryos compared to controls. *Lrp5*, *Lrp6* and *dkk1b* knockdown embryos are being evaluated in a similar manner and will also be presented.

The results of this study provide important information about the effects of *FZD6* and related  $\beta$ -catenin/Wnt pathway genes in craniofacial development and morphogenesis.



## PgmNr 1986: Kabuki syndrome: Long time follow-up of two patients and the importance of a multidisciplinary team.

### Authors:

D. Sabau <sup>1</sup>; P. Calapod <sup>2</sup>; C. Rusu <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, Sfântul Ioan Emergency Hospital, Bucharest, Romania; 2) Medical Genetics, Iași Regional Medical Genetics Centre, Iași, Romania

---

Kabuki syndrome (KS [MIM 147920]) is a rare and variable genetic disorder that affects 1 in 30,000 – 86,000 individuals. This disorder is characterized by typical dysmorphic features, failure to thrive in infancy and later obesity, skeletal anomalies and dermatoglyphic changes and intellectual disability. Some cases associate cardiac and renal defects. Mutations in genes *KMT2D* (12q13.12) [MIM 602113] and *KDM6A* (Xp11.3) [MIM 300128] are known to cause KS.

We describe 2 patients diagnosed in Iași Regional Medical Genetics Centre to illustrate the importance of the clinical findings for the diagnosis, the evolution of the clinical picture in time, as well as particular findings associated. In both cases the diagnosis of Kabuki syndrome was confirmed using DNA sequencing.

Both cases are each represented by the only child of young, healthy, unrelated couples and family history is negative. In one case the pregnancy has been uneventful and the child was born at full term, natural birth, without particular events in the neonatal period, while in other the baby was born at an early term and the mother had a varicella-zoster infection in the seventh month of pregnancy and developed chickenpox. In both of them intellectual development has been delayed.

Case 1 (male, 9 years old): growth deficiency and microcephaly, typical dysmorphic face, persistence of fetal fingertip pads and clinodactyly of the 5th digits, joint laxity, hypotonia and moderate intellectual disability. Investigations revealed congenital heart defects and hydronephrosis.

Case 2 (male, 15 years old): normal growth, typical dysmorphic face, brachydactyly in hands and feet and persistence of fetal fingertip pads, genital anomalies (unilateral cryptorchidism and phimosis) and inguinal hernia (all surgically corrected) and mild intellectual disability. In time he developed gynecomastia. Investigations revealed congenital heart defects and unilateral renal ectopia and dysplasia.

The follow-up showed tendency to particular obesity, progression of renal dysfunction and behavioural changes in both cases. The evolution over time of the dysmorphic facial features of both patients will be illustrated in the study.

In conclusion, we present two cases of KS to exemplify the changing of the clinical picture in time, to present particularities and to discuss the importance of a multidisciplinary team for a correct management of the patient and family.

## PgmNr 1987: Contribution of rare variants in the noncanonical Wnt/planar cell polarity pathway genes to nonsyndromic cleft lip and/or palate.

### Authors:

L.A. Brito; G. Hsia; J.Y.T. Wang; M.R. Passos-Bueno

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics, University of São Paulo, São Paulo, São Paulo, Brazil

---

Nonsyndromic cleft lip with or without cleft palate (NSCL/P) is a complex trait, which etiology is largely influenced by common variants. Increasing evidence has also been observed for the role of high-effect rare variants, such as those in cadherin-catenins adhesion complex genes. These pathogenic variants have been mostly detected segregating in familial cases, which comprise 30% of all NSCL/P. While genes of the canonical Wnt/b-catenin pathway have been associated with NSCL/P, Wnt/Planar Cell Polarity (PCP) pathway has been underexplored in this context. Wnt/PCP plays an important role in vertebrate craniofacial development, e.g. during neural crest cells (NCC) migration. Our aim is to determine the relevance of rare variants in Wnt/PCP genes to NSCL/P, by sequencing 9 PCP genes in 251 probands of familial NSCL/P cases. Genes included in this study presents enriched expression during craniofacial development and intolerance to loss-of-function variants (according to Sysface and gnomAD databases). Using Ampliseq sequencing platform and Miseq (Illumina), we have found, in *CELSR3*, *DVL3*, *FZD7*, *PRICKLE1*, *VANGL2*, *WNT5A* and *WNT11*, a total of 27 rare (<0.5% in gnomAD) missense variants with CADD score >20, and one private heterozygous frameshift deletion in *PRICKLE1*. Aggregation of rare variants was not observed for any PCP gene, comparing patients with Brazilian controls from ABraOM database, although borderline significance was found for *PRICKLE1* (P=0.05), and *WNT11* (P=0.06). As *PRICKLE1* is highly intolerant to LoF variants, we investigated segregation of the frameshift variant, and found segregation in proband's affected mother and brother. Exome sequencing was performed in these three individuals, and no candidate variant was found in genes previously implicated with orofacial clefting. In addition, presence of copy number variants was refuted by array CGH in the proband. We suggest that rare variants in core Wnt/PCP genes are not a prevalent cause of NSCL/P; nevertheless, a *PRICKLE1* frameshift variant is the most probable cause of NSCL/P in a Brazilian family. The role of *PRICKLE1* in NSCL/P etiology has been supported by mouse mutants, which may present cleft of the upper lip and palate. In addition, *PRICKLE1* has also been shown to regulate levels of the cleft-associated epithelial cadherin prior to NCC delamination in zebrafish. Our results may thus represent the first case of a *PRICKLE1* pathogenic variant segregating in familial NSCL/P.

## PgmNr 1988: A population-based study of gene x smoking interactions and orofacial cleft risk.

### Authors:

A.L. Petrin<sup>1</sup>; W. Lyu<sup>2</sup>; P.A. Romitti<sup>3</sup>; M.M. Jenkins<sup>4</sup>; R.G. Munger<sup>5</sup>; J.C. Murray<sup>6</sup>; A.J. Wilcox<sup>7</sup>; R.T. Lie<sup>8</sup>; G.L. Wehby<sup>2</sup>; L.M. Moreno-Urbe<sup>1,9</sup>; National Birth Defects Prevention Study

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Iowa Institute for Oral Health Research, University of Iowa, Iowa City, Iowa.; 2) Department of Health Management and Policy, University of Iowa, Iowa City, Iowa.; 3) Department of Epidemiology, College of Public Health, The University of Iowa, Iowa City, IA; 4) Centers for Disease Control and Prevention, National Center on Birth Defects and Developmental Disabilities, Atlanta, GA; 5) Department of Nutrition and Food Sciences, Utah State University, Logan, UT; 6) Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA; 7) Epidemiology Branch, National Institute of Environmental Health Sciences/National Institutes of Health, Durham, NC; 8) Department of Global Public Health and Primary Care, University of Bergen, Bergen, Norway; 9) Department of Orthodontics, College of Dentistry, the University of Iowa, Iowa City, IA

---

Nonsyndromic orofacial clefts (OFCs) are among the most prevalent birth defects, affecting on average 1 in 700 births. Maternal smoking is the most well-established environmental factor that may contribute to OFC risk; however, gene-environment interaction studies continue to show conflicting results, possibly due to heterogeneity in risk by maternal and fetal smoking detoxification gene variants. We examined the interaction effects between first trimester maternal smoking (active and passive) and variants in 7 smoking detoxification genes (maternal and fetal *CYP2E1*, *ELAVL2*, *GRID2*, *GSTP1*, *NAT2*, *TBK1*, *ZNF236*) on risk of nonsyndromic OFCs. We combined individual-level data from 5 population-based studies — Iowa Case Control Study, National Birth Defects Prevention Study, Norway Facial Clefts Study, Norwegian Mother and Child Cohort, and Utah Child and Family Health Study—to create the largest population-based case-control sample to date for this assessment, consisting of 1,330 (1,743) case children (mothers) and 2,549 (3,417) control children (mothers). We evaluated interactions using multivariable logistic regression, including fixed effects for study site, maternal age and education at delivery, pre-pregnancy body mass index, and first-trimester alcohol consumption and folic acid/multivitamin use. We show that *CYP2E1* and *TBK1* increase OFC risk through maternal genotypes in the presence of maternal active and passive smoking. *NAT2* however, seems to affect OFC risk through the fetal genotype in the presence of any maternal passive smoking. *ZNF236*, on the other hand, acts via both maternal and fetal genotypes in combination with all types of smoking behavior measured in this study. These findings suggest meaningful heterogeneity in the risk of OFC associated with maternal smoking by genetic variants in several smoking metabolism genes. This study also confirms passive smoking as a teratogenic factor for OFC and further shows its interactive effects with smoking metabolism genes to increase OFC risk.

## PgmNr 1989: A locus on 4q28 is a laterality modifier with opposite effects for left- and right-sided cleft lip.

### Authors:

E.J. Leslie <sup>1</sup>; D. Chang <sup>1</sup>; N. Harerimana <sup>1</sup>; M.R. Sun <sup>2</sup>; C. Buxo <sup>3</sup>; K. Christensen <sup>4</sup>; F. Deleyiannis <sup>5</sup>; J.T. Hecht <sup>6</sup>; L.L. Field <sup>7</sup>; L.M. Moreno Uribe <sup>8</sup>; I. Orioli <sup>9</sup>; C. Padilla <sup>10</sup>; J.R. Shaffer <sup>11</sup>; A.R. Vieira <sup>11</sup>; S.M. Weinberg <sup>11</sup>; G.L. Wehby <sup>8</sup>; E. Feingold <sup>11</sup>; J.C. Murray <sup>8</sup>; R.J. Lipinski <sup>2</sup>; J.C. Carlson <sup>11</sup>; M.L. Marazita <sup>11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Emory University; 2) University of Wisconsin Madison; 3) University of Puerto Rico; 4) University of Southern Denmark; 5) UC Health Colorado Springs; 6) UT Health Sciences Center at Houston; 7) University of British Columbia; 8) University of Iowa; 9) INAGEMP (National Institute of Population Medical Genetics) Brazil; 10) University of the Philippines Manila; 11) University of Pittsburgh

---

Despite the symmetry of the human face, unilateral orofacial clefts (OFCs) are twice as likely to occur on the left side than the right; a ratio that cannot be explained by random chance so we looked to identify factors that might contribute to this sidedness difference. Past OFC studies sought to identify general OFC risk factors and were not designed to be able to identify variants specific to laterality subtypes because they combined them and assumed identical risks. We therefore performed genome-wide association studies to identify modifiers of cleft lip laterality in a multiethnic sample of 2,142 total cases with cleft lip only (CLO) or cleft lip with cleft palate (CLP). We performed separate analyses for CL and CLP given evidence from human genetics, epidemiology, and model systems that these can have distinct etiologies. In the CL analysis (N=216 LCL, 130 RCL) we identified a novel signal on 4q28.1 ( $p=8.4 \times 10^{-8}$ ) in which the minor alleles increased risk for LCL 3.5-fold over RCL. We then compared each group of cases to 1700 independent controls, which revealed that that SNPs at this locus influenced risk in opposite directions: the allele that increased risk for LCL ( $p=2.0 \times 10^{-6}$ ; OR=1.84) was protective for RCL ( $p=0.01$ ; OR=0.84). These effects were consistent between the European and Latinos cases that comprised the largest two population groups in our sample, indicating that this result was not driven solely by a single population group. Furthermore, the 4q28.1 locus showed no effect on overall risk of CL nor for any type of CLP ( $p=0.2$ ; N=638 LCLP, 416 RCLP). The associated SNPs were located in an intergenic region ~400kb downstream of *FAT4*, a protocadherin involved in planar cell polarity. In vitro studies suggest that *FAT4* is partially responsive to SHH signaling as addition of SHH to cranial neural crest cells suppresses *FAT4* expression. We performed mouse in situ hybridization experiments and showed that *Fat4* expression localized to the mesenchyme of the “elbow” of the medial nasal processes, whereas *Shh* is expressed in the distal end. Interestingly, *SHH* mutations primarily are associated with midline cleft lips suggesting that these genes may cooperate in upper lip development. This study identifies, for the first time, genetic risk underlying cleft lip laterality subtypes by identifying the 4q28.1 locus as a modifier of CLO sidedness. This may provide insights into the development of other human organ asymmetries or their pathology.

## PgmNr 1990: Identification of *de novo* variants by WGS of European and Latino trios and their contribution towards nonsyndromic orofacial clefts.

### Authors:

M.R. Bishop <sup>1</sup>; P. Chopra <sup>1</sup>; N. Mukhopadhyay <sup>2</sup>; D.J. Cutler <sup>1</sup>; M.P. Epstein <sup>1</sup>; G.L. Wehby <sup>3</sup>; L.C. Valencia-Ramirez <sup>4</sup>; C. Restrepo <sup>4</sup>; L.M. Moreno Uribe <sup>3</sup>; J.T. Hecht <sup>5</sup>; F. Deleyiannis <sup>6</sup>; S.M. Weinberg <sup>2</sup>; T.H. Beaty <sup>7</sup>; J.C. Murray <sup>3</sup>; E. Feingold <sup>2</sup>; M.L. Marazita <sup>2</sup>; E.J. Leslie <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University, Atlanta, GA.; 2) University of Pittsburgh, Pittsburgh, PA.; 3) University of Iowa, Iowa City, IA.; 4) Fundación Clínica Noel, Medellín, Colombia.; 5) McGovern Medical School UT Health at Houston, Houston, TX.; 6) UC Health, Colorado Springs, CO.; 7) Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD.

---

Orofacial clefts (OFCs) are the most common craniofacial malformation in humans, affecting 1 in 1000 live births around the world. Approximately 70% of cleft lip with/without cleft palate cases and 50% of cleft palate only cases are classified as nonsyndromic OFCs, a complex trait with a strong genetic component. Linkage studies, GWAS, and candidate gene studies have identified a number of genetic risk variants, with GWAS being the most successful. However, GWAS approaches focus on common variants and overlook rare inherited and *de novo* variants. All types of variants (common, rare, *de novo*, etc.) appear to confer risk to OFCs but have not been thoroughly assessed for their contribution towards nonsyndromic OFCs. Therefore, we performed whole-genome sequencing (WGS) on 760 case-parent trios with nonsyndromic OFCs. In order to identify *de novo* variants that may contribute to nonsyndromic OFCs, we developed a stringent filtering pipeline to process and analyze the WGS data; following bioinformatics processing, variants were prioritized, further filtered based on allele balance, and visually inspected using Integrative Genomic Viewer to manually curate a list of high confidence *de novo* variants. We prioritized predicted loss-of-function variants found in 56 European trios with cleft palate only and 256 Latino trios with nonsyndromic cleft lip and palate for curation. We identified loss-of-function *de novo* variants in *CHD7*, *COL2A1*, *CTNND1*, *TFAP2A*, and *RPL5*. Loss-of-function mutations in the genes cause syndromes involving OFCs, but our trios notably lacked the additional phenotypic features required for clinical diagnosis of a syndrome. This suggests that variants in known syndromic OFC genes may also be contributing to nonsyndromic OFCs. Therefore, we closely examined *de novo* variants in *IRF6* and *GRHL3*, which cause the most common OFC syndrome, Van der Woude syndrome. We identified three *de novo* missense variants, two in *IRF6* and one in *GRHL3*. Additionally, we found high confidence *de novo* missense variants in OFC candidate genes previously nominated by GWAS or animal models: *FAF1*, *THADA*, and *SHROOM3*. Overall, the identification of *de novo* variants in a large cohort of case-parent trios with nonsyndromic OFCs provides the strongest evidence to date for a role for *de novo* variants in OFC etiology.

## PgmNr 1991: ***HLA-DQA1\*03:01* and *DQB1\*03:02* are protective alleles for Alzheimer disease, multiple sclerosis, and celiac disease.**

### Authors:

J.J. Farrell <sup>1</sup>; C. Zhu <sup>1</sup>; X. Zhang <sup>1,6</sup>; G. Jun <sup>1</sup>; R. Mayeux <sup>2</sup>; J. Mez <sup>7</sup>; J.L. Haines <sup>3</sup>; M.A. Pericak-Vance <sup>4</sup>; G.D. Schellenberg <sup>5</sup>; K.L. Lunetta <sup>6</sup>; L.A. Farrer <sup>1,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine (Biomedical Genetics), Boston University Medical School, Boston, Massachusetts.; 2) Department of Neurology and Sergievsky Center, Columbia University, New York, New York.; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, Ohio.; 4) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, Florida; 5) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; 6) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts.; 7) Department of Neurology, Boston University School of Medicine, Boston, Massachusetts

---

Recent Alzheimer disease (AD) genome-wide studies have found associations with multiple loci involved in immune and inflammatory pathways including the MHC II region. To pinpoint the causative variant(s) in this region, we tested the association of HLA alleles with AD in European ancestry (EA), African American (AA), and Caribbean Hispanic (CH) unrelated participants (7,024 AD cases, 6,574 controls) in the Alzheimer Disease Sequencing Project (ADSP) and with an AD proxy outcome (report of AD-type dementia in at least one parent) among individuals of British ancestry in the UK Biobank (34,127 AD cases, 239,254 controls).

The UK Biobank imputed the HLA genotypes (n=226 alleles) using HLA\*IMP:02. For the ADSP, we called the HLA genotypes (n=445 alleles) from the ADSP WES and WGS samples using the HLA-Genotyper software tool we developed. A single-variant level analysis was first performed within ancestry and sequencing platform (UKBB British (imputed), ADSP EA (WES) +ADSP EA (WGS), ADSP CH (WGS), and ADSP AA (WGS) with AD status as outcome, and age (as appropriate for the study design), sex and principle components as covariates. Meta-analysis of results across all populations and datasets was performed using SeqMeta.

Meta-analysis revealed two genome-wide significant alleles (HLA-DQA1\*03:01, OR=0.90, P=2.66x10<sup>-8</sup> and HLA-DQB1\*03:02, OR=0.88, P=2.87x10<sup>-8</sup>) that were protective in all groups (study specific ORs between 0.34 and 0.90). Previous reported findings for AD association with HLA-DRB1\*15:01 (OR=1.00, P=0.55) and HLA-DQA1\*01:02 (OR=1.00, P=0.89) did not replicate in the UKBB Cohort.

To examine the protective influence of these alleles in other diseases, we performed a PheWAS with ICD10 diagnoses in the UK Biobank cohort. DQA1\*03:01 was significantly protective for two autoimmune diseases: multiple sclerosis (OR= 0.65, P=2.33x10<sup>-12</sup>) and celiac disease (OR=0.36, P=9.14x10<sup>-55</sup>). HLA-DQB1\*03:02 was also significantly associated with celiac disease (OR=0.61, P=1.22x10<sup>-11</sup>) and multiple sclerosis (OR=0.66, P=5.10x10<sup>-7</sup>). Notably, cognitive decline has been

observed in persons affected with multiple sclerosis and to a lesser extent celiac disease. These results suggest that these diseases share a common immune or inflammatory pathway. Future studies of the protective mechanism of these HLA alleles may provide insight into the neurobiology of cognitive decline and dementia and suggest novel therapeutic approaches for these disorders' shared cognitive decline.



## PgmNr 1992: Using polygenic risk scores to identify risk for asthma in an underrepresented population.

### Authors:

C.H. Arehart<sup>1</sup>; M. Daya<sup>1</sup>; S. Chavan<sup>1</sup>; N. Rafaels<sup>1</sup>; M.P. Boorgula<sup>1</sup>; C. Cox<sup>1</sup>; T.M. Brunetti<sup>1</sup>; C. Ober<sup>2</sup>; L.A. Lange<sup>1</sup>; E. Bleecker<sup>3</sup>; D. Meyers<sup>3</sup>; C. Rotimi<sup>4</sup>; E.E. Kenny<sup>5</sup>; T.H. Beaty<sup>6</sup>; C.R. Gignoux<sup>1</sup>; R.A. Mathias<sup>7</sup>; K.C. Barnes<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, University of Colorado Denver, Aurora, CO, 80045, USA; 2) Department of Human Genetics, University of Chicago, Chicago, IL, 60637, USA; 3) Department of Medicine, University of Arizona College of Medicine, Tucson, AZ, 85724, USA; 4) Center for Research on Genomics & Global Health, National Institutes of Health, Bethesda, MD, 20892, USA; 5) Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA; 6) Department of Epidemiology, Bloomberg School of Public Health, JHU, Baltimore, MD, 21205, USA; 7) Department of Medicine, Johns Hopkins University, Baltimore, MD, 21224, USA

---

Polygenic risk scores (PRSs) have shown promise in predicting complex disease risk attributable to genetics. These risk scores have become more accurate as genome-wide association studies (GWASs) grow yet can be biased by the individuals selected in the training dataset.

Despite comprising only ~16% of the global population, Europeans account for ~79% of the current GWAS catalog. This presents a challenge for modeling PRSs in non-European cohorts because PRS analyses 1) require large discovery GWAS sample sizes to approach genetic-variant-based heritability and 2) become unreliable and unpredictably biased when using European GWASs for non-European cohorts. To address this problem, Coram et al. devised XP-BLUP which combines trans-ethnic and ethnic-specific data in a two-component linear mixed effects model. The present study applies XP-BLUP to asthma, a complex disease for which morbidity and mortality is greater among African Americans (AA) compared to European Americans.

Using European Transnational Asthma Genetics Consortium (TAGC) GWAS summary statistics, we implemented XP-BLUP to compute PRSs in an AA population available through the *Consortium on Asthma among African-ancestry Populations in the Americas* (CAAPA). After quality control, we included 3,740 unrelated individuals (1,223 cases, 2,517 controls) with GWAS array data imputed against the TOPMed reference panel. Model training was based on 80% of the cohort, and model testing (risk scoring) was based on the remaining 20%. We included 389 asthma-associated TAGC variants in the AA samples for the disease-specific variance component. While the standardized PRSs were not highly predictive of asthma, they discriminated between cases and controls better than random chance (OR [95% CI]=1.36 [1.12-1.66], p=0.002, and AUC=0.57). The PRSs ranged from -3.3 to 4.8 SD units and a 3-unit increase in PRS corresponded to a 2.5-fold increase in asthma risk. Subsetting to childhood-onset asthma yielded similar results. In contrast, standard PRS models (single component linear model trained on CAAPA and LDpred trained on TAGC) could not discriminate between cases and controls all with p>0.3.

In summary, the resultant risk scores were concordant with phenotypic outcomes and may provide a



good representation of the contribution of genetic variants to asthma. Increasing GWAS training sample sizes, especially those of African ancestry, would contribute to improved risk profiling of genetic predisposition to disease.

## PgmNr 1993: Utilizing whole genome sequence data to elucidate the genetics of total serum IgE concentrations.

### Authors:

C. Cox <sup>1</sup>; P. Kachroo <sup>2</sup>; J. Lasky-Su <sup>2</sup>; X. Li <sup>3</sup>; D. Qiao <sup>4</sup>; D.A. Meyers <sup>3</sup>; G. O'Conner <sup>4</sup>; S.T. Weiss <sup>2</sup>; N. Hansel <sup>5</sup>; R.M. Reed <sup>6</sup>; V.E. Ortega <sup>7</sup>; I. Ruczinski <sup>8</sup>; T.H. Beaty <sup>9</sup>; R.A. Mathias <sup>9</sup>; M. Daya <sup>1</sup>; K.C. Barnes <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, University of Colorado, Denver, CO; 2) Channing Division of Network Medicine, Brigham Women's Hospital and Harvard Medical School, Boston, MA; 3) Genomics and Precision Medicine, University of Arizona Tucson; 4) Department of Medicine, Boston University School of Medicine; 5) Division of Pulmonary and Critical Care Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 7) Health Sciences, Wake Forest University, Winston-Salem, NC; 8) Department of Biostatistics, Bloomberg School of Public Health, JHU, Baltimore MD; 9) Department of Epidemiology, Bloomberg School of Public Health, Baltimore, MD

**RATIONALE:** Total serum IgE (tIgE) is an important mediator of allergic disease. tIgE is highly elevated in atopic diseases such as asthma, rhinitis, conjunctivitis, atopic dermatitis and food allergy, and elevated tIgE is associated with coronary artery disease. tIgE is an important intermediary phenotype of asthma, and has a higher heritability than asthma; with heritability estimates ranging between 40-80%. In order to better understand risk factors for atopy in the context of asthma, we performed a multi-ethnic whole genome meta-analysis of tIgE using TOPMed sequence data.

**METHODS:** Data from Barbados Asthma Genetics Study, The Genetic Epidemiology of Asthma in Costa Rica and the Childhood Asthma Management Program, Severe Asthma Research Program and Framingham Heart Study (FHS) were included in our meta-analysis. Analysis strata were defined based on study, race/ethnicity and asthma status (n=6,104 total, n=1,352 African ancestry, n=4,571 European ancestry, n=2,117 asthmatics). Linear mixed effect models (which allows for adjustment of relatedness) was run through the ENCORE web server using TOPMed freeze6a data (SAIGE software) separately for each stratum. Results were combined using inverse-variance meta-analysis.

**RESULTS:** We confirmed four of the five genome-wide associations previously reported ( $P \leq 10^{-5}$ , *FCER1A*, *IL13*, *MHC class II* and *STAT6* gene loci). SNPs in the vicinity of *STAT6* achieved genome-wide significance ( $P \leq 5 \times 10^{-8}$ ). This region also achieved genome-wide significance in the previous FHS GWAS of tIgE. *STAT6* is a transcription factor that affects Th2 lymphocyte responses mediated by IL-4 and IL-13, and has been identified by genome-wide association studies (GWAS) of asthma performed by the Trans-National Asthma Genetic Consortium (TAGC) and the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA). Two novel loci with suggestive associations ( $P \leq 10^{-6}$ ) not previously reported by GWAS of tIgE were located in the vicinity of the genes *LRRC32+GUCY2EP* and *RIMBP2+STX2*, respectively. *LRRC32* has previously been reported by a GWAS of atopic dermatitis.

**FUTURE WORK:** Our next steps are to include additional TOPMed studies with available tIgE data

(Genetic Epidemiology of COPD and Genetics of Cardiometabolic Health in the Amish) in the meta-analysis and genotype array data from other sources (CAAPA, Genetics of Asthma Susceptibility to Pollution, Atopic Dermatitis Research Network) imputed using the TOPMed reference panel.

## PgmNr 1994: Identifying maximally informative mouse models of asthma through human-mouse comparative transcriptomics.

### Authors:

L.J. Donoghue <sup>1,2</sup>; Y.H. Tsai <sup>3</sup>; K.M. McFadden <sup>1</sup>; J.S. Parker <sup>3</sup>; SNP. Kelada <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 2) Curriculum in Genetics and Molecular Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 3) Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 4) Marsico Lung Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC.

---

The use of mouse models in asthma research has led to valuable insights on airway pathology and functional analyses of putative disease-related genes such as those identified through genome-wide association studies. Numerous mouse model protocols, primarily driven by exposure to allergen(s), have been developed for the goal of investigating hallmark human disease features. Despite their demonstrated utility in research, however, questions remain regarding how accurately mouse models represent human asthma. Furthermore significant protocol variation exists among published mouse model studies, yet protocol differences have never been assessed systematically with regards to how variation impacts model similarity to human disease. Therefore, we sought to evaluate how effectively mouse models of asthma mimic the gene expression differences seen between human cases and controls to aid in model selection and interpretation. To achieve this goal, we used data from eight previously published studies of differential expression in cases versus controls to create a disease-state classifier. After mining the literature for existing mouse model gene expression datasets and performing RNA sequencing on models from our laboratory, we applied our disease-state classifier to estimate a similarity score based on the concordance in expression of key human disease genes in each mouse model. Mouse models varied in similarity scores estimates by both magnitude and within-treatment group variance. Using data generated by our lab, we tested the impact of sampling gene expression from either whole lung or isolated airway tissue as well as variation in mouse genetic background across 31 inbred strains. These data demonstrated that sampling gene expression from isolated airway samples could provide stronger signals in downstream analyses and that genetic background of mouse used considerably impacts similarity to the human disease state. Finally, using hierarchical clustering and pathway analyses we identified genes that are consistently upregulated in mouse models concordantly with human disease and pathways that are discordant. These findings serve as a resource for the asthma research community to aid in the selection of mouse model parameters, evaluate the impact of mouse genetic background, and determine pathway concordance or discordance between human disease and mouse models to gain important context for maximizing the utility and accuracy of findings from mouse models.

## PgmNr 1995: Large-scale genetic analysis identifies 66 novel loci for asthma.

### Authors:

Y. Han <sup>1,2</sup>; Q. Jia <sup>1,2</sup>; P.S. Jahani <sup>3</sup>; B.P. Hurrell <sup>3</sup>; P. Huang <sup>1,2,4</sup>; J. Gukasyan <sup>1,2</sup>; N.C. Woodward <sup>1,2</sup>; E. Eskin <sup>5</sup>; F. Gilliland <sup>1</sup>; O. Akbari <sup>3</sup>; J.A. Hartiala <sup>1,2</sup>; H. Allayee <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033.; 2) Department of Biochemistry and Molecular Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033.; 3) Department of Molecular Microbiology and Immunology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033.; 4) Central South University, Xiangya School of Medicine, Changsha, China.; 5) Department of Computer Science and Inter-Departmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA 90095.

---

To further elucidate the genetic architecture of asthma, we conducted a genome-wide association study (GWAS) in UK Biobank, followed by a meta-analysis with GWAS results from the Trans-National Asthma Genetic Consortium (TAGC) that in total included 8,365,715 SNPs common to both datasets in 88,486 cases and 447,859 controls. 66 novel genomic regions were associated with asthma and replicated nearly all (142/145) previously known genomic regions, bringing the total number of susceptibility loci identified to date to 211. Altogether, the 66 novel loci explained an additional 1.5% of the heritability for asthma, reflecting the modest nature of the susceptibility alleles' effect sizes. Furthermore, the asthma risk alleles at 17 of the novel loci exhibited directionally consistent associations with decreased forced expiratory volume in one second (FEV<sub>1</sub>), of which 10 were also associated with decreased forced vital capacity (FVC). Bioinformatics analyses revealed enrichment of asthma-associated variants colocalizing to DNase I hypersensitive sites in several biologically relevant tissues, with particularly significant enrichment in B and T cells. In addition, 52 novel loci yielded at least one *cis* expression quantitative trait locus (eQTL) in one or more tissues, the most significant of which were in blood and immune cells. These observations are consistent with the enrichment analyses pointing to lymphocytes as playing important roles in asthma. Interestingly, *CD52*, which encodes a membrane glycoprotein present at high levels on the surface of lymphocytes, monocytes, and dendritic cells, was one of only two candidate causal genes that also yielded an eQTL in the lung. More notably, alemtuzumab is an anti-CD52 (αCD52) antibody that results in preferential and prolonged depletion of circulating T and B cells and is FDA approved for treating multiple sclerosis and chronic lymphocytic leukemia. Functional validation of *CD52* was demonstrated through treatment of mice with an αCD52 antibody that not only mimicked the immune cell-depleting activity of alemtuzumab but also significantly reduced pulmonary inflammation, airway epithelium thickness, and allergen-induced airway hyperreactivity. Our collective results provide additional insight into the genetic determinants of asthma, provide further evidence that the immune system plays a prominent role in its pathogenesis, and suggest that *CD52* represents a potentially novel therapeutic target for treating asthma.

## PgmNr 1996: Whole genome sequencing to study rare gene-by-air pollution interactions that influence lung function in minority children.

### Authors:

A.C.Y. Mak <sup>1</sup>; D. Hu <sup>1</sup>; E. Lee <sup>1</sup>; M.J. White <sup>1</sup>; S. Huntsman <sup>1</sup>; C. Eng <sup>1</sup>; S.S. Oh <sup>1</sup>; M.A. Seibold <sup>2</sup>; E.G. Burchard <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of California San Francisco, San Francisco, CA; 2) National Jewish Health, Denver, CO

---

**Introduction** Baseline lung function, quantified as forced expiratory volume in the first second of exhalation (FEV<sub>1</sub>), is a standard diagnostic criterion used by clinicians to identify and classify lung diseases. FEV<sub>1</sub> is strongly influenced by genetics, with an estimated heritability as high as 55%. Numerous genetic loci have been associated with FEV<sub>1</sub>, and there is clear evidence that genetic ancestry contributes to FEV<sub>1</sub> variation. Several epidemiological studies have identified early-life exposure to air pollution as a significant predictor of baseline lung function. FEV<sub>1</sub> is a complex phenotype that has been shown to be heavily influenced by both environmental and genetic factors. Although numerous studies have tried to identify the genetic and environmental factors responsible for variation in FEV<sub>1</sub>, the majority of the heritability remains undefined. We hypothesize that a portion of the “missing heritability” for FEV<sub>1</sub> may be explained by gene-by-environment (GxE) interactions.

**Methods** We performed whole genome sequencing on 3,000 ethnically diverse children with asthma, including Puerto Ricans, Mexican Americans and African Americans. We performed gene-by-air pollution analyses using early-life and lifetime exposure to air pollution (O<sub>3</sub>, NO<sub>2</sub> and PM<sub>2.5</sub>). We applied a variance component test, iSKAT, to identify interactions between sets of rare variants with each air pollution measurement. All models were adjusted for age, sex, global ancestry, and the main effects of the set of rare variants and air pollution.

**Results and Conclusions** We identified rare variants in *IDH1* interact with lifetime PM<sub>2.5</sub> to influence FEV<sub>1</sub> in Puerto Rican. Our preliminary results indicate that GxE interactions play a significant role in FEV<sub>1</sub> variation.

**Acknowledgements** We thank all NHLBI/TOPMed, NHGRI/GSP and NIMHD investigators and participating centers who have contributed to the generation, quality control, coordination and management of the data used in this study. We gratefully acknowledge Fred Lurmann for providing the air pollution data and the study participants who provided clinical and biological data for this study.

## PgmNr 1997: Effects of *ICAM1* genotype on rhinovirus species-specific illnesses: Evidence for a gene-by-asthma interaction.

### Authors:

N. Schoettler <sup>1,2</sup>; D. Jackson <sup>3</sup>; J.A. Park <sup>4</sup>; R.F. Lemanske <sup>3,5</sup>; Y. Bochkov <sup>3</sup>; J.E. Gern <sup>3,5</sup>; C. Ober <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine, University of Chicago, Chicago, IL.; 2) Human Genetics, University of Chicago, Chicago, IL.; 3) Pediatrics, University of Wisconsin, Madison, WI.; 4) Environmental Health, Harvard T. H. Chan School of Public Health, Boston, MA.; 5) Medicine, University of Wisconsin, Madison, WI.

---

Asthma is a complex disorder with both genetic and environmental contributions. In particular, rhinovirus (RV) infections are an important risk factor for childhood onset asthma and for asthma exacerbations throughout life. Three RV species (RV-A, RV-B and RV-C) cause a variety of respiratory tract infections, ranging from mild colds to severe illness requiring hospitalization. All RV-B strains and most RV-A strains use the intracellular adhesion molecule 1 (ICAM-1) receptor for entry into cells. A variant in *ICAM1*, rs5498 (A/G; Lys694Glu), results in decreased ICAM-1 expression on the cell surface. We investigated the effect of this variant on RV-associated illnesses in children 4 to 12 years of age (n=146 with asthma; n=131 without asthma) enrolled in the prospective RhinoGen study in Madison, Wisconsin. Children were observed for 5 weeks during each of 2 successive RV seasons in the fall and spring. Illnesses were recorded and RV species were identified by partial sequencing during each illness. Children were grouped as having 0 vs.  $\geq 1$  RV-A and 0 vs.  $\geq 1$  RV-B illnesses; analyses included sex as a covariate. As expected, children with asthma had overall more RV-A illnesses compared to those without asthma (49.3% vs 42.0%;  $p=0.031$ ), but unexpectedly the G allele was associated with more RV-A illnesses in both groups (additive model;  $p=0.043$ ). In contrast, RV-B illnesses were less common in children with asthma (10.6% vs 15.4%;  $p=0.020$ ). Moreover, there was no genotype effect on RV-B illnesses in children with asthma ( $p=0.805$ ), but a negative association between RV-B illnesses and the number of G alleles in children without asthma ( $p=0.016$ ). This pattern was further reflected in a strong gene-by-asthma interaction on RV-B illnesses ( $p=0.0057$  for genotype,  $p=0.0013$  for asthma status,  $p=0.029$  for the interaction). Our results show that the *ICAM1* G allele, which results in reduced ICAM-1 receptor levels, is associated with less RV-B infection in non-asthmatic children only and with more RV-A illnesses in both asthmatic and non-asthmatic children. We have therefore identified a genetic variant in a viral receptor that has opposite effects on risk for RV-A and RV-B illnesses, and different genotype effects on the occurrence of RV-B illnesses in children with and without asthma. These results highlight the genetic complexity of response to RV infection, an important risk factor for asthma onset and exacerbation.

## PgmNr 1998: Whole-genome sequencing and cell-type specific RNA sequencing of a multiplex pedigree identifies genes for juvenile idiopathic arthritis.

### Authors:

C.N. Avery <sup>1</sup>; K. Chen <sup>2</sup>; J.F. Bohnsack <sup>2</sup>; L.B. Jorde <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Utah School of Medicine, 15 North 2030 East Room 5100, Salt Lake City, UT 84112; 2) Department of Pediatrics, University of Utah School of Medicine, 175 North Medical Drive East, Salt Lake City, UT 84132

---

Pediatric autoimmune diseases are genetically complex, and to date, GWAS appears to confirm that common variants underlie susceptibility to autoimmune conditions. It remains a challenge, however, to identify high-impact pathogenic variants with a clear mechanism in autoimmune disease progression. Juvenile idiopathic arthritis (JIA) is the most common pediatric rheumatic disease, affecting approximately 1 in 1000 children. Current GWAS candidate genes account for approximately 18% of JIA susceptibility, and other susceptibility genes remain to be uncovered. Pedigrees with multiple affected first-degree relatives provide a unique opportunity to discover heritable genetic susceptibility that might be otherwise difficult to detect. Pathogenic variants, even if private to a single family, have the potential to uncover novel disease mechanisms. In this study, whole-genome sequencing (WGS) was performed on a two-generation multiplex pedigree, and RNA was isolated and sequenced from circulating CD4+ T cells, a highly implicated cell type in JIA, of the three affected children. Rare ( $AF < .01$ ) single nucleotide variants in cases were investigated with Pedigree Variant Annotation, Analysis, and Search Tool (pVAASST), a computational tool for prioritizing rare variants with predicted pathogenic effects while considering co-segregation of alleles with the phenotype within a pedigree. Top candidate genes were screened against a larger cohort ( $n = 40$ ) of WGS JIA patients to investigate their potential contribution to disease. Differential expression analysis comparing CD4+ T cells of affected siblings ( $n = 3$ ) and CD4+ T cells of unrelated controls ( $n = 3$ ) replicates previous observations of differential expression of *JUN* and *NFKBIA*, while also revealing novel differences in transcriptional profiles within this pedigree, even during clinical remission. These differentially expressed genes also suggest areas of investigation in non-coding regions of the genome, particularly within this pedigree. Our results invite further investigation into novel disease mechanisms in JIA that could involve crosstalk between lymphocytes and osteocytes. In particular, our candidate genes highlight components of inflammatory pathways not previously implicated in JIA pathology.



## PgmNr 1999: Regulation of Janus kinase 2 by an inflammatory bowel disease causal noncoding SNP.

### Authors:

C.J. Cardinale <sup>1</sup>; M.E. March <sup>1</sup>; X. Lin <sup>2</sup>; L.A. Spruce <sup>1</sup>; J.P. Bradfield <sup>3</sup>; Z. Wei <sup>2</sup>; S.H. Seeholzer <sup>1</sup>; S.F.A. Grant <sup>1</sup>; H. Hakonarson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.; 2) New Jersey Institute of Technology, Newark, NJ; 3) Quantinuum Research LLC, San Diego, CA

---

Among the >240 genetic loci described to date which confer susceptibility to inflammatory bowel disease, a small subset have been fine-mapped down to an individual, noncoding single nucleotide polymorphism (SNP). To illustrate a model mechanism by which a presumed-causal trait-associated noncoding SNP can function, we analyzed rs1887428, located in the promoter region of the Janus kinase 2 (*JAK2*) gene. Using affinity purification-mass spectrometry (AP-MS), we determined that the risk/G allele is bound preferentially by the transcription factor (TF) RBPJ, while the protective/C allele is bound preferentially by the homeobox TF CUX1, impacting reporter gene expression. This SNP did not have significant expression quantitative trait locus (eQTL) associations. To identify target genes of the variant, we constructed subclones of the Jurkat cell line by CRISPR/Cas9 genome editing that were homozygous for the risk or the protective allele of the SNP of interest. Our results at the transcriptome level show that while rs188748 only has a very modest influence on *JAK2* expression, this effect was amplified downstream through the expression of pathway member *STAT5B* (>4-fold). In a CpG methylation study of human donor samples, the risk allele of rs1887428 was associated with increased DNA methylation of the *JAK2* promoter. These results reveal that even in the absence of a consensus TF binding motif or eQTL evidence, a putatively causal SNP can be characterized to yield insight into inflammatory bowel disease mechanisms.

## PgmNr 2000: Single cell transcriptome analysis of circulating plasmacytoid dendritic cells and switched memory B-cells in SLE patients reveals transcriptional subsets within the classical cell lineages.

### Authors:

Y. Ghodke-Puranik <sup>1</sup>; A. Puranik <sup>1</sup>; R. Tapon <sup>1</sup>; M.A. Jensen <sup>1</sup>; A. Gupta <sup>1</sup>; U. Sankaramanchi <sup>1</sup>; J. Paredes <sup>1</sup>; I. Nln <sup>1</sup>; A. Saxena <sup>2</sup>; H.M. Belmont <sup>2</sup>; P.M. Izmirly <sup>2</sup>; R.M. Clancy <sup>2</sup>; J.P. Buyon <sup>2</sup>; T.B. Niewold <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Colton Center for Autoimmunity, NYU Langone Medical Center, New York, NY; 2) New York University Medical Center, New York, NY

---

**Background/Purpose:** Both plasmacytoid dendritic cells (pDCs) and switched memory B cells (SMBCs) are considered to be key effector cells in systemic lupus erythematosus. It seems likely that within these classical cell lineages, additional diversity of function will exist that will contribute to disease pathogenesis. To explore this question, we performed single-cell RNA sequencing in pDCs and SMBCs from SLE patients and controls to assess gene expression patterns and cellular sub-groupings within these lineages.

**Methods:** pDCs and SMBCs from SLE patients (n=10) and Healthy controls (n=5) were purified by magnetic separation. For deep sequencing, we used the Fluidigm C1 HT system with 800 capture site chips to capture single cells. Single cell capture was verified by direct visualization using the Array Scan system, allowing us to remove empty wells and wells with multiple cells. After quality control and adaptor trimming, the data was analyzed using SeqGeq software. pDCs and SMBCs were clustered using UMAP and pseudo-time analysis was performed using the Monocle program. Type I IFN activity in SLE plasma was measured using reporter cell assay.

**Results:** A total of 2774 pDCs and 2578 SMBCs from SLE and healthy controls passed the quality control and were used for further analysis. In pDCs, we observed unique clusters for patients with high interferon, low interferon and controls, indicating that the IFN response is a major determinant of overall gene expression patterns in SLE patient pDCs. IFN signature in pDCs correlated with circulating type I IFN activity in the SLE patients measured at the same time. Other genes upregulated in pDCs included the type I interferon regulator *AXL* and *MACC1*. The SMBCs were heterogenous in patients and controls, and in contrast to the pDCs the overall clustering pattern was independent of the IFN score. SMBC clusters were predominantly defined by genes indicating cellular activation or proliferation such as *HLA-DRs* and *CREB1*, or genes associated with nucleic acid processing such as *DNASE1* and *SNORD3B-1*.

**Conclusions:** We find distinct clusters of cells defined transcriptionally within the pDC and SMBC lineages, and the transcripts which define these subgroups differ between cell lineages. Type I IFN induced transcripts are important to pDC diversity, while in SMBCs transcripts related to cellular activation and nucleic acid processing are critical markers of transcriptional heterogeneity.

## PgmNr 2001: Complement contributes to sex differences in risk of lupus and schizophrenia.

### Authors:

N. Kamitaki <sup>1,2</sup>; A. Sekar <sup>1,2</sup>; R.E. Handsaker <sup>1,2</sup>; H. de Rivera <sup>1,2</sup>; K. Tooley <sup>1,2</sup>; D.L. Morris <sup>3</sup>; C. Whelan <sup>1,2</sup>; P. Tomblinson <sup>3</sup>; L.M. Olde Loohuis <sup>4,5</sup>; M. Boehnke <sup>6</sup>; R.P. Kimberly <sup>7</sup>; J.B. Harley <sup>8</sup>; C.D. Langefeld <sup>9</sup>; M.T. Pato <sup>10</sup>; C.N. Pato <sup>10</sup>; R. Ophoff <sup>4,5</sup>; L.A. Criswell <sup>11</sup>; R.R. Graham <sup>12</sup>; T.J. Vyse <sup>3</sup>; S.A. McCarroll <sup>1,2</sup>; Schizophrenia Working Group of the Psychiatric Genomics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Harvard Medical School, Boston, MA, USA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Department of Medical and Molecular Genetics, King's College London, London, UK; 4) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA; 5) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA; 6) Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA; 7) Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL, USA; 8) Center for Autoimmune Genomics and Etiology (CAGE), Department of Pediatrics, Cincinnati Children's Medical Center & University of Cincinnati and the US Department of Veterans Affairs Medical Center, Cincinnati, OH, USA; 9) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA; 10) SUNY Downstate Medical Center, Brooklyn, NY, USA; 11) Rosalind Russell / Ephraim P Engleman Rheumatology Research Center, Division of Rheumatology, UCSF School of Medicine, San Francisco, CA, USA; 12) Human Genetics, Genentech Inc, South San Francisco, CA, USA

---

Many illnesses differentially affect men and women, in many cases for unknown reasons. For example, lupus affects nine times more women than men, whereas schizophrenia affects more men and tends to affect men more severely. We found that alleles that increase expression of complement component 4 (C4) genes offer strong protection from lupus while increasing risk for schizophrenia, and that both effects are stronger in men than in women. Common inherited combinations of C4A and C4B gene copy number associated with 7-fold variation in risk for lupus ( $p < 10^{-117}$ ), with C4A genes offering strong protection and C4B more modest protection, and both contributing independently of other variation in the MHC locus. The effects of C4 alleles were stronger in men, among whom common C4 alleles generated 14-fold variation in risk (vs. 4-fold among women). Similarly, alleles of C4 affected schizophrenia risk twice as strongly in men as in women. In cerebrospinal fluid and plasma, C4 and its effector protein (C3) were more abundant in adult men than adult women ( $p < 10^{-5}$ ). Sex differences in ambient levels of C4 and C3 may help explain the larger effects of C4 alleles in men, women's greater risk of SLE, and men's greater vulnerability to schizophrenia.

## PgmNr 2002: Association of FUT2 and ABO with Crohn's disease in Koreans.

### Authors:

B.M. Kim <sup>1</sup>; S.G. Jung <sup>1</sup>; H.S. Lee <sup>1</sup>; J.W. Moon <sup>1</sup>; J.W. Baek <sup>1</sup>; E.H. Oh <sup>2</sup>; S.H. Park <sup>2</sup>; S.K. Yang <sup>2</sup>; K. Song <sup>1</sup>; B.D. Ye <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine; 2) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

---

Fucosyl transferase 2 gene (*FUT2*) at 19q13 is a well-known Crohn's disease (CD) susceptibility locus in Caucasians. *FUT2* encodes the alpha-1,2-fucosyltransferase responsible for the synthesis of the H antigen, the precursor of the ABO histo-blood group antigens in body fluids and the intestinal mucosa. Homozygotes for any nonfunctional *FUT2* allele fail to express ABO antigens on their mucosa, called non-secretors. Non-secretor phenotype has been associated with CD. The two most common alleles causing the non-secretor phenotype are rs601338 (W143X) in Europeans and rs1047781 (I129F) in East Asians. The *ABO* gene encodes for glycosyltransferase enzymes which transfer specific sugar residues to the H antigen. Given that Crohn's disease is thought to arise from dysregulated mucosal immune responses to the gut flora and both the ABO blood group and the *FUT2* secretor status affect the composition of the gut microbiota, the goal of this study was to evaluate the associations of variants of *FUT2* and *ABO* with CD in Koreans. Previous candidate gene association study in a Caucasian population found no effect of *ABO* variants in CD risk, but found that non-secretors and non-O blood carriers are at higher CD risk. We examined the association of rs1047781 in 1,735 patients with CD and 8,074 healthy controls and confirmed its association at genome-wide significance:  $OR_{combined} = 1.30$ , 95% confidence interval (CI) = 1.21-1.40,  $P_{combined} = 3.52 \times 10^{-12}$ . The *ABO* locus showed genome-wide significant association with CD in Asians ( $P_{meta} = 2.35 \times 10^{-8}$ ). Following stratification on the basis of *FUT2* genotype, carriers of the secretor O blood group were significantly protective against CD compared to the secretor non-O blood group ( $OR = 0.63$ , 95% CI = 0.54-0.73,  $P = 2.86 \times 10^{-9}$ ). In conclusion, we confirmed the previously reported association of *FUT2* with CD in Korean population at genome-wide significance, and report that the O blood group and *FUT2* secretor status are protective factors against Crohn's disease in Asians.

## PgmNr 2003: Epstein-Barr virus (EBV) molecular mechanisms suggest pathogenesis for systemic lupus erythematosus.

### Authors:

V. Laurynenka<sup>1</sup>; S. Eswar<sup>1</sup>; S. Parameswaran<sup>1</sup>; X. Chen<sup>1</sup>; K. Kaufman<sup>1,5,6</sup>; B. Namjou<sup>1</sup>; L. Kottyan<sup>1,2,3,4</sup>; M. Weirauch<sup>1,2,3,5</sup>; J.B. Harley<sup>1,2,3,6</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Center for Autoimmune Genomics & Etiology (CAGE), Cincinnati Children's Hospital Medical Center; Department of Pediatrics, University of Cincinnati, Cincinnati, OH.; 2) Divisions of Immunobiology, Cincinnati Children's Hospital Medical Center; Department of Pediatrics, University of Cincinnati, Cincinnati, OH.; 3) Developmental Biology, Cincinnati Children's Hospital Medical Center; Department of Pediatrics, University of Cincinnati, Cincinnati, OH.; 4) Allergy & Immunology, Cincinnati Children's Hospital Medical Center; Department of Pediatrics, University of Cincinnati, Cincinnati, OH.; 5) Biomedical Informatics, Cincinnati Children's Hospital Medical Center; Department of Pediatrics, University of Cincinnati, Cincinnati, OH.; 6) US Department of Veterans Affairs Medical Center, Cincinnati, OH.

---

Systemic lupus erythematosus (SLE) is an autoimmune disease affecting millions worldwide and is now diagnosed 10 times more often than 50 years ago in Western countries. Genome-wide association studies (GWAS) have identified more than 500 associated single nucleotide polymorphisms. Most of these associations localize to incompletely understood regulatory regions. EBV was nominated as a potential environmental factor by immunochemistry and epidemiological studies, but the mechanism is not known.

For characterization of the intersections of transcription factor (TF) DNA binding sites at SLE risk loci we applied the RELI (Regulatory Element Locus Intersection) simulation (Nat Genet 50:699, 2018). The 347 risk alleles for SLE in European ancestry curated from published genome wide association studies (GWASs) and candidate gene studies, all at  $p < 5 \times 10^{-8}$ , were reduced to 83 loci after pruning as a consequence of linkage disequilibrium. We evaluated the available 53 virally encoded TF ChIP-seq (chromatin immunoprecipitation with DNA sequencing) datasets and complemented this analysis with the results from 11,483 human TF ChIP-seq datasets.

We found that, 52 (62.7%) of 83 the European SLE risk loci are occupied by the Epstein-Barr virus Nuclear Antigen 3C (EBNA3C) protein (OR=3.7,  $P=1.96 \times 10^{-18}$  after Bonferroni correction), EBNA2 (OR=4.2,  $P=4.17 \times 10^{-18}$ ), EBNA leader protein (EBNALP) (OR=3.3,  $P=5.53 \times 10^{-18}$ ), EBNA3B (OR=6.0,  $P=4.19 \times 10^{-18}$ ). The 3 viral (EBNALP, EBNA3C, EBNA2) and more than 300 human TFs (all with  $P < 1 \times 10^{-6}$ ) cluster together in an optimal subset of approximately 56% known loci in SLE with  $p < 1 \times 10^{-100}$ , revealing possible gene-environment interaction and identifying the genetic regulatory mechanisms that alter disease risk. More than 85% of the top 100 associated TF ChIP-seq datasets were collected from EBV transformed B cell lines in the Latency III program of viral expression, for which EBNALP, EBNA3C and B, EBNA2 are viral gene products. Therefore, these results further

nominate EBV for a role in the pathogenesis of SLE by a mechanism operating in transformed B cells through the EBV Latency III program of viral expression.

## PgmNr 2004: Shared associations across autoimmune diseases improves fine-mapping resolution in 129,124 samples.

### Authors:

M.R. Lincoln <sup>1</sup>; N. Connally <sup>1</sup>; C. Cotsapas <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology, Yale School of Medicine, New Haven, CT.; 2) Department of Genetics, Yale School of Medicine, New Haven, CT.

Genome-wide association studies in the autoimmune and inflammatory diseases (AID) have identified hundreds of susceptibility loci, but fine-mapping these to underlying causal variants remains challenging. Fine-mapping resolution is a function of statistical power and, ultimately, of sample size; meaningful resolution increase requires large increases in sample size, which are unlikely where samples and resources are limited. As an alternative approach, we sought to leverage the extensive sharing of associations between AID by combining cases and controls across diagnoses to increase sample size.

To do so we obtained and analyzed genotype data for controls and cases diagnosed with one of six AID, all genotyped on the Immunochip, a fine-mapping array densely interrogating 188 AID-associated loci. After quality control, population outlier removal, and resolution of duplicates and relatives, we analyzed data from 52,542 healthy controls and 76,582 individuals with one of celiac disease, inflammatory bowel disease, multiple sclerosis, rheumatoid arthritis, systemic lupus erythematosus or type 1 diabetes.

We computed conditionally independent case/control association statistics in each disease by grouping samples into country-level strata, performing logistic regression in each stratum with principal component covariates, and conditioning on sequentially identified lead SNPs. We then used joint likelihood mapping (JLIM) to identify conditionally independent associations shared between pairs of the six diseases, which we could combine in a cross-disease meta-analysis. Of the 269 unique associations in 77 loci that met JLIM inclusion criteria (P value thresholds and overlapping  $r^2 \leq 0.5$  windows), 50 were shared between two diseases, and 11 between at least three of six diseases (JLIM  $P < 0.05$ ). Nine loci exhibited effects that differed in direction between diseases.

In each of the 77 loci, we meta-analyzed all strata drawn from diseases sharing an effect, and assessed fine-mapping resolution increase as the change in the number of SNPs included in the 95% credible interval (95CI). In each disease separately, we observed a mean of 45.6 (s.d. 89.0) SNPs in the 95CI; after cross-disease fine-mapping, this improved three-fold to a mean 15.1 (s.d. 21.9) SNPs per 95CI.

We therefore show that effects are shared widely across AID, and that cross-disease meta-analysis can meaningfully increase fine-mapping resolution.

## PgmNr 2005: Genotypic diversity observed within a large cohort of Armenian patients with late-onset familial Mediterranean fever.

### Authors:

C. Oberkanins<sup>1</sup>; G. Kriegshäuser<sup>2,3</sup>; H. Hayrapetyan<sup>4,5</sup>; S. Atoyan<sup>4,5</sup>; S. Nemeth<sup>1</sup>; T. Sarkisian<sup>4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) ViennaLab Diagnostics, Vienna, Austria; 2) Institute of Clinical Chemistry and Laboratory Medicine, General Hospital Steyr, Austria; 3) Clinical Institute of Medical and Laboratory Diagnostics, Medical University of Graz, Austria; 4) Center of Medical Genetics and Primary Health Care, Yerevan, Armenia; 5) Department of Medical Genetics, Yerevan State Medical University, Yerevan, Armenia

---

**Introduction:** Familial Mediterranean fever (FMF) as an autoinflammatory disease, results from mutations in the *MEFV* gene mainly with an autosomal recessive mode of inheritance. The age of onset of FMF varies, with about 60% and 90% of patients experiencing their first attack before the age of 10 and 20 years, respectively. Hence, FMF with the first attack occurring at the age of  $\geq 40$  years (i.e. late-onset FMF) is rare and only a few small studies have addressed this disease subset.

**Objectives:** This work aimed at investigating the molecular genetic characteristics of Armenian patients diagnosed with late-onset FMF.

**Methods:** Genomic DNA isolated from 354 Armenian late-onset FMF patients were analysed for the 12 most common *MEFV* mutations plus SAA1 isoforms 1.1, 1.3 and 1.5 using multiplex PCR and reverse-hybridisation. Mutational spectra and resulting genotypes were then matched against the clinico-demographic profiles collected for these patients.

**Results:** Of all 354 patients, 194 (54.80%) were female and 160 (45.20%) were male. The following genotypes were significantly associated with the late-onset variant: M680I/E148Q ( $P=0.004$ ), M694V/E148Q ( $P<0.001$ ) and V726A/V726A ( $P=0.001$ ). Of note, 12/354 (3.39%) patients were found to be homozygous for the M694V mutation.

**Conclusions:** Our data suggest that late-onset FMF is more prevalent in women and is of greater genetic diversity than previously reported. Further studies including late-onset FMF patients homozygous for *MEFV* mutation M694V are ongoing and may lead to the identification of novel disease-modifying mechanisms.



## PgmNr 2006: Identifying the genetic basis of myasthenia gravis through analysis of genome-wide data.

### Authors:

A. Topaloudi <sup>1</sup>; F. Tsetsos <sup>2</sup>; Z. Zagoriti <sup>3</sup>; K. Kleopa <sup>4</sup>; G. Lagoumintzis <sup>3</sup>; E. Zamba-Papanicolaou <sup>4</sup>; Y. Christou <sup>4</sup>; K. Poulas <sup>3</sup>; S. Tzartos <sup>5,6</sup>; M. Georgitsi <sup>2,7</sup>; P. Paschou <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biological Sciences, Purdue University, West Lafayette, IN, United States; 2) Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupoli, Greece; 3) Department of Pharmacy, University of Patras, Rio, Greece; 4) Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 5) Hellenic Pasteur Institute, Athens, Greece; 6) Tzartos NeuroDiagnostics, Athens, Greece; 7) Department of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

---

Myasthenia Gravis (MG) is a heritable autoimmune disorder that affects the neuromuscular junction. One characteristic of the disease is its sex-specific onset at different ages. At early onset (<50 years old) it affects mainly females, whereas at late onset (>50 years old) it affects mostly males. The purpose of this study is to elucidate the genetic factors that contribute to the susceptibility of MG. Here, we are presenting the results of i) the largest GWAS meta-analysis for MG on the largest MG dataset to date (1183 cases and 2509 controls- including a novel dataset and two previously published datasets); ii) the first ever Copy-number Variation (CNV) analysis of rare variants on MG; and, again for the first time iii) a cross-disorder meta-analysis of MG with other autoimmune disorders (Systemic Lupus Erythematosus, Multiple Sclerosis, Autoimmune Thyroid Disease and Rheumatoid Arthritis). We observe the strongest association in the HLA-B gene (rs2905714, OR=2.235,  $p=3.02 \times 10^{-7}$ ). We also identify a novel locus (rs4142030, OR=1.946,  $p=4.9 \times 10^{-7}$ ) at the THEMIS gene. Our rare CNVs analysis detects an enrichment of CNV burden in cases for CNV count (OR 1.17 [1.13–1.22],  $p=2 \times 10^{-16}$ ) and CNV gene count (OR 1.2 [1.18–1.24],  $p=2 \times 10^{-16}$ ) metrics. In the cross-disorder meta-analysis, we detect variants significantly associated across multiple different autoimmune disorders, as well as significant novel SNPs arising in the meta-analysis but not in the individual studies. We also describe independent analyses focusing separately on early and late-onset cases, to investigate the genetic variants implicated in each onset. Our analysis confirms previous findings and reveals novel genetic variants that contribute to the development of MG. Our results shed light into the genetic architecture of MG and point to the potential of a shared genetic mechanism across different autoimmune disorders.

## **PgmNr 2007: Comprehensive analyses of genetic, epigenetic, and transcriptomic data to elucidate the mechanisms of systemic lupus erythematosus (SLE).**

**Authors:**

W. Yang; Y.F. Wang; T.Y. Wang; M. Guo; Y.L. Lau

[View Session](#) [Add to Schedule](#)

**Affiliation:** Paediatrics & Adolescent Med, Univ Hong Kong, Hong Kong, China

---

Systemic lupus erythematosus (SLE) is a prototype autoimmune disease with extreme heterogeneity. There are significant differences between different major ethnicities in disease prevalence, average onset age, and severity. Understanding the disease mechanisms and the mechanisms underlying disease association is crucial for targeted treatment of the disease and precision medicine in SLE. In a recent effort to identify susceptibility genes for SLE, we have assembled the largest sample collection on Han Chinese from several cohorts from Hong Kong, Guangzhou, and Anhui, China. Through meta-analysis of genome-wide association studies of these data, we have identified 40 new susceptibility loci for SLE reaching genome-wide significance. Interestingly, a number of these loci showed clear specificity for Chinese populations compared to European data, despite that there is adequate power in Europeans cohorts. The shared loci between Chinese and European ethnicities for SLE showed close correlation in their signature of recent positive selection measured by iHS(the integrated haplotype Score), while the loci with ethnic specificity demonstrated little correlation in their selection signature. The differences in susceptibility loci suggest that there might be differences in the disease mechanisms between major ethnicities. We present results on a comprehensive analysis of the susceptibility genes by integrating data on eQTL, gene expression profiles, cell-type specific open chromatin and histone markers for enhancers, and differential DNA methylation in SLE patients compared to that of controls. Analyses of multi-omics data allow us to have a better understanding of disease mechanisms and the mechanisms of genetic associations.

# PgmNr 2008: Electronic health records-based GWAS and PheWAS for C3 and C4 levels as markers of complement pathway activity.

## Authors:

A. Khan <sup>1</sup>; N. Shang <sup>2</sup>; L. Petukhova <sup>2</sup>; J. Zhang <sup>2</sup>; Y. Shen <sup>2</sup>; S.J. Hebring <sup>3</sup>; H. Moncrieffe <sup>4</sup>; L.C. Kottyan <sup>4</sup>; B.N. Khales <sup>4</sup>; R. Knevel <sup>5</sup>; S. Raychaudhuri <sup>6</sup>; E.W. Karlson <sup>6</sup>; J.B. Harley <sup>4</sup>; J.C. Denny <sup>7</sup>; D. Crosslin <sup>8</sup>; M.S.V Elkind <sup>2</sup>; A. Gharavi <sup>1</sup>; G. Hripcsak <sup>2</sup>; C. Weng <sup>2</sup>; K. Kiryluk <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Medicine, Columbia University , New York, New York.; 2) Columbia University , New York, New York; 3) Marshfield Clinic Research Institute, Marshfield; 4) Cincinnati Children's Hospital Medical Center; 5) Leiden University Medical Center, Leiden, Netherlands; 6) Harvard Medical School, Boston; 7) Vanderbilt University Medical Center; 8) Washington University, Seattle

## Introduction

Blood levels of C3 and C4 are used in clinical practice to screen for evidence complement activation. Here, we hypothesize that EHR-derived C3 and C4 levels can be used as quantitative endophenotypes to identify genetic regulators of the complement pathway.

## Methods

We performed a GWAS for C3 and C4 levels within the eMERGE Network, which consists of 12 medical centers contributing EHR and genotype data for 98,918 patients. C3 and C4 levels were available for a total of 3,903 and 3,952 individuals, respectively. Genetic association tests were performed using linear regression within 3 major ancestral groups. Complement levels were log-transformed and adjusted for sex, center, and genetic ancestry. Meta-analysis was performed using the inverse-variance-weighted method. We used linkage disequilibrium score regression to estimate the heritability of C3 and C4, and to test for genome-wide genetic correlations.

## Results

We detected two genome-wide significant loci for C3 levels: *CFH* locus on chr.1q31.3 (rs3753396-A, Beta=0.05, 95%CI 0.04-0.15, P=5.97x10<sup>-9</sup>) and C3 locus on chr.19p13.3 (rs11569470-A, Beta=-0.05, 95%CI -0.07 to -0.03, P=4.90x10<sup>-08</sup>). For plasma C4 levels, we confirmed large effects of the C4 copy number variant locus on chr.6p21.32 (rs3135353-C, Beta=0.17, 95%CI 0.14-0.20, P=1.24x10<sup>-36</sup>) and detected a new locus on chr.11p13 (rs58520479-G, Beta=0.08, 95%CI 0.04-0.11, P=4.35x10<sup>-08</sup>). The overall SNP-based heritability of C3 and C4 was estimated at 15% and 63%, respectively. C3 levels had negative genetic correlation with systemic lupus erythematosus (rg=-0.39) and positive correlation with CAD (rg=0.38), essential hypertension (rg=0.42), and albuminuria (rg=0.33). C4 levels had positive genetic correlation with LDL (rg=0.25), total cholesterol (rg=0.35), triglycerides (rg=0.20), CAD (rg=0.22) and essential hypertension (rg=0.20). Moreover, C4 levels had negative genetic correlation with SLE (rg=-0.77), primary sclerosing cholangitis (rg=-0.69), type 1 diabetes (rg=-0.43), celiac disease (rg=-0.51), Crohn's disease (rg=-0.26), plantar warts (rg=-0.25), scarlet fever (rg=-0.37), positive tuberculosis test (rg=-0.35), cold sores (rg=-0.34) and shingles (rg=-0.17). These results were further refined by locus-specific PheWAS analyses.

## Conclusions

Our study indicated that C3 and C4 levels have significant heritability and exhibit genetic correlations with a wide range of autoimmune, inflammatory, infectious, and cardio-metabolic traits.

## PgmNr 2009: Cross-population polygenic risk score predictions improved by prioritization of variants in predicted cell-type-specific regulatory elements.

### Authors:

T. Amariuta<sup>1,2,3,4,5</sup>; K. Ishigaki<sup>1,2,3,4,5</sup>; E. Kawakami<sup>6,7</sup>; S. Raychaudhuri<sup>1,2,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Graduate School of Arts and Sciences, Harvard University, Cambridge, Massachusetts, 02138, USA; 2) Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, 02115, USA; 3) Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA; 4) Center for Data Sciences, Harvard Medical School, Boston, Massachusetts, 02115, USA; 5) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142, USA; 6) Medical Sciences Innovation Hub Program, RIKEN, Yokohama, Japan; 7) Artificial Intelligence Medicine, Graduate School of Medicine, Chiba University, Chiba, Japan

Genome-wide association studies (GWAS) have implicated thousands of associations between complex traits and genetic variants, most of which reside in the noncoding genome and have poorly understood regulatory function. We and others have identified noncoding regulatory elements harboring large proportions of genetic variation. However, our growing knowledge of genetic susceptibility factors risks benefitting mostly European populations due to preferential bias in genetic studies. For example, polygenic risk scores (PRSs) strongly predict polygenic traits but translate poorly between ancestries, possibly due to improper modeling of ancestry-specific linkage disequilibrium (LD).

We hypothesized that PRSs would better translate between ancestries if we prioritized variants with predicted regulatory function. To this end, we identified regulatory elements associated with transcription factors (TFs), which regulate genome-wide programs such as cell-type-specific gene expression and chromatin dynamics.

We collected all 3,158 TF occupancy profiles (ChIP-seq) from NCBI. After merging experiments of the same TF and cell type, we applied our previously published strategy called IMPACT to predict cell-state-specific regulatory elements from 732 unique TF/cell-type binding profiles. To identify shared regulatory mechanisms between Europeans and East Asians, we partitioned the heritability of 12 polygenic traits with stratified-LD score regression using a European GWAS (N=38,242) and an East Asian GWAS (N=22,515). Notably, for rheumatoid arthritis (RA), we observed that 34 IMPACT annotations captured significant proportions ( $p < 0.05$ , Bonferroni corrected) of European RA heritability, 29 (74%) of which also captured East Asian RA heritability. Motivated by evidence of shared biology between populations, we selected these 29 IMPACT annotations to evaluate the predictive power of a cross-population PRS. Compared to a PRS based on all SNPs, prioritizing the top 5% of IMPACT variants resulted in a significant increase in predictive performance from European to East Asian ancestry (19%  $R^2$  increase, 96% decrease in  $R^2$  using deprioritized 95%) and vice versa (20% average  $R^2$  increase, 91% decrease in  $R^2$  using deprioritized 95%) averaged across these 29

annotations.

In conclusion, we find that our strategy may be used to improve the cross-population predictive performance of PRSs and narrow the gap between ancestry-specific disparities in genetics research and clinical practice.

## PgmNr 2010: Exploration of shared genetic susceptibility loci between type 1 diabetes and rheumatoid arthritis in the Pakistani population.

### Authors:

M.M. Aslam <sup>1,2</sup>; P. John <sup>1</sup>; K. Fan <sup>2</sup>; A. Bhatti <sup>1</sup>; E. Feingold <sup>2</sup>; F.Y. Demirci <sup>2</sup>; M.I. Kamboh <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST), Islamabad, Pakistan; 2) Department of Human Genetics, Graduate School of Public Health (GSPH), University of Pittsburgh, Pittsburgh, Pittsburgh, Pennsylvania, USA

---

**Objective.** Type 1 diabetes (T1D) and rheumatoid arthritis (RA) are autoimmune diseases. It is known that certain genetic loci and factors that increase the overall autoimmunity risk can be shared among different autoimmune diseases. We sought to replicate seven T1D-related SNPs that have been previously reported to be associated with RA susceptibility in a small set of mixed family-based and case-control Pakistani sample in a relatively large and independent RA case-control sample from the same population.

**Methods.** Seven T1D-associated SNPs (*GLIS3*/rs7020673, *BACH2*/rs11755527, *SKAP2*/rs7804356, *GDSMB*/rs2290400, *C6orf173*/rs9388489, *LOC399716*/rs947474 and *DLK1-MEG2*/rs941576) were genotyped in a large Pakistani RA case-control sample (n=1,959) using TaqMan® SNP genotyping assays.

**Results.** None of the tested SNPs showed statistically significant association with RA susceptibility, however, one SNP (*GLIS3*/rs7020673) showed a trend for association (OR= 0.88, p=7.99E-02). *GLIS3* is expressed in the cartilage and regulated by suppressor of the fused homologous gene that encodes a negative regulator of hedgehog signaling.

**Conclusion.** Our study has failed to replicate the previously reported association of seven T1D-associated SNPs with RA risk in a large sample from the same population. Thus, our results do not support a major role of these T1D SNPs in affecting RA susceptibility in the Pakistani population.

## PgmNr 2011: Multi-ancestry fine mapping of interferon lambda and the outcome of acute hepatitis C virus infection.

### Authors:

C. Vergara <sup>1</sup>; P. Duggal <sup>1</sup>; A. Valencia <sup>2,3</sup>; G. Wojcik <sup>4</sup>; M. Taub <sup>1</sup>; HCV Genetics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetic Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA; 2) Johns Hopkins University, School of Medicine, Baltimore, MD, USA; 3) Universidad Pontificia Bolivariana, Medellin, Colombia.; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

---

Clearance of acute infection with hepatitis C virus (HCV) is associated with the presence of the TT allele of the rs368234815 (TT/ΔG) dinucleotide polymorphism in *IFNL4* on chromosome 19q13.13. We aimed to fine-map this region to detect possible causal variants that may contribute to clearance of HCV infection. First, we performed targeted sequencing of the genomic segment containing the *IFNL1-IFNL4* genes in 64 individuals sampled according to rs368234815 genotype: TT/clearance (N=16) and ΔG/persistent (N=15) (genotype-outcome concordant) or TT/persistent (N=19) and ΔG/clearance (N=14) (discordant). Sequencing primers were designed to precisely map paired reads to increase accuracy of genotyping calls in this complex region. We identified 25 SNPs with a difference in counts of alternative allele > 5 between clearance and persistence individuals. Then, we evaluated those markers in an association analysis of HCV clearance conditioning on rs368234815 in two independent groups of European (N=1717, 692 clearance/1025 persistence) and African ancestry (N=1835, 320 clearance/1515 persistence) individuals and meta-analyzed the results. Logistic regression adjusting for HIV status and three principal components of ancestry was performed in each ancestry group and meta-analyzed using a fixed effect model. 10/25 variants located in the *IFNL3-IFNL4* region were associated with HCV clearance ( $P < 0.05$ ) in the conditioned analysis with lead SNPs rs4803221 (Meta-analysis  $P=4.9 \times 10^{-4}$ ) and rs8099917 ( $P=5.5 \times 10^{-4}$ ). In the European ancestry group, the individuals carrying the haplotype rs368234815ΔG/rs4803221C were 1.7x more likely to spontaneously clear HCV than those with the rs368234815ΔG/rs4803221G haplotype (95% CI: 1.3-2.3,  $P=3.6 \times 10^{-5}$ ). For another nearby SNP, the haplotype of rs368234815ΔG/rs8099917T was significantly associated with HCV clearance compared to rs368234815ΔG/rs8099917G (OR: 1.6, 95% CI: 1.3-2.2,  $P$  value:  $1.8 \times 10^{-4}$ ). We identified four variants as possible causal variants in the region: rs368234815, rs12982533, rs10612351 and rs4803221 (posterior probability 0.27-0.61). Our results are consistent with a main signal of association being shared across populations driven by one or more functional variants represented by rs368234815, with additional contributions from rs4803221, and/or nearby SNPs in linkage disequilibrium including rs8099917 in the European ancestry population.



## PgmNr 2012: Fine mapping the 17q12-21 asthma locus using whole genome sequences from African Americans.

### Authors:

C. Washington; A. Biddanda; C. Ober

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, University of Chicago, Chicago, Illinois.

---

Asthma is a common complex disease characterized by lower airway inflammation, airflow obstruction and symptoms of wheezing, coughing, and breathlessness. Asthma GWASs have identified over 60 childhood-onset asthma loci, including a locus on chromosome 17q12-21 that is the most highly replicated and most significant locus identified in European ancestry (EA) populations. However, associations between SNPs at this locus and asthma in African Americans (AAs) have been equivocal, despite increased asthma severity and prevalence in AAs. One potential reason for this discrepancy may be due to the extensive linkage disequilibrium (LD) block at this locus in EA populations (~200kb) being significantly less in AAs. As a result, SNPs tagging causal risk alleles in EA populations may not tag causal risk alleles in AAs. Here, we sought to characterize the genetic variation and haplotype structure at this locus in whole genome sequences from EA and AA individuals. To this end, we curated publicly available data from the Consortium on Asthma in African-ancestry Populations in the Americas (CAAPA) and the EVE Consortium; 258 AA and 100 EA individuals with asthma were included in this study. We first phased variants across the locus using the Michigan Imputation Server and then used ChromoPainter to assign haplotypes. We used 5 SNPs that tagged the risk and protective haplotypes in EA populations to determine the number of additional recombinant haplotypes in AAs. As expected in the EA sample, only two common (>5% frequency) 5-SNP haplotypes were observed: one corresponding to asthma risk and one to asthma protection (59.0% and 34.5%, respectively); the remaining 6.5% included 5 additional rare (recombinant) haplotypes. In contrast, we observed 5 common and 8 rare haplotypes in AAs. The EA risk and protective haplotypes accounted for 41.7% and 10.7%, respectively of AA haplotypes, with 3 additional common haplotypes occurring at 17.6%, 15.7%, and 5.4%, and 8 rare haplotypes accounting for the remaining 8.9%. One common haplotype (15.7%) was absent in the EA sample. The decreased frequency of the protective haplotype in AAs compared to the EA sample (10.7% v. 34.5%) may contribute to the increased prevalence of asthma in AAs, and the high frequency of recombinant haplotypes could potentially contribute to increased severity. CW is supported by the Howard Hughes Medical Institute through the James H. Gilliam Fellowships for Advanced Study program.

## PgmNr 2013: Trans-ancestral dissection of urate and gout associated major loci *SLC2A9* and *ABCG2* reveals primate-specific regulatory effects.

### Authors:

W. Wei <sup>1</sup>; R. Takei <sup>1,2</sup>; M. Cadzow <sup>2</sup>; D. Markie <sup>3</sup>; M. Bixley <sup>2</sup>; C. Li <sup>4,5</sup>; H.K. Choi <sup>6</sup>; Z. Li <sup>7,8</sup>; H. Hu <sup>9</sup>; H. Guo <sup>10</sup>; M. He <sup>9</sup>; Y. Shi <sup>4,5,7,8</sup>; S. Robertson <sup>1</sup>; T.R. Merriman <sup>2</sup>; Eurogout Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Women's and Children's Health, University of Otago, Dunedin, Otago, New Zealand; 2) Department of Biochemistry, University of Otago, Dunedin, Otago, New Zealand; 3) Department of Pathology, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand; 4) Shandong Gout Clinical Medical Center, Qingdao 266003, China; 5) Shandong Provincial Key Laboratory of Metabolic Disease, The Affiliated Hospital of Qingdao University, Qingdao 266003, China; 6) Massachusetts General Hospital, Harvard Medical School, Boston, USA; 7) Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai 200030, China; 8) Institute of Social Cognitive and Behavioral Sciences, Shanghai Jiao Tong University, Shanghai 200240, China; 9) Department of Occupational and Environmental Health and the Ministry of Education Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430030, Hubei, China; 10) Center for Biostatistics, School of Health Sciences, The University of Manchester, Manchester, UK

**Background:** Gout is a complex inflammatory arthritis affecting ~20% of people with an elevated serum urate level (hyperuricemia). Gout and hyperuricemia are essentially specific to humans and other higher primates, with varied prevalence across ancestral groups. *SLC2A9* and *ABCG2* are major loci associated with both urate and gout in multiple ancestral groups, with abundant epistatic interactions observed in urate particularly in the *SLC2A9* adjacent region. However, fine-mapping has been challenging due to extensive linkage disequilibrium underlying the associated regions.

**Results:** Leveraging trans-ancestral meta-analyses, we generated single-variant resolution mapping for *SLC2A9* (rs3775948 for urate and rs4697701 for gout) and *ABCG2* (rs2622621 for gout). Both *SLC2A9* and *ABCG2* were heavily involved in chromatin interactions with multiple regulators including super-enhancers in a tissue-specific manner, and colocalized with ten and three primate-specific elements respectively. The *SLC2A9* variant rs4697701 was within an ancient enhancer whereas the *ABCG2* variant rs2622621 was within a primate-specific transposable element. Further, the *SLC2A9* super-enhancers and primate-specific elements were all within the *SLC2A9*-adjacent epistatic interaction region. The epistatic interactions were statistically replicated in the gout cohorts, and supported by recent functional evidence of enhancer-promoter interactions and promoters with distal enhancer effects in human cell lines.

**Conclusions:** The interplay of common functional variants and multiple level regulators drives the associations in *SLC2A9* and *ABCG2* and regulates their urate transport function in a primate-specific manner.

## PgmNr 2014: Identification of rare variants contributing to disseminated coccidioidomycosis susceptibility after infection.

### Authors:

J. Berghout<sup>1,2</sup>; G.R. Thompson<sup>3,4</sup>; D. Bays<sup>4</sup>; S. Waldman<sup>4</sup>; E. Stewart<sup>4</sup>; A.P. Hsu<sup>5</sup>; D.A. Powell<sup>6</sup>; A.L. Chaput<sup>6</sup>; N. Pouladi<sup>1,2</sup>; F. Vitali<sup>1,2</sup>; L.F. Shubitz<sup>6,7</sup>; S.M. Holland<sup>5</sup>; J.A. Frelinger<sup>6</sup>; J.N. Galgiani<sup>6</sup>; Y.A. Lussier<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Biomedical Informatics and Biostatistics, The University of Arizona, Tucson, Arizona.; 2) Department of Medicine, University of Arizona, Tucson, Arizona; 3) Department of Medical Microbiology and Immunology, University of California Davis, Sacramento, CA; 4) Department of Medicine, Division of Infectious Disease, University of California Davis, Sacramento, CA; 5) Immunopathogenesis Section, National Institutes of Allergy and Infectious Disease, National Institutes of Health, Bethesda, MD; 6) Valley Fever Center for Excellence, University of Arizona, Tucson, AZ; 7) Animal and Comparative Biomedical Science, University of Arizona, Tucson, AZ

**Introduction:** Disseminated coccidioidomycosis (DCM) is a rare and serious complication of infection with the soil-resident fungal pathogens *Coccidioides immitis* or *C. posadasii*. The majority of symptomatic infections (est. 50,000/yr) are limited to the lung (pulmonary coccidioidomycosis; PUL), presenting as community acquired pneumonia and self-resolving. However, in 600-1000 cases per year, the infection spreads to skin, liver, spleen, bone, and brain, among others. Genetic variation and ancestry are major risk factors for dissemination in immunocompetent patients, with DCM observed in 7-50% of African-American patients in contrast to ~1-2% of Caucasian or Hispanic patients. However, the mechanism(s) underlying dissemination remain unknown. We hypothesized that genetic susceptibility factors would be infrequent pathogenic variants with allele and gene-level heterogeneity that functionally converge on shared biological processes and pathways that are differentially burdened between DCM cases and PUL controls.

**Methods:** We recruited the largest ever cohort of patients for investigation of the genetics underlying coccidioidomycosis pathophysiology, comprised of 535 putatively immunocompetent patients from Arizona and California (n=147 DCM; n=388 PUL) with IRB-approved informed consent for genetic analysis, rigorous clinical screening, and laboratory-confirmed coccidioidomycosis. DNA was extracted from blood, and exomes were sequenced by the Broad Institute. Variants were jointly called across all samples using Genome Analysis Toolkit and resulting VCF files were filtered using VCF tools for variant call quality, read depth, and allele frequency. Processed files were annotated using SnpEff/SnpSift and additional analyses performed in R.

**Results:** 546,022 qualifying variants were observed in the complete dataset including rare mutations in candidate immune response genes in DCM cases, and genes previously implicated in susceptibility to fungal infections from immunocompromised patient case studies (IFNGR1, STAT1, IL12RB1), and in animal models (CLEC7A/DECTIN1, CARD9). Work is ongoing to control for demographic variation and confounding, and future work on this dataset will include gene set burden testing using sequence kernel analysis (SKAT-O).

**Conclusion:** This preliminary work describes a new genetically characterized cohort for disseminated

coccidioidomycosis and suggests that DCM is associated with discrete mutations in genes associated with immune function.

## PgmNr 2015: The NIAID Centralized Sequencing Initiative: An integrated framework for research sequencing with clinical validation and return of results for immune disorders.

### Authors:

J. Yan <sup>1</sup>; M. Similuk <sup>1</sup>; L. Jamal <sup>1</sup>; H. Hullfish <sup>1</sup>; K. Jevtich <sup>1</sup>; C. Hong <sup>2</sup>; P. Littel <sup>1</sup>; C. Lowe <sup>1</sup>; X. Cheng <sup>3</sup>; S. Huse <sup>4</sup>; G. Fahle <sup>5</sup>; A. Oler <sup>3</sup>; S. Xirasagar <sup>3</sup>; M. Walkiewicz <sup>1</sup>; S. Holland <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland.; 2) Clinical Genomics Section, National Human Genome Research Institute National Institutes of Health, Bethesda, Maryland.; 3) Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland.; 4) NIAID Collaborative Bioinformatics Resource, Frederick National Laboratory for Cancer Research; 5) Department of Laboratory Medicine, National Institutes of Health Clinical Center, Bethesda, Maryland.

---

**Aims:** Immunological disorders encompass a wide spectrum of phenotypes of both multifactorial and complex Mendelian etiology characterized by reduced penetrance and variable expressivity. Efforts to understand these disorders require multifaceted approaches to phenotypic refinement, variant interpretation, and functional studies in a large cohort of patients. In light of these needs, the Centralized Sequencing Initiative (CSI) was established at the National Institute of Allergy and Infectious Diseases (NIAID) to provide streamlined genomic and phenotypic data generation, bioinformatics, quality control, clinical variant interpretation, results reporting, and genetic counseling in order to facilitate gene discovery and patient care for patients with immunological conditions at the National Institutes of Health (NIH).

**Methods:** Process development of the protocol included hiring clinical and research staff, creating a clinic scheduling and referral process from 35 protocols at NIAID studying a diverse group of immune disorders, engaging in education and outreach, collaboration with multiple teams for data processing and analyses, and integrating exome sequencing orders and results into the NIH electronic health record (EHR).

**Results:** Since the start of the protocol two years ago, 2029 patients have been enrolled with 448 patients receiving reports uploaded into the EHR. Informed consent is obtained in person at the NIH or remotely. Research exome sequencing is followed by clinical Sanger validation in a CLIA CAP laboratory at the Department of Laboratory Medicine at the NIH, with a subset of patients undergoing clinical CMA optimized for exonic coverage of immune system genes. Phenotyping using HPO terms in PhenoTips, integrated with genomic data analysis using Seqr, is enabled by the NIAID Genomic Research Integration System (GRIS). All analyses are completed by a clinical molecular geneticist, followed by review and possible functional studies with principal investigators at NIAID. Results include primary and secondary findings. Patients receive genetic counseling and coordination of care upon clinical confirmation. Data is shared in order to facilitate discovery of monogenic and polygenic

contributions to immune system disorders.

Conclusion: By integrating multiple systems at the NIH, this protocol represents a systematic multidisciplinary approach for large-scale reliable and consistent research sequencing with clinical validation and return of results.

## PgmNr 2016: Integrative analysis of disease association and blood eQTL data of Korean patients with Crohn's disease identifies *TNFSF15* and *GPR35* at two reported loci.

### Authors:

S. Jung <sup>1</sup>; W. Liu <sup>2</sup>; J. Baek <sup>1</sup>; J. Moon <sup>1</sup>; B. Ye <sup>3</sup>; H. Lee <sup>1</sup>; S. Park <sup>3</sup>; S. Yang <sup>3</sup>; B. Han <sup>4</sup>; J. Liu <sup>2</sup>; K. Song <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Korea Korea, Republic of; 2) Human Genetics, Genome Institute of Singapore, Singapore; 3) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 4) Department of Medicine, Seoul National University College of Medicine, Seoul, Korea

---

Recent genome-wide association studies (GWAS) have identified over 200 loci that are associated with inflammatory bowel diseases, but the mechanisms by which these variants influence the disease are largely unknown. Recently, methods have been developed to predict disease associated genes by integrating GWAS and expression quantitative trait locus (eQTL) data. As integration of SNPs in established loci and eQTL data for Asian Crohn's disease (CD) has never been conducted before, we first constructed eQTL database using RNA sequencing data of whole blood tissue from 101 Korean patients with CD. We identified 135,164 cis-eQTL and 3,816 eGenes with FDR less than 0.05. A significant proportion (96.5-98.7 %) of overlapped eGenes with those identified in previous eQTL studies showed the same direction of allelic effects. The significantly enriched pathways of these 3,816 eGenes include neutrophil degranulation, small molecule biosynthetic process, etc. We then used eQTL and GWAS CAusal Variants Identification in Associated Regions (eCAVIAR) to colocalize of lead SNPs in CD GWAS (899 cases and 3,805 controls) and cis-eQTL analysis of 101 CD patients. The integrated analysis identified likely target genes, *TNFSF15* and *GPR35*, at two reported loci with co-localization posterior probability (CLPP) greater than 0.01 and 0.95 for total credible set posterior probability. Their risk alleles were associated with lower expression level of *TNFSF15* and *GPR35*, respectively. This resource would be useful to understand of gene expression regulation and prioritize findings from genome-wide association studies in Asians. Our eQTL browser can be accessed at "http://asan.crohneqtl.com/".

## PgmNr 2017: Multiple HLA loci contribute to spontaneous clearance of acute hepatitis C virus infection in populations of diverse ancestry.

### Authors:

A. Valencia-Duarte <sup>1,2</sup>; C. Vergara <sup>3</sup>; C. Thio <sup>1</sup>; N. Vince <sup>4,5</sup>; P.-A. Gourraud <sup>4,5</sup>; V. Douillard <sup>4,5</sup>; M. Carrington <sup>6,7</sup>; P. Duggal <sup>3</sup>; HCV Genetics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Johns Hopkins University, Baltimore, USA; 2) Universidad Pontificia Bolivariana, School of Medicine, Medellin, Colombia; 3) Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA; 4) Université de Nantes, INSERM, Centre de Recherche en Transplantation et Immunologie, UMR 1064, ATIP-Avenir, Equipe 5, Nantes, France; 5) CHU de Nantes, Institut de Transplantation Urologie Néphrologie (ITUN), Nantes, France; 6) Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, MD, USA.; 7) Ragon Institute of MGH, MIT and Harvard, Cambridge, MA, USA.

---

Hepatitis C virus (HCV) is a major cause of chronic liver disease, hepatocellular carcinoma, and end-stage liver disease. Genome-wide association studies have identified three loci with independent effects on HCV spontaneous clearance, Major Histocompatibility Complex (MHC), interferon lambda locus *IFNL4-IFNL3* (19q13.2) and G-protein-coupled receptor 158 gene at 10p12.1, in both African ancestry and European ancestry populations. We fine-mapped the MHC region to determine the contribution of classical alleles and amino acid changes for Class I and II HLA. 1600 individuals of European ancestry (594 clearance/1006 persistence) and 1869 individuals of African ancestry (338 clearance/1526 persistence) were included. Samples were genotyped with Illumina Omni1-Quad BeadChip and imputation of classical HLA alleles and amino acids was performed using HIBAG. We tested for association with SNPs, classical alleles and amino acids using logistic regression models, adjusting for HIV status and ancestry principal components, and meta-analyzed the results using META. The top association signal was located upstream of the Class II genes *DQB1* and *DQA2* (rs2647011 C>A,  $P_{\text{meta}}=1.1 \times 10^{-18}$ ,  $OR_{\text{EA}}: 0.59$   $P=3.78 \times 10^{-11}$ ;  $OR_{\text{AA}}: 0.57$   $P=3.7 \times 10^{-9}$ ). Two classical HLA alleles were independently associated with clearance in that region in both populations: *DQB1*\*03:01 ( $P_{\text{meta}}=3.8 \times 10^{-12}$ ,  $OR_{\text{EA}}: 1.61$ ,  $P=7.6 \times 10^{-8}$ ,  $OR_{\text{AA}}: 1.62$ ,  $P=9.5 \times 10^{-6}$ ) and *DRB1*\*01:01 ( $P_{\text{meta}}=1.47 \times 10^{-7}$ ,  $OR_{\text{EA}}: 1.68$ ,  $P=2.0 \times 10^{-4}$ ,  $OR_{\text{AA}}: 2.36$ ,  $P=2.0 \times 10^{-4}$ ). In *DQB1* the residues 9, 13, 26, 45, 53, 55, 71 and 87 were significantly associated ( $P_{\text{meta}} < 4 \times 10^{-4}$ ) with either spontaneous clearance or persistence of HCV infection. In *DRB1*, there were 13 residues associated with infection outcome, of which 71 and 86 were most significantly associated, but neither 71 nor 86 explained the *DRB1*\*01:01 signal. Global association models based on long range Ancestral HLA haplotypes will be included in the analysis. Our findings suggest that two classical alleles completely account for the main association of HCV clearance with the HLA locus observed in populations of European and African ancestry. We are evaluating the associated amino acids individually and together to determine their contributions. This will be critical for vaccine design and assessment of vaccine efficacy at preventing chronic infection.



## PgmNr 2018: Genomic architecture of Stevens-Johnson syndrome from whole-genome sequencing.

### Authors:

Y. Kawai <sup>1,2</sup>; Y. Hitomi <sup>2,3</sup>; M. Ueta <sup>4</sup>; S.S. Khor <sup>1,2</sup>; K. Nakatani <sup>2</sup>; C. Sotozono <sup>5</sup>; S. Kinoshita <sup>4</sup>; M. Nagasaki <sup>6,7</sup>; K. Tokunaga <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Medical Science Project, National Center for Global Health and Medicine, Tokyo, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo; 3) Department of Microbiology, Hoshi University School of Pharmacy and Pharmaceutical Sciences, Tokyo, Japan; 4) Department of Frontier Medical Science and Technology for Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 5) Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 6) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 7) Human Biosciences Unit for the Top Global Course Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Kyoto, Japan

---

Stevens-Johnson syndrome (SJS) and its severe condition with extensive skin detachment and a poor prognosis, toxic epidermal necrolysis (TEN) are immunologically mediated acute inflammatory vesiculobullous reactions on the skin and mucous membranes. Previous studies utilizing SNP array-based genotyping identified the susceptibility loci of cold medicine-related SJS/TEN with severe ocular complications (CM-SJS/TEN with SOC). However, little is known about the contribution of rare variants and structural variants to development of CM-SJS/TEN with SOC. We performed the whole-genome sequencing of 133 patients with CM-SJS/TEN with SOC and 418 healthy control subjects to discover the broad spectrum of genetic variation. Whole-genome resequencing discovered 18.8 millions SNPs, 1.7 million insertion/deletions, 670 thousand microsatellites and 10 thousands large-scale structural variants after filtrations. Genome-wide association tests on these variants identified the association of microsatellite polymorphism near CDH12 gene and the aggregation of rare coding variants on TRPM8 gene. Majority of the common variants were found in non-coding regions suggesting the regulatory role of genetic variations in the pathogenesis of CM-SJS/TEN with SOC.

## PgmNr 2019: Multiethnic GWAS identifies novel genetic regulators of the IgA system with relevance to human disease.

### Authors:

L. Liu <sup>1</sup>; A. Khan <sup>1</sup>; Y. Li <sup>1</sup>; N. Steers <sup>1</sup>; O. Balderes <sup>1</sup>; J. Zhang <sup>1</sup>; P. Krithivasan <sup>1</sup>; Z. Moldoveanu <sup>2</sup>; B. Julian <sup>2</sup>; R. Wyatt <sup>3</sup>; H. Suzuki <sup>4</sup>; J. Xie <sup>5</sup>; N. Chen <sup>5</sup>; H. Zhang <sup>6</sup>; L. Hammarström <sup>7</sup>; A. Viktorin <sup>8</sup>; P. Magnusson <sup>8</sup>; N. Shang <sup>9</sup>; G. Hripacsak <sup>9</sup>; C. Weng <sup>9</sup>; T. Rundek <sup>10</sup>; M. Elkind <sup>11</sup>; E. Oelsner <sup>1</sup>; R. Barr <sup>12</sup>; I. Ionita-Laza <sup>13</sup>; J. Novak <sup>2</sup>; A. Gharavi <sup>1</sup>; K. Kiryluk <sup>1</sup>; eMERGE Network Investigators, STOP-IGAN Study Investigators

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Vagelos College of Physicians & Surgeons, Columbia University, New York, USA.; 2) Department of Microbiology and Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA.; 3) Div. of Pediatric Nephrology, University of Tennessee Health Sciences Center, Memphis, Tennessee, USA.; 4) Department of Nephrology, Juntendo University Faculty of Medicine, Tokyo, Japan.; 5) Department of Nephrology, Institute of Nephrology, Shanghai Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.; 6) Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China.; 7) Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden.; 8) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.; 9) Department of Biomedical Informatics, Vagelos College of Physicians & Surgeons, Columbia University, New York, USA.; 10) Department of Neurology, University of Miami, Miami, Florida, USA.; 11) Department of Neurology, Vagelos College of Physicians & Surgeons, Columbia University, New York, USA.; 12) Department of Medicine, Vagelos College of Physicians & Surgeons and Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, USA.; 13) Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, USA.

---

**Background:** Immunoglobulin A (IgA) is one of the major effectors of the adaptive immune system and plays a role in several common diseases, including mucosal infections, inflammatory bowel disease, celiac disease, and IgA nephropathy.

**Method:** We performed a multiethnic GWAS for serum IgA levels in 41,263 individuals of European, African, East Asian and Hispanic/Latino ancestry. The log-transformed IgA levels were tested using linear model controlling for age, sex, and genetic ancestry. A meta-analysis with dense imputation across 18 cohorts was performed using a fixed-effects model, followed by fine-mapping of significant loci, and co-localization with blood eQTLs. The genetic relationships between serum IgA levels and human disease traits were tested using genome-wide genetic correlation analyses. Additionally, we performed PheWAS for genome-wide polygenic score for IgA levels within the eMERGE Network, which consists of 12 medical centers contributing EHR and genotype data for a total of 98,918 patients.

**Results:** African ancestry was associated with higher serum IgA levels, while European ancestry was associated with lower levels compared to other ancestries. This was confirmed by genetic admixture analysis in the US-based cohorts. In the trans-ethnic GWAS analysis, we replicated 9 known loci and

discovered 10 novel loci associated with serum IgA levels. Cumulatively, these loci explained 1.2% of variance in serum IgA levels. Co-localization analysis with whole blood eQTL dataset revealed 9 loci with strong co-localization (posterior probability > 0.8), prioritizing several novel candidate genes. Genome-wide genetic-correlation analysis pointed to significant correlations between serum IgA levels and IgA nephropathy ( $r^2=0.32$ ), multiple sclerosis ( $r^2=0.23$ ), type 2 diabetes ( $r^2=0.17$ ), and body mass index ( $r^2=0.13$ ), and negative correlations with celiac disease ( $r^2=-0.27$ ), inflammatory bowel disease ( $r^2=-0.12$ ), as well as infections, including bacterial meningitis ( $r^2=-0.35$ ), rubella ( $r^2=-0.22$ ) and shingles ( $r^2=-0.23$ ).

**Conclusion:** We report a multiethnic GWAS for serum IgA levels with discovery of 10 novel loci. Our findings provide novel insights into the genetic regulation of the IgA system and its potential role in human disease. While individuals who inherit higher levels of IgA may be protected against pathogens and gut inflammation, they may be at a higher risk of IgA nephropathy, multiple sclerosis, diabetes and obesity.

## PgmNr 2020: Multivariate analysis of deep phenotype data: A novel approach to discovery.

### Authors:

J. Partanen<sup>1</sup>; S. Ruotsalainen<sup>1</sup>; I. Surakka<sup>1,2</sup>; M. Pirinen<sup>1,3,4</sup>; C. Benner<sup>1</sup>; A. Cichonska<sup>1,5,6</sup>; M. Salmi<sup>7,8</sup>; S. Jalkanen<sup>7,8</sup>; V. Salomaa<sup>9</sup>; A. Palotie<sup>1,10,11</sup>; M.J. Daly<sup>1,10,11</sup>; S. Ripatti<sup>1,3,10</sup>; J. Koskela<sup>1,10</sup>; FinnGen

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland; 2) Department of Internal Medicine, University of Michigan, Ann Arbor, MI, US; 3) Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland; 4) Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland; 5) Department of Computer Science, Helsinki Institute for Information Technology HIIT, Aalto University, Espoo, Finland; 6) Department of Future Technologies, University of Turku, Turku, Finland; 7) Medicity Research Laboratory, University of Turku, Turku, Finland; 8) Institute of Biomedicine, University of Turku, Turku, Finland; 9) National Institute for Health and Welfare, Helsinki, Finland; 10) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, US; 11) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

---

Multivariate methods increase power of association detection. We aimed to discover disease associations by analyzing deep phenotype data using multivariate GWAS followed by finemapping and phenome-wide analysis of proposed causal variants.

Assessing 66 quantitative traits in the population-based FINRISK study, hierarchical clustering identified one highly correlated cluster of 12 inflammatory markers – mainly interleukins and growth factors (n=6,890) – for which we performed multivariate GWAS applying Canonical Correlation Analysis. We then performed a phenome-wide association study (PheWAS) in the FinnGen study (n=135,638) across 1,801 disease endpoints for variants suggested causal by a newly developed multivariate finemapping approach.

Within the multivariate results we identified 11 independent genome-wide significant loci, 3 of which (*F5*, *ABO*, *C1orf140*) were not detected by univariate analyses. 8 loci had previously been associated with at least one of the cluster traits while 3 were novel. Multivariate finemapping identified 18 variants as likely causal variants underlying the 11 loci. These variants were associated with 25 disease endpoints in FinnGen ( $p < 1e-4$ ), spread across 5 of the 18 variants. A majority of these disease associations (19/25) were for two variants located in the *F5* and *ABO* loci. These 19 associations would have gone undetected had we used univariate GWAS. Neither of these two variants had previously reported disease associations in GWAS catalog. In the *F5* locus rs3820060 was protective (OR=0.91,  $p=5.9e-5$ ) for venous thromboembolism (VTE) and increased the risk of demyelinating disease (OR=1.52,  $p=1.6e-5$ ). The effect was independent of the well-known *F5* Leiden mutation (rs6025) and the protective effect for VTE was replicated in UK Biobank (UKBB). In the *ABO* locus rs550057 increased the risk of 17 diseases in FinnGen, including 11 diseases related to VTE, pulmonary embolism or other disorders of veins as well as 6 other diseases, such as heart failure and endometriosis. Most of these associations (13/17) were replicated in UKBB ( $p < 1e-6$ ). These findings

suggest an even broader scope of functions phenome-wide for *ABO* and *F5* than previously described.

Our results demonstrate that compared to univariate analyses, multivariate analysis of biomarker data reveals a considerably increased amount of genetic associations with several diseases, and combined with large biobank-based PheWAS provides a new approach to disease association discovery.

## PgmNr 2021: Genome-wide association study for serum galactose-deficient IgA1 in IgA nephropathy.

### Authors:

Y. Wang <sup>1,2</sup>; X. Zhou <sup>1,2</sup>; H. Zhang <sup>1,2</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Renal Division, Peking University First Hospital, No.8 Xi Shi Ku Street, Beijing, China; 2) Peking University Institute of Nephrology, No.8 Xi Shi Ku Street, Beijing, China

**Objectives:** IgA nephropathy (IgAN) is the most common primary glomerulonephritis worldwide. Galactose-deficient IgA1 (Gd-IgA1) plays a key role in the pathogenesis of IgAN. Although the heritability of serum Gd-IgA1 levels is high (ranging from 54% to 80%), the genetic association between Gd-IgA1 and IgAN has not yet been clearly determined. To further identify novel susceptibility loci, we carried out a genome wide association study (GWAS) for serum Gd-IgA1 levels in IgAN patients.

**Methods:** We performed a quantitative trait GWAS for serum Gd-IgA1 levels, with discovery and follow-up in 1,127 IgAN patients in a Chinese population. Gd-IgA1 levels were measured using a *Helix aspersa* lectin-based ELISA method. The mRNA levels of susceptibility genes in peripheral blood mononuclear cells (PBMCs) were evaluated by mRNA microarrays (Affymetrix PrimeView Human Gene Expression Array and Illumina HT-12 v4 Expression BeadChip). To search for genotype-phenotype associations in IgAN, linear regression was used for quantitative traits and logistic regression for binary and ordinal variables.

**Results:** We identified two loci passing genome-wide significance, including *GALNT12* ( $P = 1.67 \times 10^{-8}$ , Beta = 0.68) and *C1GALT1* ( $P = 3.10 \times 10^{-8}$ , Beta = 0.24). Additionally, we confirmed reported association of *C1GALT1* with serum Gd-IgA1 levels, including rs1008897 ( $P = 9.75 \times 10^{-3}$ ) and rs13226913 ( $P = 3.89 \times 10^{-2}$ ), which are common variants in Europeans but rare in East Asians (MAF 34% vs. 5% and 58% vs. 7%). *C1GALT1* variant associated in our study is in partial linkage disequilibrium with rs1008897 ( $D' = 0.92$ ,  $r^2 = 0.07$ ) and rs13226913 ( $D' = 0.44$ ,  $r^2 = 0.02$ ). Compared with healthy controls ( $n = 61$ ), *GALNT12* and *C1GALT1* showed lower mRNA expression in PBMCs from IgAN patients ( $n = 94$ ) (0.86-fold change,  $P = 1.00 \times 10^{-6}$  and 0.90-fold change,  $P = 1.92 \times 10^{-4}$ , respectively). Sub-phenotype analysis showed that the risk allele of *GALNT12* variant was associated with decreased serum C3 levels ( $P = 0.02$ , Beta = - 0.05).

**Conclusion:** Our study identifies two loci, which encode two enzymes (*GALNT12* and *C1GALT1*) involved in O-linked glycosylation, are associated with serum levels of Gd-IgA1 in IgAN. Down-regulation of these two enzymes may contribute to the generation of aberrantly glycosylated IgA1 in IgAN.

## PgmNr 2022: Study of differences in susceptibility genes for high-altitude polycythemia in Highland people.

### Authors:

L.L. Kang <sup>1,2,3</sup>; Z.Y. Zhang <sup>1,2</sup>; L.J. Liu <sup>1,2</sup>; F.C. Zhao <sup>1,2</sup>; L.F. Ma <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Medicine, Xizang Minzu University, Xianyang, Shaanxi, China; 2) Key Laboratory for Molecular Genetic Mechanisms and Intervention Research on High Altitude Disease of Tibet Autonomous Region,; School of Medicine, Xizang Minzu University, Xianyang, 712082,;Shaanxi, China; 3) Key Laboratory of High Altitude Environment and Genes Related to Diseases of Tibet Autonomous Region, School of Medicine, Xizang Minzu University, Xianyang, 712082,;Shaanxi, China

---

**Background:** High-altitude polycythemia (HAPC) is a chronic high-altitude disease characterized by excessive proliferation of red blood cells. HAPC primarily affects people who for prolonged periods live in hypoxic environments. Common over time. The most frequent symptoms of HAPC include headache, dizziness, breathlessness and sleep disorders. Although chronic hypoxia is the leading cause of HAPC, the fundamental pathophysiologic processes and molecular mechanisms responsible for its development are unclear. We, therefore, sought to explore the hereditary factors related to HAPC susceptibility in highland populations.

**Methods:** This case-control association study involved 141 people (70 patients and 70 healthy controls). In our study, we aimed to investigate the associations between susceptibility to HAPC and 529 candidate genes that are related to the oxygen metabolism in red blood cells.

**Results/conclusion:** Using the unconditional logistic regression analysis, false discovery rate (FDR) calculation, and Bonferroni correction we found that PDK1, RUNDC3B, EPO, RELN, MET, PTK2, TDRD1, TCL1A, STAT3, STAT5A, IL12RB1, and NF2 gene polymorphisms were associated with increased risk of HAPC. We determined gene-specific differences and found 12 gene enrichments on the pathway of 16, the main enrichment in the PI3K-AKT pathway, JAK-STAT pathway, and HIF-1 pathway.

**Keywords:** High-altitude polycythemia; highland people; SNPs

## PgmNr 2023: Association of ADAM17 SNPs with TNF $\alpha$ inhibitor response in rheumatoid arthritis patients.

### Authors:

H.J. Kim <sup>1</sup>; H.R. Yeon <sup>1</sup>; S.O. Kang <sup>1</sup>; K.H. Min <sup>1</sup>; H.A. Kim <sup>2</sup>; E. Kim <sup>3</sup>; K.E. Lee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) College of Pharmacy, Chungbuk National University, Cheongju, Republic of Korea; 2) Department of Rheumatology, Ajou University School of Medicine, Suwon, Republic of Korea; 3) College of Pharmacy, Ajou University, Suwon, Republic of Korea

---

**Background:** Patients with rheumatoid arthritis receive treatment modulating TNF and interleukins. However, about 30% of these patients respond poorly and genetics have a potential role in this variability. A disintegrin and metalloproteinase 17(ADAM17), also known as tumor necrosis factor- $\alpha$  converting enzyme(TACE) is a membrane bound enzyme that cleaves immune cell surface cytokine and cytokine receptor, such as TNF $\alpha$  and IL-6, which led to the speculation that inhibition of this metalloprotease might have beneficial effects in autoimmune diseases. Thus, we evaluated the influence of ADAM17 genetic polymorphisms on the treatment response of TNF $\alpha$  inhibitors.

**Method:** Seven SNPs of ADAM17 were selected based on <http://grch37.ensembl.org/index.html> with minor allele frequency of greater than 10% in Japanese and Han Chinese. Genotyping analysis was performed using SNaPshot assay and Taqman genotyping assay after extracting gDNA from blood or buffy coat obtained from patients, who received TNF inhibitor in Ajou University Hospital and Chungbuk National University Hospital in Korea. Clinical data of Patients were collected from electronic medical records and statistic analysis were performed using IBM SPSS Statistics 25.

**Result:** There were six patients (1%) who had medication changes of more than once within the TNF $\alpha$  inhibitors. Among the seven SNPs of our interest, two SNPs revealed statistically significant differences in medication changes (rs117645314 and rs117179141 both  $p=0.018$ ). These two SNPs were found to be in linkage disequilibrium(LD) relation. Multivariate regression analysis was performed to examine the independent effect of variables. Body weight and rs117645314 were significantly associated with medication change after adjusting for age, weight, sex, and rs117645314. ( $r^2=0.367$ )

**Conclusion:** This study showed that ADAM17 can be a possible marker for the treatment response of TNF $\alpha$  inhibitor.



## PgmNr 2024: Association of response to TNF- $\alpha$ inhibitors with rheumatoid arthritis-related SNPs.

### Authors:

N. Trinh <sup>1</sup>; K.H. Min <sup>1</sup>; S.O. Kang <sup>1</sup>; H.J. Kim <sup>1</sup>; H.R. Yeon <sup>1</sup>; I.A. Choi <sup>2</sup>; K.E. Lee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Chungbuk National University, Cheongju, Korea, Republic of Korea; 2) Division of Rheumatology, Department of Internal Medicine, Chungbuk National University Hospital, Cheongju, Korea

---

**Background:** Rheumatoid arthritis (RA) is an autoimmune disease characterized by chronic inflammation that mainly affects the joints. The pathogenesis of RA remains unknown, however, a number of studies suggest that genetic component may be responsible for up to 60% of susceptibility to RA. This suggests that genetics could play an integral part in disease susceptibility and treatment. **Method:** Eleven candidate SNPs were selected by using PubMed text-mining related to RA. We collected clinical data to measure the response to TNF- $\alpha$  antagonist treatment for remission using 28-joint count Disease Activity Score (DAS28) and subcomponents [tender joint count (TJC), swollen joint count (SJC), erythrocyte sedimentation rate (ESR) and patient's general health (GH)]. Our primary endpoint was the difference in DAS28 at baseline and 6 months follow-up ( $\Delta$ DAS28). Secondary outcome was the difference in TJC28, SJC28, ESR, and GH at baseline and 6 months follow-up ( $\Delta$ TJC28,  $\Delta$ SJC28,  $\Delta$ ESR, and  $\Delta$ GH).

**Results:** Eleven SNPs revealed no statistically significant differences in  $\Delta$ DAS28. However in secondary outcome, RETN rs3745367 showed statistically significant difference in both  $\Delta$ TJC28 and  $\Delta$ SJC28 (p value = 0.039 and 0.038, respectively). According to independent t-test results,  $\Delta$ SJC28 in patients with CC or CT were significantly higher when compared to TT-carriers in RETN rs7408174. Patients carrying GG in RETN rs3745367 and CC in CD226 rs763361 showed statistically higher DAS28 difference when compared to patients with other alleles ( $3.77 \pm 0.69$  in GG/CC and  $2.32 \pm 1.74$  in other alleles,  $p = 0.034$ ). GG/CC-carriers when compared with patients carrying other alleles in rs3745367/rs763361 also revealed statistically significant differences in  $\Delta$ TJC28 ( $18.00 \pm 4.24$  versus  $5.94 \pm 7.57$ ,  $p < 0.001$ ) and  $\Delta$ SJC28 ( $14.00 \pm 8.08$  versus  $4.34 \pm 5.44$ ,  $p = 0.019$ ).

**Conclusion:** This study showed that RA patients carrying GG/CC in rs3745367/rs763361 showed favorable response to TNF- $\alpha$  inhibitor treatment, using the endpoint of the difference of DAS28, TJC28 and SJC28 between at baseline and 6-month follow-up. Our results revealed that RETN and CD226 SNPs could be determinants of response in TNF- $\alpha$  inhibitor treatment.

## **PgmNr 2025: Identifying novel high-impact rare disease-causing mutations, genes, and pathways in exomes of Ashkenazi Jewish inflammatory bowel disease patients.**

### **Authors:**

Y. Wu; C. Sevim Raybak; Y. Itan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

---

The Ashkenazi Jewish (AJ) population is predisposed to inflammatory bowel disease (IBD) arising via a founder effect approximately 30 generations ago. Most IBD genomic studies were done in Europeans, following the genome-wide association studies approach of common alleles/complex disease architecture. In this study we aim to detect IBD-related rare high-impact variants, genes and pathways within the AJ population. We genetically identified in whole exome sequencing (WES) data of the IBD Genetics Consortium 4,968 Ashkenazi Jewish samples, of them 1,905 IBD cases and 3,063 controls. We employed functional impact and population genetics filters to collect credible genetic variants in the IBD and controls WES data. Based on filtered high-impact variants, gene burden cases-controls analysis revealed several novel and highly significant IBD-related mutations and genes, as well as well-established Crohn's disease genes, including NOD2. We then performed Phenome-Wide Association Studies (PheWAS) analyses on Mount Sinai's BioMe Biobank and tested classification performance of polygenetic risk scores (PRS) derived from filtered variants. The PheWAS analyses replicated known IBD associations verifying the novel mutations and genes results, and discovered new phenotype associations as well. PRS exhibited promising power in identifying IBD in AJs (AUC=0.8209). Finally, we estimated biological pathways that were significantly enriched in AJ IBD cases' significant genes. Our results provide insights into high-impact rare variants, their harboring genes and associated pathways that contribute to IBD in the AJ population.

## PgmNr 2026: Unmapped exome sequencing reads reveal a potential role for virome variation in childhood HIV-1 disease progression.

### Authors:

S. Mwesigwa <sup>1</sup>; G. Mboowa <sup>1</sup>; L. Williams <sup>2</sup>; G. Retshabile <sup>2</sup>; B. Mlotshwa <sup>2</sup>; E. Katagirya <sup>1</sup>; D. Kateete <sup>1</sup>; E. Wampande <sup>1</sup>; S. Kyobe <sup>1</sup>; M. Wayengera <sup>1</sup>; W.S. Mpoloka <sup>2</sup>; A. Mirembe <sup>3</sup>; I. Kasvosve <sup>2</sup>; K. Morapedi <sup>2</sup>; G.P. Kisitu <sup>3</sup>; A. Kekitiinwa <sup>3</sup>; M.L. Joloba <sup>1</sup>; C.W. Brown <sup>6</sup>; G. Mardon <sup>5</sup>; M. Matshaba <sup>4</sup>; N. Hanchard <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) College of Health Sciences, Makerere University, Kampala Uganda; 2) University of Botswana, Gaborone Botswana; 3) Baylor College of Medicine Children's Foundation Uganda (Baylor Uganda), Kampala Uganda; 4) Botswana Baylor Children's Clinical Centre of Excellence, Gaborone Botswana; 5) Department of Molecular and Human Genetics Baylor College of Medicine, Houston, TX; 6) University of Tennessee Health Science Center, LeBonheur Children's Hospital and St. Jude Children's Research Hospital, Memphis, TN

**SIGNIFICANCE:** Human immunodeficiency virus (HIV) infection remains a major public health burden across sub-Saharan Africa. Children perinatally infected with HIV exhibit extremes in their temporal progression to AIDS - Rapid Progressors (RPs) develop AIDS 3-6 months of HIV infection, whereas Long-Term Non-Progressors (LTNPs) show no signs of AIDS  $\geq 10$  years after infection, even without therapy. The genetic factors underpinning this variability in African children remain unknown, but viral co-infection has been postulated as a potential contributor.

**METHODS:** We used VirusFinder v.2 pipeline to detect viral sequences from unmapped exome sequencing reads from blood samples of 812 HIV+ children from Botswana and Uganda, equally distributed between LTNPs and RPs.

**RESULTS:** Among the  $\sim 1\%$  of reads ( $N=2.9 \times 10^8$ ) that didn't map to the human genome, we found evidence for 19 viral species (median identity  $> 90\%$ ). Consistent with population ascertainment and sampling conditions, phiX174, human Herpes, and HIV were most commonly observed but didn't differ significantly by group. Ugandan samples showed a higher diversity index for viral communities than Botswana ( $p=4.6 \times 10^{-13}$ ), characterized by the relatively common occurrence of Togaviridae, Herpesviridae, and Anelloviridae. LTNPs were more likely to have detectable viral sequences than RPs (78% vs 65%;  $p=0.00004$ ; OR, 1.9; 95%CI (1.4 - 2.6)), specifically, a significant excess of the Anelloviridae family ( $p=0.0003$ ;  $q=0.004$ ; OR, 3.98; 95%CI (1.7-10.2)), and predominantly the Torque Teno virus (TTV) species ( $p=0.0003$ ; OR, 3.1 95%CI (1.7-5.9)), a viral class that has recently been implicated in immune suppression during HIV co-infection. This was independent of covariates including gender, contig counts, and other viral families present. This trend was still evident when stratified by country although it was no longer statistically significant in Uganda (Botswana;  $p=0.00004$ ; Uganda;  $p=0.06868$ ).

**CONCLUSION:** We demonstrate the potential of next-generation sequencing to inform complex disease studies in African populations. Our results imply that the blood-virome could act as a proxy biomarker for disease progression, and that temporal variation in exposure and infection with TTV could play an important role in determining HIV disease progression. Studies of host variants influencing HIV disease progression in African children will need to take account for concomitant viral

co-infection.

## PgmNr 2027: An Epstein-Barr virus (EBV) transcription co-factor, EBNA2, concentrates at variants in the risk loci of Kawasaki's disease.

### Authors:

B. Emadi <sup>1</sup>; V. Laurynenka <sup>1</sup>; S.H. Eswar <sup>1</sup>; S. Parameswaran <sup>1</sup>; X. Chen <sup>1</sup>; M. Hostetter <sup>2</sup>; L. Kottyan <sup>1,5</sup>; M. Weirauch <sup>1,4,6</sup>; J. Harley <sup>1,3,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Autoimmune Genomics & Etiology (CAGE), CCHMC, Cincinnati, OH.; 2) Department of Pediatrics, University of Cincinnati, Cincinnati, OH.; 3) Divisions of Immunology Biology, CCHMC, Cincinnati, OH.; 4) Developmental Biology, CCHMC, Cincinnati, OH.; 5) Allergy & Immunology, CCHMC, Cincinnati, OH.; 6) Biomedical Informatics, CCHMC, Cincinnati, OH.; 7) US Department of Veterans Affairs Medical Center, Cincinnati, OH, USA.

---

Kawasaki disease (KD), also known as mucocutaneous lymph node syndrome, is an idiopathic, acute, multisystem, inflammatory, and self-limited vasculitis, usually affecting infants and children younger than 5 years of age. Occasionally, the coronary artery vasculitis of KD is lethal. Many viruses and bacteria have been suggested as potential causes of the disease. Epstein-Barr virus (EBV) has been discarded as an etiology by multiple authors because of lowered frequencies of serologic responses in KD at presentation.

From published genome wide association studies (GWASs) and candidate gene studies in KD patients we extracted 11 single nucleotide polymorphisms (SNPs) in Asian ancestry studies, all at  $p < 5 \times 10^{-8}$ , which reduce to 5 independent loci by pruning the set markers in linkage disequilibrium with each other. We evaluated 11,482 TF ChIP-seq (chromatin immunoprecipitation with DNA sequencing) datasets, which include 53 viral TFs. To determine whether the intersections of transcription factor and co-factor (TF) DNA binding sites concentrate at variants contained in the KD risk loci, we applied the RELI (Regulatory Element Locus Intersection) (Nat Genet 50:699, 2018).

Significant binding was found in 14 TF ChIP-seq datasets for 9 human TFs (BACH2, RELA, H3F3A, EP300, POLR2A, MED1, SPI1, POU2F2 and NFXL1) and one viral TF, all with  $P < 1 \times 10^{-6}$  after Bonferroni correction. Unexpectedly, the binding of EBV Epstein-Barr virus Nuclear Antigen 2 (EBNA2) to DNA in the 5 KD risk loci is >27-fold enriched compared to the whole genome with  $P = 9.6 \times 10^{-14}$ . Among 5 loci only 4 are occupied by human and viral TFs consistent with their having a regulatory role. Twelve (85.7%) of 14 ChIP-seq datasets were obtained from B cell lines, 9 (64.3%) of them are from EBV-transformed B cell lines in the Latency III program of viral expression in which EBNA2 is an expressed viral gene products. These results show that features characteristics of the EBV transformed B cell expressing EBNA2 correlate with the genetic mechanism(s) of pathogenesis in at least some children with Kawasaki disease. These results are consistent with, but do not establish, causation.

## PgmNr 2028: A GWAS of neonatal sepsis shows significant genetic differences between males and females.

### Authors:

S. Williams <sup>1</sup>; T. Ciesielski <sup>1</sup>; X. Zhang <sup>1</sup>; J. Bartlett <sup>1</sup>; G. Sirugo <sup>2</sup>; NeoMero Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Population and Quantitative Health Sciences, Case Western Reserve Univ, Cleveland, Ohio.; 2) Division of Translational Medicine and Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Late Onset Neonatal Sepsis is a serious perinatal condition, involving widespread infection, immune disruption, organ dysfunction, and in 15-30% of cases, death. The initiation of this disorder requires contact with infectious agents in the perinatal period and this presents opportunities to prevent exposure, but these behaviors cannot be 100% effective. Thus, it is critical to identify host susceptibility factors, gain etiologic insights, and develop interventions that predict and help the neonates at greatest risk. Here we performed a case control genome wide association study using the Illumina MEGA Consortium V2 Beadchip to identify genetic factors that associate with neonatal sepsis. Participants were identified in hospitals from several European countries (NeoMero Consortium: Estonia, Greece, Italy, Lithuania, the Netherlands, Spain, the UK, and France), and after initial quality control processing there were 224 cases and 273 controls. Because neonatal sepsis is more common in males and some etiologic factors may differ between the sexes, we performed combined and sex-stratified association analyses. Additionally, because the X chromosome encodes a number of immune related genes, we included it in the GWAS and performed sex-stratified quality control and association analyses in the X chromosome data. We identified multiple SNPs, 71 in total, that associated with neonatal sepsis at  $p < 1 \times 10^{-4}$  in at least one of the five adjusted logistic regression analyses, but did not detect any genome wide significant associations ( $p < 5 \times 10^{-8}$ ). However, we observed large differences in the findings between males and females. None of the top SNPs in males were associated with sepsis in females (at  $p < 0.05$ ), and none of the top SNPs in females were associated among the males (at  $p < 0.05$ ). These statements hold for both the autosomal and X chromosome variants. When the top SNPs from these sex-stratified analyses were analyzed in the full study population, most of the sex\*SNP interaction terms were significant and they were highly significant for all female top SNPs. Finally, we note that the association between *IL-10* variation and neonatal sepsis previously reported, was also observed with 5 SNPs in our combined analysis (at  $p < 0.05$ ). These results indicate that genetic susceptibility factors for neonatal sepsis likely differ by sex and future etiologic research should routinely involve sex-stratified sensitivity analyses.

## PgmNr 2029: Causal role of L-glutamine and 3-ureidopropionate in sickle cell disease complications: A Mendelian randomization analysis.

### Authors:

Y. Ilboudo <sup>1</sup>; M. Garrett <sup>2</sup>; P. Bartolucci <sup>4</sup>; F. Galactéros <sup>4</sup>; A. Ashley-Koch <sup>2</sup>; M. Telen <sup>3</sup>; G. Lettre <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biochemistry, University of Montréal, Montreal, Quebec, Canada; 2) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, North Carolina, United States of America; 3) Department of Medicine, Division of Hematology, Duke University Medical Center, Durham, North Carolina, United States of America; 4) Red Cell Genetic Disease Unit, Hôpital Henri-Mondor, Assistance Publique-Hôpitaux de Paris (AP-HP), Université Paris Est, IMRB - U955 - Equipe no 2, Créteil, France

In 2017, the FDA approved for the first time in 20 years a new drug to treat sickle cell disease (SCD). The drug, Endari, is an oral L-glutamine metabolite that acts by reducing oxidative stress in erythrocytes. The conclusions from the phase 3 randomized clinical trial suggested that treatment with L-glutamine reduces pain crises and hospitalization rate (Niihara *et al.*, NEJM, 2018). Our goals were to confirm the clinical trial findings using an alternative, genetics-based strategy, and to identify other metabolites causally implicated in the clinical heterogeneity observed in SCD patients.

Employing a targeted metabolomics platform, we profiled 129 metabolites from the plasma of two SCD cohorts from Paris, France (GEN-MOD (n=406)) and North Carolina, USA (OMG-SCD (n=300)). To prioritize metabolites, we performed correlation analyses between metabolite levels and 6 SCD complications (pain crises, cholecystectomy, retinopathy, leg ulcer, priapism, aseptic necrosis) or estimated glomerular filtration rate (eGFR). We assessed statistical significance via permutation test and used 2-sample Mendelian randomization (MR) methods to test for causality. We obtained MR instruments associated with metabolites of interest from two large metabolite GWAS performed in healthy Europeans (Shin *et al.*, Nature Genetics, 2014, n=7,824; Long *et al.*, Nature Genetics, 2017, n=1,960). We derived SNP associations with SCD clinical outcomes from the large African-American CSSCD cohort (n=1,278).

We identified a causal relationship between L-glutamine levels and pain crises requiring hospitalization (odds ratio (OR) [95% confidence interval]=0.68 [0.52 - 0.89], P=0.005). Our MR study estimates a 32% reduction in pain crises with one standard deviation increase of L-glutamine. This estimate is comparable to the 33% reduction in hospitalization due to emergency room visit observed in SCD patients treated with L-glutamine vs. placebo in the clinical trial. We identified 71 other significant associations (FDR<5%) between metabolite levels and SCD-related complications or eGFR. Two-sample MR analyses revealed a causal association between 3-ureidopropionate and eGFR: a standard deviation increase in 3-ureidopropionate leads to an eGFR increase of 0.07 mL/min per 1.172 m<sup>2</sup> (P=9.7x10<sup>-4</sup>).

MR analyses provide additional evidence to support the causal role of L-glutamine in modulating pain crises in SCD. Additionally, we found a tentative causal link between 3-ureidopropionate levels and

eGFR.



## PgmNr 2030: A metagenome-wide association study of gut microbiome revealed novel etiology of rheumatoid arthritis in the Japanese population.

### Authors:

T. Kishikawa<sup>1,2</sup>; Y. Maeda<sup>3,4</sup>; T. Nii<sup>3,4</sup>; D. Motooka<sup>5</sup>; Y. Matsumoto<sup>5</sup>; M. Matsushita<sup>6,7</sup>; H. Matsuoka<sup>7</sup>; M. Yoshimura<sup>7</sup>; S. Kawada<sup>8</sup>; S. Teshigawara<sup>7</sup>; E. Oguro<sup>3</sup>; Y. Okita<sup>7</sup>; K. Kawamoto<sup>8</sup>; S. Higa<sup>8</sup>; T. Hirano<sup>3</sup>; M. Narazaki<sup>3</sup>; A. Ogata<sup>8</sup>; Y. Saeki<sup>7,9</sup>; S. Nakamura<sup>5</sup>; H. Inohara<sup>2</sup>; A. Kumanogoh<sup>3,10</sup>; K. Takeda<sup>4,11</sup>; Y. Okada<sup>1,12</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan.; 2) Department of Otorhinolaryngology - Head and Neck Surgery, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan.; 3) Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of Medicine, Suita 565-0871, Japan.; 4) Laboratory of Immune Regulation, Department of Microbiology and Immunology, Osaka University Graduate School of Medicine, Suita 565-0871, Japan.; 5) Department of Infection Metagenomics, Research Institute for Microbial Diseases, Osaka University, Suita 565-0871, Japan.; 6) Department of Rheumatology and Allergology, Saiseikai Senri Hospital, Suita 565-0862, Japan.; 7) Rheumatology and Allergology, NHO Osaka Minami Medical Center, Kidohigasi-machi, Kawachinagano, Osaka, 586-8521, Japan.; 8) Division of Rheumatology, Department of Internal Medicine, Daini Osaka Police Hospital, 2-6-40 Karasugatsuji Tennoji-ku, Osaka 543-8922 Japan.; 9) Clinical Research, NHO Osaka Minami Medical Center, Kidohigasi-machi, Kawachinagano, Osaka, 586-8521, Japan.; 10) Department of Immunopathology, Immunology Frontier Research Center, Osaka University, Suita 565-0871, Japan.; 11) WPI Immunology Frontier Research Center, Osaka University, Suita 565-0871, Japan.; 12) Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, Japan.

**Objective** The causality and pathogenic mechanism of microbiome composition remain elusive in many diseases, including autoimmune diseases such as rheumatoid arthritis (RA). This study aimed to elucidate gut microbiome's role in RA pathology by a comprehensive metagenome-wide association study (MWAS).

**Methods** We conducted MWAS of the RA gut microbiome in the Japanese population ( $n_{\text{case}} = 82$ ,  $n_{\text{control}} = 42$ ) by utilizing whole-genome shotgun sequencing of high depth (average 13 Gb per sample). Our MWAS consisted of three major bioinformatic analytic pipelines (phylogenetic analysis, functional gene analysis, and pathway analysis).

**Results** Phylogenetic case-control association tests showed high abundance of multiple species belonging to the genus *Prevotella* (but other than previously reported *Prevotella copri*) in the RA case metagenome. The non-linear machine learning method efficiently deconvoluted the case-control phylogenetic discrepancy. Gene functional assessments showed that the abundance of one redox reaction-related gene was significantly decreased in the RA metagenome compared to controls. A variety of biological pathways including those related to metabolism (e.g., fatty acid biosynthesis and glycosaminoglycan degradation) were enriched in the case-control comparison. A population-specific link between the metagenome and host genome was identified by comparing biological pathway

enrichment between the RA metagenome and the RA genome-wide association study (GWAS) results. No apparent discrepancy in alpha- or beta-diversities of metagenome was found between RA cases and controls.

**Conclusion** Our shotgun sequencing-based MWAS highlights a novel link among the gut microbiome, host genome, and pathology of RA, which contributes to our understanding of the microbiome's role in RA etiology.

# PgmNr 2031: The effects of common *TMPRSS6* and *TF* variants on iron status in healthy Gambian women of reproductive age.

## Authors:

M. Jallow <sup>1</sup>; S. Campino <sup>2</sup>; A. Prentice <sup>1</sup>; C. Cerami <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) The Nutrition Theme, Medical Research Council Unit The Gambia at London School of Hygiene & Tropical Medicine, Atlantic Road Fajara, P.O. Box 273 Banjul, The Gambia; 2) Department of Infection Biology, Faculty of Infectious and Tropical Diseases, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT, UK

## Background

Anaemia is a global health problem affecting women of reproductive age and children in low-and middle-income countries the most. Nutritional deficiencies, infection and genetic risk factors are the major drivers. However, the role of genetic factors particularly in settings where the prevalence of anaemia is high, has not been fully investigated. Genome-wide association studies have identified numerous single nucleotide polymorphisms in the *TMPRSS6* and *TF* genes which are linked to impaired iron status. However, the effects of these SNPs on anaemia in West African populations have not been investigated.

## Objectives

To investigate the effects of SNPs in *TMPRSS6* and *TF* genes on iron status in healthy Gambian women of reproductive age.

## Methods

We analyse data from women of reproductive age (15 to 49years), n=227, with genotype data on *TMPRSS6* (rs2235321, rs855791, rs4820268, rs2235324, rs2413450 and rs5756506) and *TF* (rs3811647 and rs1799852), and iron biomarker data (serum iron, unsaturated iron binding capacity (UIBC), transferrin, ferritin soluble transferrin receptor (sTfR), transferrin saturation (TSAT) and total iron binding capacity (TIBC) and hepcidin) collected prospectively as part of the MRC Keneba Biobank, in The Gambia. We investigated the effects of genotype on the iron status.

## Results

The minor allele (G) of *TMPRSS6* rs4820268 showed elevated serum iron [17.02 (5.25) vs 12.63 (4.82) *p*-value=0.039], TSAT [27.9 (10.3) vs 19.1 (6.7); *p*-value=0.022], and ferritin [33.6 (16.0) vs 32.7 (36.1); *p*-value=0.025], and decreased TIBC [62.2 (6.9) vs 67.8 (9.5); *p*-value=0.046], GG vs AA genotypes respectively. Inversely, the minor allele (A) of *TF* rs3811647 showed reduced TSAT [22.3 (11.2) vs 26.7 (12.9), *p*-value=0.024], elevated transferrin [3.1 (0.8), *p*-value=0.019], TIBC [68.2 (10.7) vs 62.2 (11.2), *p*-value=0.001] and UIBC [53.5 (13.5) vs 46.2 (13.9), *p*-value=0.001], GA vs GG genotypes respectively.

**Conclusion**

*TMPRSS6* rs4820268 minor allele (G) may be protective against low iron status whereas, the minor allele (A) of *TF* rs3811647 may predispose to low iron status. We aim to further analyse a larger dataset and with more SNPs both individually and in combination with other ones, and different in populations groups to further gain more insight into the effects of SNPs within the iron regulatory genes on the risk of anaemia. This may enable the development of a genetic risk score for anaemia.

## PgmNr 2032: Deep phenotyping of individuals with 3q29 deletion syndrome: Results from The Emory 3q29 Project.

### Authors:

J.G. Mulle <sup>1,7</sup>; M.M. Murphy <sup>1</sup>; T.L. Burrell <sup>2,3</sup>; J.C. Cubells <sup>1,4,5</sup>; R.A. Espana <sup>6</sup>; M.J. Gambello <sup>1</sup>; K.C.B. Goines <sup>5,6</sup>; C. Klaiman <sup>2,3</sup>; S. Koh <sup>2</sup>; R.L. Sanchez Russo <sup>1</sup>; C.A. Saulnier <sup>2</sup>; E.F. Walker <sup>6</sup>; The Emory 3q29 Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University School of Medicine, Whitehead Research Building, 615 Michael Street, Suite 300, Atlanta, GA, 30322, USA.; 2) Department of Pediatrics, Emory University School of Medicine, 1920 Briarcliff Road NE, Atlanta, GA, 30322, USA; 3) Marcus Autism Center, 1920 Briarcliff Road NE, Atlanta, GA, 30322, USA.; 4) Emory Autism Center, 1551 Shoup Court, Atlanta, GA, 30322, USA.; 5) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, 12 Executive Park Drive, 2nd floor, Atlanta, GA, 30329, USA.; 6) Department of Psychology, Emory University, 36 Eagle Row, Atlanta, GA, 30322, USA.; 7) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, 30322,

---

3q29 deletion syndrome is a rare (1 in 30,000) genomic disorder associated with cognitive disability, autism, generalized anxiety disorder, and a 40-fold increased risk for schizophrenia (SZ). The Emory 3q29 Project (<http://genome.emory.edu/3q29/>) unites an interdisciplinary team toward the common goal of understanding the phenotypic spectrum, natural history, and molecular mechanism of the 3q29 deletion. To facilitate investigation of the 3q29 deletion phenotype, our team has developed a comprehensive, transdiagnostic phenotyping protocol for both diagnostic and dimensional assessments (Murphy MM et al, 2018). We have now directly evaluated 20 individuals with the 3q29 deletion (average age = 12 years, range 6-26) for cognitive ability, anxiety, attentional deficits/executive function, fine motor function/cerebellar dysregulation, social disability, and prodrome/psychosis. Using strict diagnostic criteria, our study subjects qualify for diagnoses as follows: 56% intellectual disability, 30% autism spectrum disorder (ASD), 30% generalized anxiety disorder, and 10% psychosis or SZ. This enrichment of psychosis is noteworthy given that the majority of our study subjects have not yet approached the age at risk for SZ. 63% of our study subjects qualify for a diagnosis of Attention Deficit/Hyperactivity Disorder (ADHD), which was not previously appreciated as part of the 3q29 deletion phenotype. When we consider quantitative measures, we find that cognitive ability is mildly impacted (average IQ = 73, range 99-46) but more than 60% of our study subjects have subtest scores that differ by more than 1 standard deviation. Verbal ability is typically a strength (average = 80), and spatial reasoning is a weakness (average = 66). 75% of study subjects with 3q29 deletion exhibit significant graphomotor weakness, and this is driven by motor coordination deficits. Males have spatial reasoning scores that are on average 12 points lower than females, and motor coordination deficits are also more pronounced in males, suggesting sex-dependent effects in this domain. Taken together, these data a) illuminate the phenotype of 3q29 deletion syndrome; b) provide specific direction for clinical and educational support; c) will be crucial for connecting phenotypic deficits to molecular mechanism; and d) inspire future directions to understand whether second hit mutations, polygenic background, or environmental exposures contribute to the heterogeneity of the disorder.



## PgmNr 2033: *In vivo* modeling of the 2q13 reciprocal CNV reveals multiple phenotypic contributors and genetic interactions.

### Authors:

G. Kellaris <sup>1,2</sup>; C. Golzio <sup>3,4</sup>; N. Katsanis <sup>2</sup>; E. Davis <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC 27701, USA; 2) Pediatrics department, Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University, Chicago, IL 60611, USA; 3) Institut de Génétique et de Biologie Moléculaire et Cellulaire, Department of Translational Medicine and Neurogenetics; Centre National de la Recherche Scientifique, UMR7104; 4) Institut National de la Santé et de la Recherche Médicale, U964, Université de Strasbourg, 67400, Illkirch, France

---

Copy number variants (CNV), a subclass of structural variation, alter gene dosage through either haploinsufficient or triplosensitive pathomechanisms and are an established source of mutational burden in neurodevelopmental traits. Here, we investigate an experimentally tractable reciprocal CNV at 2q13. Over 50 deletion and 20 duplication carriers have been reported, and affected individuals display a constellation of variable phenotypes including head size defects, psychiatric phenotypes, seizures, facial dysmorphisms, and cardiac anomalies. The minimum overlapping region among reported cases comprises eight genes (*FBLN7*, *BUB1*, *TMEM87B*, *MERTK*, *Z3CH8*, *ACOXL*, *ANAPC1*, *BCL2L11*), and for a majority of loci, evidence is lacking from single nucleotide variants in healthy or diseased individuals to support a role in phenotype(s). Dosage gain models, generated by overexpression of each human transcript in zebrafish embryos, did not identify any mRNAs capable of inducing a detectable head size or craniofacial phenotype in isolation. Pair-wise overexpression of all possible gene combinations showed that three discrete pairs resulted in significant craniofacial defects. However, we observed thirteen pairs that resulted in microcephaly, with the majority of *ACOXL* pairs leading to head size reduction. Dosage loss zebrafish models for *tmem87a*, *anapc1* and *bcl2l11* individually displayed relevant phenotypes, such as microcephaly and craniofacial abnormalities. Further, pair-wise co-suppression of genes that demonstrated no detectable phenotype in sub-effective doses revealed epistatic events, similar to the overexpression studies and consistent with the mirror phenotypes of humans. Finally, placement of *ANAPC1* or *TMEM87A* on a protein affinity network comprised of phenotypically-related genes implicated in autism, intellectual disability, and head size phenotypes showed putative involvement in discrete functional modules governing cell cycle progression and endoplasmic reticulum targeting, respectively. Our data suggest that phenotype-specific epistatic interactions both within the 2q13 locus and also in *trans* with functionally relevant genes outside the 2q13 region likely drive neuroanatomical and/or craniofacial abnormalities associated with this CNV.

## PgmNr 2034: The emerging role of *RBFOX1* in visual impairment genetics.

**Authors:**

F. Bolduc; S. Weber

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dept Pediatrics, Univ Alberta, Edmonton, Alberta, Canada

---

**BACKGROUND.** Several individuals with neurodevelopmental disability (NDD) also present with clinically relevant visual impairment (VI). While VI can sometimes be difficult to recognize early on, its impact on daily functioning is significant for the individual, the family and educators. VI is not systematically reported in individuals with NDD. Here, we identified 2 individuals from the same family with an *RBFOX1* intragenic heterozygous deletion with visual processing defects.

**METHODS:** We used comparative genomic hybridization to identify a copy number variation (CNV) in individuals with NDD seen in our neurogenetic clinic. We also reviewed the published literature to develop a comprehensive clinical description of *RBFOX1*. Next, we conducted a pathway analysis to identify possible mediators of visual impairment in individuals with an *RBFOX1* CNV.

**RESULTS:** The role of *RBFOX1* role in intellectual disability, anxiety, sleep difficulty and focal epilepsy as well as renal and cardiac malformations has been previously reported. We identified a novel potential role with cortical visual impairment in 2 individuals with an intragenic *RBFOX1* deletion with overlapping clinical characteristics. Our pathway analysis revealed several genetic links between *RBFOX1* and VI.

**DISCUSSION:** Our findings highlight the need to further document the presence of visual impairment in analysis of NDD phenotypes as it has a potentially relevant clinical phenotype. Our results also support the recent findings in a mice model of *RBOX1* where visual impairment was reported.



## PgmNr 2035: Polygenic score profiling of quantitative lipid traits: Lessons from a specialized dyslipidemic cohort.

### Authors:

J.S. Dron<sup>1,2</sup>; J. Wang<sup>1</sup>; M.A. Iacocca<sup>3</sup>; A.D. McIntyre<sup>1</sup>; M.R. Ban<sup>1</sup>; H. Cao<sup>1</sup>; J.F. Robinson<sup>1</sup>; R.A. Hegele<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada; 2) Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada; 3) Department of Biomedical Data Science, Stanford School of Medicine, Stanford University, Stanford, CA, United States of America; 4) Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada

**Introduction:** Polygenic scores (PSs) are often calculated in prospective cohorts representing generally healthy populations. Because of this, characterizing polygenic underpinnings of a quantitative trait's full phenotypic spectrum can be challenging given the rarity of phenotypic extremes, which are of interest in the context of human health and disease. In dyslipidemia patients from a tertiary referral clinic, we aimed to establish polygenic profiles for the three main blood lipid traits—low-density lipoprotein cholesterol (LDL-C), triglyceride (TG), and high-density lipoprotein cholesterol (HDL-C)—with a focus on the extremes of each trait.

**Materials and Methods:** We performed PS analyses on patients using a 10-SNP LDL-C score (N=1,226), 16-SNP TG score (N=1,406), and 9-SNP HDL-C score (N=1,298). For clinical utility, we utilized smaller PSs that incorporate SNPs with the largest phenotypic effects. "Extreme scores" reflect an excess of trait-raising SNPs and are defined as scores  $\geq 90^{\text{th}}$  percentile, as calculated in the 1000 Genomes cohort. For each PS analysis, patients were stratified by clinically relevant thresholds for the associated trait (i.e. for the LDL-C score, patients were stratified by LDL-C level). Odds ratios were calculated between stratification groups of a trait, and differences in mean PS for each stratification group were assessed with a Kruskal-Wallis test using Dunn's multiple test adjustment.

**Results:** As LDL-C levels increased, the prevalence of extreme LDL-C PSs increased, except for patients with the highest LDL-C levels, as their PSs were similar to those with normal levels. As TG levels increased, the prevalence of extreme TG PSs increased; patients with the highest TG levels were 5.30-fold [3.47-8.06;  $P < 0.0001$ ] more likely to have an extreme PS compared to those with normal levels. As HDL-C levels increased, the prevalence of extreme HDL-C PSs increased; patients with the highest HDL-C levels were 2.37-fold [1.32-4.33;  $P = 0.003$ ] more likely to have an extreme PS compared to those with normal levels.

**Conclusions:** Polygenic profiles of each lipid trait were unique. High LDL-C had a less pronounced polygenic basis, suggesting a more prominent monogenic component. Conversely, high TG and HDL-C had a stronger polygenic basis. Evaluation of patients at phenotypic extremes allows for more efficient use of resources to define clinically relevant genetic determinants for diagnosis and possibly tailored interventions.

## PgmNr 2036: Unexplained familial hypercholesterolemia in African Americans: No detectable mutation does not necessarily imply a polygenic basis.

### Authors:

O. Soylemez<sup>1,2,3</sup>; Z. Ahmad<sup>4</sup>; E. Koch<sup>1,2,3</sup>; S. Sunyaev<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetics, Department of Medicine, Brigham & Women's Hosp, Boston, Massachusetts.; 2) Harvard Medical School, Boston, Massachusetts.; 3) Broad Institute, Cambridge, Massachusetts.; 4) Division of Endocrinology, Department of Internal Medicine, UT Southwestern Medical Center.

---

Extreme values of quantitative traits or rare severe phenotypic presentations can be caused by individual mutations of very large effects segregating in Mendelian fashion. Alternatively, such phenotypes may reflect extremes of polygenic or oligogenic genetic inheritance. Beyond importance for basic genetics, this distinction is practically important for the design of genetic studies that can follow either Mendelian or complex trait paradigm. Currently, sequencing studies of rare phenotypes attribute only a fraction of cases to an obvious monogenic cause. Frequently, individuals with extreme phenotype have family members with milder presentation. Thus, it is expected that family history is informative about genetic architecture of severe phenotypes. We investigated this hypothesis in a series of population genetics computer simulations and the analysis of cases of severe familial hypercholesterolemia (FH) in African American individuals.

Recent work on FH in patients without known causative mutations in *LDLR*, *APOB* and *PCSK9* suggests that a large fraction of such cases may be attributed to polygenic inheritance whereby a large number of small-effect mutations collectively result in extreme LDL-C levels that are comparable to levels in patients with monogenic FH. It remains unclear to what extent the family history is informative to distinguish monogenic and polygenic FH patients with extreme LDL-C levels. We developed a population and statistical genetics framework to simulate families of individuals with extreme LDL-C levels and calculated the fraction of monogenic FH cases conditional on population genetics parameters, varying prevalence, and pedigree structure.

We applied our simulation framework to real pedigrees from 124 African American families with individuals that are clinically ascertained to have autosomal dominant FH with no causative mutations in known FH genes. Families with and without a causative mutation were indistinguishable under a wide range of models. Moreover, family history did not provide substantial information to distinguish likely monogenic or non-monogenic basis for unexplained FH in these families. These findings have significant implications for study design of rare and unexplained genetic conditions as well as genetic testing strategies for new gene discovery, prevention and treatment of FH in underrepresented African American individuals.

## PgmNr 2037: Genetics of hypertriglyceridemic acute pancreatitis: Analysis of a prospective acute pancreatitis cohort and a complex family.

### Authors:

P.J. Greer<sup>1</sup>; A. Wilson<sup>1</sup>; J. Castaneda<sup>1</sup>; E. Orlova<sup>3</sup>; C.A. Shelton<sup>1</sup>; S. O'Keefe<sup>1</sup>; E.E. Kershaw<sup>2</sup>; G.I. Papachristou<sup>1</sup>; C. Gonzaga-Jauregui<sup>4</sup>; D.C. Whitcomb<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, University of Pittsburgh, Pittsburgh, PA.; 2) Division of Endocrinology, Department of Medicine, University of Pittsburgh School of Medicine, University of Pittsburgh, Pittsburgh, PA.; 3) Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania.; 4) Regeneron Genetics Center, Tarrytown, NY

**Background.** Acute pancreatitis (AP) is a sudden, life-threatening inflammatory disorder. Hypertriglyceridemia (HTG) associated AP (HTG-AP) is a common subtype that is associated with higher rates of morbidity and mortality. The risk of HTG-AP is proportional to plasma triglyceride levels driven by genetic and environmental factors. While familial HTG and chylomicronemia syndromes are well defined, the underlying genetics of well-defined AP cohorts are unknown.

**Methods.** The Pancreatitis-associated Risk Of Organ Failure (PROOF) trial includes consecutive subjects with AP admitted to UPMC Presbyterian Hospital between 2006 and 2016. Among the first 500 patients, 68 (13.6%) had TG>200 mg/dL identified within 72 hours of admission. These included TG≥200 and <500 mg/dL (n=26); TG>500 and <1000 mg/dL (n=12); TG≥1000 and <2000 mg/dL (n=6); and TG ≥2000 mg/dL (n=25). HTG-AP patients were matched ~2:1 with AP without HTG (AP). The final analysis cohort included 92 HTG-AP cases, 188 (AP) and a HTG-AP proband with 8 HTG relatives. Whole exome sequencing and whole-genome genotyping using the GSA microarray (Illumina, San Diego, CA, USA) were conducted in collaboration with the Regeneron Genetics Center.

**Results:** A GWAS in our cohort replicated association between HTG-AP and a high-risk *FADS1* locus haplotype ( $p<1e-5$ ). The homozygous risk haplotype was in 3.9% of AP and 13.5% of HTG-AP with TG>1000 mg/dL: allele dose correlated with TG category ( $P<0.05$ ). Pathogenic variants in *LPL* were seen in 7/92 HTG-AP cases and 0/188 AP controls ( $p<0.0005$ ). In HTG-AP cases: *APOA5* (n=8, one complex with an *LPL* variant) *LMF1* (n=7 VUS); *APOC2* (n=2); and *GPIHBP1* variants (n=2, one complex with an *LPL* variant). A HTG-AP proband with a family history of HTG had a complex genotype with two *SPINK1* pancreatitis variants, and HTG risk variants in *APOA*, *ADRB2*, *GCKR*, *FABP1*, and *FABP2* shared in various combinations in affected family members. Pathogenic variants in AP patients with TG>200 and <500 mg/dL, *APOA5* (n=3/26) & *GPIHBP1* (n=1/26); TG≥500 to <1000 mg/dL, *APOA5* (n=2); TG>1000 and <2000 mg/dL *LPL* (1/6); and TG >2000 mg/dL variants in *LPL* (n=6/25) & *APOA5* (n=1/25). No patients with BMI>40 had a pathogenic coding variant (n=9).

**Conclusions:** Patients with HTG-AP and BMI under 40 harbor multiple genetic risks, with pathogenic *LDL* having the strongest effect on TG levels >1000 mg/dL. The *FADS1* risk haplotype is associated with

HTG-AP. Effective therapies may require a precision medicine approach.

## PgmNr 2038: Genetic structure underlying lipids metabolites linked to human diseases.

### Authors:

H. Aschard <sup>1</sup>; Y. Chunzi <sup>1</sup>; H. Julienne <sup>1</sup>; A. Ko <sup>2</sup>; M. Laakso <sup>3</sup>; P. Pajukanta <sup>2</sup>; N. Zaitlen <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computational Biology - USR 3756 CNRS, Institut Pasteur, Paris, France, Institut Pasteur, Paris, France; 2) Department of Human Genetics, University of California, Los Angeles, CA, USA; 3) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 4) Department of Medicine, University of California, San Francisco, CA, USA

---

Systematic genome-wide association studies of 'omics phenotypes highlighted shared genetic components between many molecular traits and human diseases. However, our understanding of the genetic structure underlying molecular traits and their role in common human diseases remains very limited. Here we used 10,000 participants from the Finnish Metabolic Syndrome In Men (METSIM) cohort, a unique resource including measurements for over 150 correlated lipid metabolites along genome-wide genetic data, to assess potential links between complex human phenotypes and genetic structure underlying these lipids metabolites. We first applied top-of-the-art univariate and multivariate approaches in METSIM to select a subset of approximately 500 independent SNPs associated with at least one metabolite. We next performed a clustering of these SNPs with the objective of identifying directional signatures across the 150 metabolite dimensions, while circumventing the issue of choosing a coded allele (i.e. to avoid symmetric clusters due to random allele coding) and accounting for correlation between association statistics due to phenotypic correlation. We achieved this goal by applying a *K-medoids* clustering based on absolute cosine dissimilarity, where each single SNP was weighted by its Euclidean norm to account for association signal strengths. After validating our approach through simulations, we assessed the biological relevance of the identified clusters of SNPs by estimating their enrichment for association with dozens of human diseases and phenotypes using independent GWAS summary statistics. We found several clusters to be highly enriched for association with very specific outcomes, including in particular coronary heart disease, schizophrenia, celiac disease, and years of education. These strong features suggest new hypothetical genetic mechanisms on those outcomes, and highlight the importance of integrative approaches to decipher complex structure underlying multidimensional phenotypes.

## PgmNr 2039: A large effect, Polynesian-specific, stop-gained variant in *BTNL9* is associated with atherogenic lipid profiles.

### Authors:

J.C. Carlson<sup>1</sup>; E.M. Russell<sup>2</sup>; J.Z. Zhang<sup>1</sup>; N.L. Hawley<sup>3</sup>; J. Moors<sup>4</sup>; H. Cheng<sup>5</sup>; N. Dalbeth<sup>6</sup>; J. de Zoysa<sup>6</sup>; J. Harré Hindmarsh<sup>7</sup>; R. Murphy<sup>6,8</sup>; T. Naseri<sup>9</sup>; M.S. Reupena<sup>10</sup>; L. Stamp<sup>11</sup>; J. Tuitele<sup>12</sup>; R. Deka<sup>5</sup>; S.T. McGarvey<sup>13,14</sup>; T.R. Merriman<sup>4,8</sup>; D.E. Weeks<sup>1,2</sup>; R.L. Minster<sup>2</sup>; TOPMed Lipids Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America; 3) Department of Epidemiology (Chronic Disease), School of Public Health, Yale University, New Haven, Connecticut, United States of America; 4) Department of Biochemistry, University of Otago, Dunedin, New Zealand; 5) Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, Ohio, United States of America; 6) Department of Medicine, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand; 7) Ngāti Porou Hauora Charitable Trust, Te Puia Springs, Tairāwhiti East Coast, New Zealand; 8) Maurice Wilkins Centre for Molecular Biodiscovery, University of Auckland, Auckland, New Zealand; 9) Ministry of Health, Government of Samoa, Apia, Samoa; 10) Lusia i Puava ae Mapu i Fagalele, Apia, Samoa; 11) Department of Medicine, University of Otago Christchurch, Christchurch, New Zealand; 12) Department of Public Health, Government of American Samoa, Pago Pago, American Samoa; 13) International Health Institute, Department of Epidemiology, Brown University, Providence, Rhode Island, United States of America; 14) Department of Anthropology, Brown University, Providence, Rhode Island, United States of America

Current understanding of lipid genetics has come mainly from studies in European-ancestry populations; limited effort has focused on Polynesian populations, whose unique population history may provide insight into the biological foundations of variation in lipid levels. Here, we performed an association study of 4 lipid traits: total cholesterol, high-density lipoprotein (HDL), low-density lipoprotein, and triglycerides (TG) and genetic variants in 5q35, a region that has shown suggestive association with HDL in Micronesian and Polynesian peoples and no previous evidence of association in Europeans.

We performed association testing between variants on 5q35 and lipid levels in the discovery cohort comprising 2,851 Samoans using linear mixed modeling with imputed genotypes based on Samoans sequenced via TOPMed and adjusting for age, sex, principal components of ancestry, and relatedness. Fine-mapping analyses highlighted an association between a stop-gained variant (rs200884524; c.652C>T, p.R218\*) in *BTNL9* and HDL (discovery cohort:  $\beta = -1.99\text{mg/dL}$ ,  $p = 4.49 \times 10^{-8}$ ; replication cohort of 1,522 Samoans and American Samoans:  $\beta = -1.35\text{mg/dL}$ ,  $p = 0.0017$ ; replication cohort of 2,378 Māori and Pacific individuals living in Aotearoa/New Zealand:  $\beta = -0.90\text{mg/dL}$ ,  $p = 0.003$ ; meta-analysis of all cohorts:  $\beta = -1.35\text{mg/dL}$ ,  $p = 4.75 \times 10^{-11}$ ). rs200884524 is also associated with TG (discovery cohort:  $\beta = 14.38\text{mg/dL}$ ,  $p = 3.63 \times 10^{-6}$ , Samoan replication cohort:  $\beta = 8.29\text{mg/dL}$ ,  $p = 0.061$ , Aotearoa/New Zealand replication cohort:  $\beta = 6.67\text{mg/dL}$ ,  $p = 3.51 \times 10^{-5}$ ; meta-analysis:  $\beta = 8.3\text{mg/dL}$ ,  $p = 1.02 \times 10^{-9}$ ). This variant accounts for 1.02% and 0.7% of the variation in HDL and TG respectively in

the discovery cohort.

rs200884524 has a MAF of  $<0.0002$  in gnomAD; of the 50 alleles seen in gnomAD, 45 are from East Asians. However Samoans have a MAF of 0.218 in the discovery cohort (0.215 and 0.101 in the Samoan and Aotearoa/New Zealand replication cohorts respectively). Consistent with these disparate allele frequencies, this variant showed evidence of positive selection (PBS score of 0.477, 99.6 percentile in the Samoan genome;  $nS_L$  score of 1.78, 98.3 percentile in the Samoan genome).

BTNL9 is expressed in adipose, breast, and lung tissue (GTEx). Additional work is necessary to characterize the relationship between variation in *BTNL9* and lipid levels; however, these initial findings fine map the suggestive association in 5q35, providing evidence of a new contributor to the genetic architecture of lipids in Polynesian peoples.

## PgmNr 2040: A novel variant in *CETP* is associated with higher HDL-cholesterol in people of Polynesian ancestry.

### Authors:

M. Krishnan<sup>1</sup>; M. Leask<sup>2</sup>; T.J. Major<sup>2</sup>; J.C. Carlson<sup>3</sup>; J.Z. Zhang<sup>3</sup>; E.M. Russell<sup>1</sup>; R.L. Minster<sup>1</sup>; D.E. Weeks<sup>1,3</sup>; N.L. Hawley<sup>4</sup>; T. Naseri<sup>5</sup>; M.S. Reupena<sup>6</sup>; R. Deka<sup>7</sup>; H. Cheng<sup>7</sup>; S.T. McGarvey<sup>8,9</sup>; N. Dalbeth<sup>10</sup>; J. Zoysa<sup>10</sup>; R. Murphy<sup>10</sup>; L. Stamp<sup>11</sup>; J.H. Hindmarsh<sup>12</sup>; T.R. Merriman<sup>2</sup>; J. Moors<sup>2</sup>; TOPMed Lipids Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, University of Pittsburgh, Pittsburgh, PA.; 2) Department of Biochemistry, University of Otago, Dunedin, New Zealand.; 3) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.; 4) Department of Epidemiology (Chronic Disease), School of Public Health, Yale University, New Haven, Connecticut, USA.; 5) Ministry of Health, Government of Samoa, Apia, Samoa.; 6) Lutia i Puava ae Mapu i Fagalele, Apia, Samoa.; 7) Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, Ohio, USA.; 8) International Health Institute, Department of Epidemiology, School of Public Health, Brown University, Providence, Rhode Island, USA.; 9) Department of Anthropology, Brown University, Providence, Rhode Island, USA.; 10) Department of Medicine, Faculty of Medical and Health Sciences, University of Auckland, Auckland, New Zealand; Maurice Wilkins Centre.; 11) Department of Medicine, University of Otago Christchurch, Christchurch, New Zealand.; 12) Ngati Porou Hauora, Te Puia Springs, New Zealand.

Common variants in Cholesteryl ester transfer protein (*CETP*) are associated with differences in HDL-cholesterol (HDL) levels in most ancestral groups. However, no studies have examined whether genetic variation in *CETP* associates with HDL among people of Polynesian ancestry.

An HDL GWAS using the Illumina CoreExome platform was conducted in 2,594 people of Polynesian ancestry (Aotearoa/New Zealand [A/NZ] Māori, Cook Island Māori, Sāmoan, Tongan, Niuean and Pukapukan) living in A/NZ. Using 55 *CETP* gene sequences, genetic variants common in people of A/NZ Polynesian ancestry (>5%) but extremely rare (<0.001%) in gnomAD were selected for follow-up. A set of 2,851 Sāmoans (cohort I) using imputed genotypes based on Sāmoans sequenced via TOPMed and a set of 1,510 directly genotyped Sāmoans and American Sāmoans (cohort II) were used to replicate initial findings. The A/NZ analysis was adjusted for age, sex and principal components of ancestry, whereas the Sāmoan cohorts were adjusted for age, sex, principal components of ancestry and relatedness. Selection was calculated via the haplotype-based statistic (nSL) using the selscan program.

Three variants within *CETP* (rs183130, rs247617, rs9939224) were genome-wide significant in the A/NZ Polynesian. Closer inspection found a novel Polynesian-specific variant (NC\_000016.10:g56971035C>T, *CETP*:p.P177L) associated with higher HDL levels in the A/NZ ( $\beta$ =11.3 [9.1, 13.5] mg/dL,  $p=9.1\times10^{-24}$ ), Sāmoan cohort I ( $\beta$ =8.7 [7.4, 10.1] mg/dL,  $p=3.0\times10^{-37}$ ) and cohort II ( $\beta$ =8.3 [6.7, 9.9] mg/dL,  $p=4.5\times10^{-25}$ ). The MAF for the A/NZ Polynesian, Sāmoan cohorts I and II were 3.4%, 4.4% and 4.9% respectively but this variant is unobserved in gnomAD. The



variances in HDL explained by the variant in the A/NZ cohort and S?moan cohort I were 7.7% and 5.1% respectively. Analysis of this variant showed moderate evidence of positive selection (nSL score of 1.52 with an empirical percentile of 92%) in the S?moan cohort I. A CADD score of 25.6 places the variant in the top 1% of predicted deleteriousness. Secondary analysis conditional on this variant shows that all other variants associated with HDL within or close to *CETP* did not reach genome-wide significance.

Our results show that a unique variant in *CETP* contributes strongly to the phenotypic variation of HDL in people of Polynesian ancestry. This improved understanding of *CETP* variation in people of Polynesian ancestry allows insights into differences in control of HDL in this population.

## PgmNr 2041: Genetic drivers of serum lipid and liver disease co-heritability are enriched in the FXR/RXR pathway.

### Authors:

S.K. Handelman <sup>1,2</sup>; X. Du <sup>1</sup>; Y. Chen <sup>1</sup>; E.K. Speliotes <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Gastroenterology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109; 2) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109

**Background:** Nonalcoholic fatty liver disease (NAFLD) is epidemiologically associated with high triglycerides (TG), high low-density lipoprotein cholesterol (LDL-C) levels, and low high-density lipoprotein cholesterol (HDL-C) levels. Part of the global obesity epidemic, NAFLD is projected to be the number one cause of liver disease worldwide by 2020. Whether NAFLD and blood lipid levels may share a common genetic etiology is unknown.

**Methods:** We used gene set enrichment analysis (GSEA) implemented in MAGENTA to identify pathways enriched for lipid GWAS associations from published summary statistics, then tested these pathways for enrichment with NAFLD associations from two populations with computed tomography measured NAFLD. Effects of variants in enriched pathways were independently tested on liver function tests and serum lipids in participants from the UKBiobank.

**Results:** European genetic associations from the Global Lipids Genetics Consortium (N >95,000) were tested for pathway enrichment using MAGENTA, identifying 58 pathways enriched for HDL-C, LDL-C, or TG GWAS associations. CT measured NAFLD GWAS associations from individuals of European-ancestry (N=7,126) were then tested for enrichment in these 58 lipid-associated pathways, and three pathways (FXR/RXR Activation, Hepatic Cholestasis, and Chylomicron Mediated Lipid Transport) were found to be enriched (P < 0.01). Of these three pathways, only FXR/RXR Activation replicated this enrichment for CT measured NAFLD genetic associations in individuals of African-ancestry (N=3,124) (P=0.0058). The FXR/RXR pathway was not enriched for associations with other comorbid traits including blood pressure (N=275,000), body mass index (N= 249,796), and waist to hip ratio (N=77,167) obtained from publically available data, suggesting that the enrichment was specific to NAFLD and serum lipids. Within the FXR/RXR pathway, genes showing high credibility (Bayes factor > 3 vs. other genes in the pathway) include PPARA, APOB and SLCO1B1. Variants in the FXR/RXR pathway replicate liver damage decreasing alleles increasing LDL cholesterol in the UKBiobank (N=389,565).

**Conclusions:** These results identify the FXR/RXR pathway as having an overlapping genetic etiology for dyslipidemia and fatty liver. Variants in the FXR/RXR pathway have opposite effects on NAFLD and serum lipids consistent with reported FXR agonist effects [PMID 28145001, PMID 27109453], suggesting that human genetics can help inform drug effects.

## PgmNr 2042: Phenome-wide copy number variation burden association study on 23 phenotypes in 10,000 participants from the eMERGE Network.

### Authors:

M. Shivakumar <sup>1</sup>; J. Glessner <sup>2</sup>; A. Lucas <sup>3</sup>; Y. Bradford <sup>3</sup>; S. Verma <sup>3</sup>; J. Haines <sup>4</sup>; P. Peissig <sup>5</sup>; J. Denny <sup>6</sup>; M. Basford <sup>6</sup>; D. Crosslin <sup>7</sup>; J. Harley <sup>8</sup>; L. Rasmussen-Torvik <sup>9</sup>; I. Kullo <sup>10</sup>; E. Larson <sup>11</sup>; P. Avillach <sup>14,15</sup>; M. Brilliant <sup>5</sup>; J. Connolly <sup>2</sup>; G. Tromp <sup>12</sup>; P. Crane <sup>7</sup>; J. Pacheco <sup>9</sup>; H. Hakonarson <sup>2</sup>; C. McCarty <sup>13</sup>; M. Ritchie <sup>3</sup>; D. Kim <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania; 2) The Center for Applied Genomics, Abramson Research Center, Children's Hospital of Philadelphia; 3) Department of Genetics, University of Pennsylvania; 4) Department of Population and Quantitative Health Sciences, Case Western Reserve University; 5) Marshfield Clinic Research Institute; 6) Department of Biomedical Informatics, Vanderbilt University; 7) Biomedical Informatics and Medical Education, University of Washington; 8) Center for Autoimmune Genomics & Etiology, Cincinnati Children's Hospital; 9) Center for Genetic Medicine, Northwestern University; 10) Department of Cardiovascular Medicine, Mayo Clinic; 11) Kaiser Permanente Washington Health Research Institute, Kaiser Foundation Health Plan of Washington; 12) Department of Biomedical Sciences, Stellenbosch University; 13) Department of Family Medicine and Biobehavioral Health, University of Minnesota Medical School; 14) Boston Children Hospital; 15) Center of Biomedical informatics, Harvard Medical School, Boston

---

Recent studies on copy number variations (CNV) have suggested that an increasing burden of CNVs is associated with susceptibility or resistance to disease. However, no systematic CNV burden association analysis has been done on a wide range of phenotypes. We performed a phenome-wide CNV burden association analysis to explore global CNV burden effects associated with multiple phenotypes in subjects from the electronic Medical Records and Genomics (eMERGE) Network. The CNV calls were generated using PennCNV and QauntisNP and retained only the regions with at least 50% intersection. We tested for significant associations between CNV burden, measured as base pairs of duplication and deletion, in 10,525 samples and 23 phenotypes based on ICD9 diagnosis and lab values from five eMERGE sites. Logistic or linear regression models were fit to the data to evaluate the associations between CNV burden and phenotypes, adjusting for age, sex, the first three principal components of ancestry, and eMERGE sites. The datasets per phenotype were split into discovery and replication dataset based on the genotyping center. However, LDL and HDL, which were only available from a single genotyping center, were randomly and equally split into discovery and replication dataset. Then, we ran association tests for the discovery and replication dataset as well as the combined dataset. We found several statistically significant associations replicated between cholesterol phenotypes and duplication burden in Total Cholesterol Median (discovery FDR =  $2.7 \times 10^{-4}$ , replication FDR =  $9.6 \times 10^{-2}$ , combined FDR =  $1.3 \times 10^{-2}$ ) and LDL (discovery FDR =  $3.7 \times 10^{-2}$ , replication FDR =  $9.6 \times 10^{-2}$ , combined FDR =  $2.7 \times 10^{-3}$ ). Triglyceride Median was also significantly associated in discovery and combined dataset (discovery FDR =  $3.7 \times 10^{-2}$ , combined FDR =  $5.5 \times 10^{-2}$ ), but was not replicated. Further, we performed a pathway-based CNV burden analysis to better understand the

biological functions of the genes within duplication for 2 significant cholesterol phenotypes. We found 11 pathways associated with Total Cholesterol Median (FDR < 0.05), including fat digestion and absorption, systemic lupus erythematosus, and PPAR signaling, which are already known to be associated with cholesterol phenotypes. These findings indicate that cholesterol-related phenotypes might be more associated than other traits with global CNV burden. This study warrants further exploration in other complex diseases caused by high cholesterol.

# PgmNr 2043: A phenome-wide and genome-wide analysis of triglyceride levels in the Healthy Nevada Project cohort.

## Authors:

J.J. Grzyski<sup>1,2</sup>; R. Read<sup>1,2</sup>; G. Elhanan<sup>1,2</sup>; W.J. Metcalf<sup>1,2</sup>; K.A. Schlauch<sup>1,2</sup>; A. Slonim<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Institute for Health Innovation, DRI, Reno, NV; 2) Institute for Health Innovation, Renown Health, Reno, NV

---

Background: In 2017, Renown Health and DRI conceived the Healthy Nevada Project (HNP): merging an extensive Electronic Health Record (EHR) database and genotypic data of 10,000 individuals based on the Illumina Human OmniExpress-24 BeadChip in a single Northern Nevada cohort. Methods: Here, a genome-wide association study (GWAS) was performed on 4,960 HNP participants to assess genotypic association of triglyceride levels against 498,709 quality-controlled genotypes. Standard quality control measures were implemented (>95% SNP call rates, >95% individual call rates, Hardy-Weinberg Equilibrium  $p > 1 \times 10^{-6}$ , minor allele frequency >.5%). PLINK v. 1.9 performed associations using age, gender, BMI, and Type 2 Diabetes (DM2) as covariates, and the standard log-additive genetic model. Two independent PheWAS analyses were performed: one examined associations between significant SNPs identified in the GWAS and EHR phenotypes based on ICD codes; the second PheWAS identified associations between triglyceride levels and ICD-based diagnoses. Both used the R package **PheWAS**. ICD9 and ICD10 codes were aggregated via a mapping from the Center for Medicare and Medicaid services. A polygenic weighted risk score was computed via SNP effect sizes and minor allele counts; linear regression on transformed data measured its contribution on phenotypic variance. Results: This study identified 48 statistically significant SNPs associated ( $p < 5 \times 10^{-8}$ ) with triglyceride levels, many reported in prior studies. SNP rs17159164 in the *DOCK4* gene ( $p = 7 \times 10^{-14}$ ); SNPs rs1260326, rs780094, and rs780093 in *GCKR* ( $p < 2 \times 10^{-11}$ ); and rs2544164, rs426968 and rs2598272 in *LRGUK* ( $p < 2 \times 10^{-8}$ ) were previously linked with triglyceride response, hypertriglyceridemia, and DM2, respectively. Strong associations with SNPs and lipid levels, metabolic syndrome, and Crohn's disease were also identified. The risk score based on the 48 SNPs explained 11% of the cohort variation in triglyceride levels ( $p = 3 \times 10^{-14}$ ). PheWAS results indicated that triglyceride levels associate strongly with DM2 ( $p < 1 \times 10^{-35}$ ), an established association, obesity ( $p < 1 \times 10^{-28}$ ), hypertension ( $p < 1 \times 10^{-18}$ ), and other phenotypes not yet previously reported. SNPs identified in the GWAS associated with other phenotypes: abnormal glucose levels, anemia, DM2, gout, and vascular disease. Conclusion: This HNP study validates prior genomic and phenotypic associations with triglyceride levels and points to additional novel links using the Healthy Nevada Project.

## PgmNr 2044: Information management systems for biobanking in Taiwan Biobank.

**Authors:**

H. Chu; C. Liang; T. Lee

[View Session](#) [Add to Schedule](#)

**Affiliation:** Taiwan Biobank, Academia Sinica, Taipei, Taiwan

---

Taiwan biobank (TWB) is a large population-based cohort study and biobank that was established as a resource for research on complex interactions between environmental, phenotypic and genomic factors in the development of chronic diseases and healthy ageing. Our aim is to improve the health of future generations. TWB plans to recruit a large and comprehensive prospective study with 200,000 participants between the age of 30 and 70 from the general population with no history of cancer, and another 100,000 patients with chronic diseases of public health importance. This study has collected extensive phenotypic and genotypic details from individual participants, including the data from questionnaires, physical examination, biomedical test, image inspection, experimental information, biospecimens, omics information, and longitudinal follow-up for a wide range of health-related outcomes.

In order to manage the recruitment and information status of disease participants, TWB has developed a variety electronic systems, including Central Authentication Management System (CAMS), Subject Personal Information Management System (SPIMS), Subject Tracking Management System (STMS), Clinical Data Management System (iQuestion), and Laboratory Information Management System (LIMS). With the help of these systems, TWB employees can quickly and systematically synchronize data with each recruitment station and center. To address the importance of privacy protection issue, these systems depend on different authorizations for different employees with different roles in TWB.

TWB is available for open access to any bona fide researchers who wish to use it to conduct research. To achieve our goal, TWB will start to release the information more actively. Through international cooperation, TWB will to be a resource for the national and international scientific community to data request. The release of this large-scale population data will greatly benefit human biomedical research.

## PgmNr 2045: Age-related changes in gene expression of muscle and cartilage reflect overlapping aging biology.

### Authors:

E. Quillen

[View Session](#) [Add to Schedule](#)

**Affiliation:** Molecular Medicine, Wake Forest School of Medicine, Winston Salem, North Carolina

Aging is a key risk factor for nearly all common, complex diseases. However, age-related changes in gene expression across tissues are poorly understood, hindering our ability to distinguish between shared and tissue-specific alterations in transcription that may contribute to diseases of aging. This study examines this question in muscle and cartilage – two musculoskeletal tissues of distinct developmental origins – in 19 baboons (*P. anubis*).

At necropsy, tissue samples were collected from the distal vastus lateralis and unblemished lateral articular cartilage of the distal femur. No live animals were used or euthanized for the purposes of this study. RNA was extracted and mRNA libraries sequenced on the Illumina HiSeq2500. Reads with a Phred score greater than 30 were retained and aligned to the Panu 3.0 (Ensembl 92) reference genome using the STAR aligner in Partek Flow. Transcripts were collapsed to gene level and TPM normalized. Genes with non-zero counts in at least 7 animals were included in subsequent analyses resulting in 14,174 genes for the muscle analysis and 14,178 genes for cartilage.

Differential expression between young ( $N = 10$ ,  $12.9 \pm 1.44$  years, human equivalent 38.7 years) and old ( $N = 9$ ,  $19.7 \pm 1.33$  years, human equivalent 59.1) animals was evaluated using a Student's t-test for each tissue. Ranked by p-value, 43 of the 100 most differentially expressed (DE) genes in muscle and cartilage overlapped with all showing the same direction of change (69.8% decreasing with age). These values are nearly identical when the top 500 most DE genes are considered. The most DE gene between the young and old animals for both tissues was *GCAT*, a nuclear gene involved in glycine production in mitochondria that has been linked to biological aging. Based on DAVID annotation, the 43 overlapping genes are significantly enriched for the GO term neuromuscular junction, suggesting some specificity in musculoskeletal aging. Strong DE was also seen for *ERCC8*, a gene involved in DNA repair, and transcription factor *NFAT5*, both of which are ubiquitously expressed. Interestingly, while the top 100 DE genes in the muscle sample fell within the top 218 most DE cartilage samples, 25 of the 100 most DE cartilage genes are ranked in the bottom half of muscle genes.

Future work will examine the expression of DE transcripts in trabecular bone and incorporate proteomic data into a multi-omic analysis to further evaluate tissue-specific aging in the musculoskeletal system.

## PgmNr 2046: Executive function GWAS in a multi-ethnic cohort implicates region on chromosome 1.

### Authors:

N. Dueker<sup>1</sup>; H. Gardener<sup>2</sup>; L. Gomez<sup>1</sup>; A. Beecham<sup>1</sup>; L. Wang<sup>1,3</sup>; S.H. Blanton<sup>1,3</sup>; C. Dong<sup>2</sup>; S. Tom<sup>4</sup>; R.L. Sacco<sup>2,5,6</sup>; T.R. Rundek<sup>2,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida.; 2) Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL USA; 3) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL USA; 4) Departments of Neurology and Epidemiology, Columbia University, New York, NY USA; 5) Department of Public Health Sciences, Miller School of Medicine, University of Miami, Miami, FL USA; 6) Evelyn F. McKnight Brain Institute, Department of Neurology, University of Miami, Miami FL USA

Cognitive abilities have a significant impact on quality of life and cognitive impairment is a recognized public health burden. Cognition is comprised of several domains including executive function, which represents higher-order cognitive abilities. Measures of executive function are heritable. A large GWAS in European Ancestry individuals identified *CADM2* to be significantly associated with executive function. However, genetic studies in minority populations are lacking. Therefore, we performed a GWAS in the Northern Manhattan Study, a multi-ethnic cohort to identify common variants (CVs) influencing executive function. A neuropsychological exam was administered to participants and executive function was assessed using the difference in time to complete color trails test forms 1 and 2 as well as the odd-man-out subtests 2 and 4. A Z score for executive function was calculated by averaging the Z-transformed test scores on these tests. Participants were genotyped using the Affymetrix 6.0 array and these data were imputed to the 1000 genomes phase 1, v3 reference panel. Linear regression, implemented in PLINK, was performed to test the association of CVs (minor allele frequency > 5%) with the executive function Z score, adjusting for age, sex, education and principal components of ancestry. Analyses were performed separately within the three race/ethnic groups (Hispanic, non-Hispanic Black and non-Hispanic White) and then meta-analyzed. A p-value <  $5.0 \times 10^{-8}$  was considered significant. Our GWAS revealed an intergenic region on chromosome 1 to be significantly associated with increased executive function (top SNP rs2788328  $\beta=0.22$ ,  $p=8.1 \times 10^{-10}$ ). A replication cohort with a similar race/ethnic distribution was not available; however, we performed a look-up in the CHARGE European Ancestry sample and found no association between this region and executive function. To explore the findings in our study further, we additionally tested rs2788328 for association with each of the individual tests comprising the combined executive function Z-score. These analyses revealed a strong association with color trails test form 2 ( $\beta=0.26$ ,  $p=2.2 \times 10^{-8}$ ) and performance on the odd-man-out tests ( $\beta=0.17$ ,  $p=5.6 \times 10^{-5}$ ). No association was observed between rs2788328 and color trails test form 1 ( $\beta=0.02$ ,  $p=0.66$ ). Together, these findings present new insight into the genetic underpinnings of cognition, as it pertains to executive function, using an underrepresented population.



## PgmNr 2047: Hacking the aging genetic code: Analysis of large whole exome datasets in two extreme long-lived populations.

### Authors:

P. Sin-Chan <sup>1</sup>; N. Gosalia <sup>1</sup>; C. Van Hout <sup>1</sup>; B. Ye <sup>1</sup>; D. Liu <sup>1</sup>; A. Marcketta <sup>1</sup>; G. Abecasis <sup>1</sup>; A. Baras <sup>1</sup>; J. Overton <sup>1</sup>; A. Franke <sup>2</sup>; N. Barzilai <sup>3</sup>; C. O'Dushlaine <sup>1</sup>; A. Shuldiner <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Regeneron Genetics Centre, Regeneron Pharmaceuticals, Inc., Tarrytown, New York, USA; 2) Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Kiel, Germany; 3) Department of Genetics, Albert Einstein College of Medicine, New York, New York, USA

---

Aging is a complex process characterized by progressive degeneration in cellular and organismal functions, leading to increased disease susceptibility and death. While our understanding of biological drivers of aging has improved dramatically, most studies have focused on animal models with extremely short lifespans. Large-scale, high-throughput DNA sequencing studies of human aging are still at their infancy. Deciphering the genetic mechanisms of aging may help understand causes of many age-related pathologies and healthy aging. To define sex-specific age thresholds of long-lived individual (LLI) status, we utilized data from the United States National Vital Statistics Reports (1997-2015) and calculated the top 10% of extreme-long lived females and males lived  $\geq 95$  and  $\geq 92$  years, respectively. We performed genome-wide association analyses in the largest dataset of LLI to date, including two independent cohorts of LLI of Ashkenazi Jewish and German descent. Exome sequence data were merged with genome-wide SNP chip genotypes and imputation performed. A total of 12,975,494 and 14,612,329 variants with MAF  $>1\%$  were analyzed from German and Ashkenazi Jewish cohorts, respectively. We next performed a Case:Control analysis in each cohort with 663 Ashkenazi Jews and 1,294 German LLI compared to 410 and 2,149 controls (Females  $<95$ , Males  $<92$ ), respectively, followed by meta-analysis. We observed significant associations to variants at *APOE* and *FOXO3A* after Bonferroni correction, as well as nominally associated population-specific variants in genes involved in lipid transport and mitochondrial function. Additionally, we compared between LLI and controls a genome-wide sum of rare (MAF  $<1\%$ ) predicted loss-of-function variants (pLoFs) and predicted deleterious missense variants. Interestingly, the LLI showed a significant small increase in heterozygous pLoF and predicted deleterious missense variants, but a significant small decrease of homozygous variants. These findings suggest protective effects of haploinsufficiency for certain genes and deleterious effects of complete loss of function for other genes. Ongoing work is focused on determining the genes with variants in LLI and control individuals and follow-up of the identified genes and variants to gain mechanistic insights with the goal of informing therapeutics or other interventions to promote healthy human aging.

## PgmNr 2048: The genome-wide association study for the age distribution of facial metrics in Korean populations.

**Authors:**

K.W. Hong; H.J. Lee; B.R. Park

[View Session](#) [Add to Schedule](#)

**Affiliation:** TheragenEtex Bioinstitute Co. Ltd., Suwon, Suwon, Korea, Republic of

---

Growth and alteration in craniofacial morphology arouses interest in many fields of science, especially physical anthropology, genetics and in the scope of forensic sciences. We performed the analysis of craniofacial morphology alteration by the gender and the aging stage in Korean populations. We studied 15 facial metrics using two large Korean populations (1,926 samples of Korea medicine Data Center cohort and 5,643 of Ansan-Ansung cohort). Among them, total 12 metrics showed the gender differences and tendencies of aging alteration. Increasing tendencies according to aging were observed in Brow ridge height, Upper lip height, Nasal tip height, Profile nasal length and decreasing tendencies were observed in Outer canthal width, Right palpebral fissure height, Left palpebral fissure height, Right upper lip thickness, Left upper lip thickness, Nasal tip protrusion, Facial base width, Lower facial width, consistent in both independent populations. We conducted the genome-wide association studies for the facial metrics using Asian Precision Medicine Research Array. In conclusion, our findings suggest that aging (over 40 ages) might affect sizes of eyes, nose length, the upper lip thickness, and facial width, possibly due to loss of elasticity on the face. Therefore, the facial metric change could be applied to individual's age prediction and the esthetic facial care.

## PgmNr 2049: *BET1L* is associated with survival time in the Japanese population.

### Authors:

M. Akiyama<sup>1,2,3</sup>; S. Sakaue<sup>1,4</sup>; A. Takahashi<sup>1,5</sup>; K. Ishigaki<sup>1</sup>; M. Hirata<sup>6</sup>; K. Matsuda<sup>7</sup>; Y. Momozawa<sup>8</sup>; Y. Okada<sup>1,4</sup>; C. Terao<sup>1,2</sup>; Y. Murakami<sup>9</sup>; M. Kubo<sup>10</sup>; Y. Kamatani<sup>1,2,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.; 2) Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.; 3) Department of Ocular Pathology and Imaging Sciences, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Fukuoka 812-8582, Japan.; 4) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, Japan.; 5) Department of Genomic Medicine, Research Institute, National Cerebral and Cardiovascular Center, Osaka 565-8565, Japan.; 6) Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan.; 7) Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical Sciences, Graduate school of Frontier Sciences, The University of Tokyo, Tokyo 108-8639, Japan.; 8) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.; 9) Division of Molecular Pathology, The Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan.; 10) RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan.; 11) Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan.

---

Human lifespan is considered to be a heritable trait. Whereas previous studies have reported dozens of lifespan-associated genetic loci through genome-wide association studies (GWAS), the majority of these studies tested an association between parental death information and genotype of their children. Furthermore, most of such studies have been conducted in the European population. Here, we report the result of GWAS for survival time in 137,693 Japanese individuals who participated in the Biobank Japan (BBJ) project. Of these individuals, 31,324 (22.7%) were died in the average follow-up period of 7.44 years after the registration of BBJ. Following the genotype imputation using the EAS samples of 1000 Genomes project as a reference, we performed sex-stratified GWASs using 6,108,833 single nucleotide polymorphisms. A meta-analysis of the single genomic-control (GC) corrected sex-stratified GWASs revealed one locus satisfied genome-wide significance (*BET1L*;  $P_{\text{meta}} = 2.39 \times 10^{-8}$ , Hazard ratio = 0.92). Intriguingly, a statistically significant dominance effect was observed ( $P$  for dominance = 0.02). The lead variant rs76612380 is common in the Japanese population (minor allele frequency [MAF] = 11%); however, low-frequency (MAF < 5%) in non-East Asian populations of the 1000 Genomes project. When we categorized the causes of death into five categories (malignant neoplasms, cardiovascular diseases, cerebrovascular diseases, pneumoniae, others), the identified locus was not associated with them ( $P = 0.29$ ). By integrating the eQTL data of GTEx project, we observed a significant overlap between the association signal of GWAS and eQTL of *BET1L* gene in muscle skeletal. Furthermore, we found that the variant suggested as a protein truncating variant for *BET1L* in muscle skeletal by the splice disruption model of GTEx project was in linkage disequilibrium with the lead variant of the identified locus ( $r^2 = 0.98$ ). A gene-set enrichment analysis using PASCAL with reconstituted gene sets implemented by DEPICT implicated that genes

relevant to *BCAR1* protein-protein interaction (PPI) subnetwork were associated with survival time ( $P = 1.54 \times 10^{-7}$ ; false discovery rate =  $2.22 \times 10^{-3}$ ). These results suggest that individual differences in expression of *BET1L* in muscle skeletal may influence survival time, and provide insights into the genetic architecture underlying human lifespan.

## PgmNr 2050: Novel genes associated with facial skin features in Chinese population.

### Authors:

Y. Li; Y. Shan; D. Wang; X. Liu

[View Session](#) [Add to Schedule](#)

**Affiliation:** BGI-Shenzhen, Shenzhen, Guang Dong, China

---

### Background

Skin is our body's largest organ. It's affected by every aspect of our life, from genetics to diet and environments. Facial skin can signal youth, fertility and health. Recent studies have revealed that up to 60% of the facial skin variation between individuals can be attributed to genetic factors. However, existing GWAS studies have been using many subjective features of skins as phenotypes which may introduce biases and noises. This study is aiming to objectively assess various features of facial skin and explore genes contributed to skin variations in Chinese population.

### Method

The skin features were measured using the VISIA® Complexion Analysis System in 800 individuals. The photographic images were captured with standard, cross-polarized, parallel polarized, and ultraviolet light. Images were taken in two different close-up views (front and left lateral 37°) for each subject to quantify the scores for spots, wrinkles, pores, texture, and erythema. 30X whole genome sequencing were performed. A GWAS across 7,236,472 common variants (MAF >0.01) were conducted adjusting for gender and top two PCs of population stratification.

### Result

Three genome-wide significant signals were identified from variant-based association. One significant gene was identified from gene-based association. The associations between *OCA2-HERC2* region and pigmentation was confirmed (rs75295597,  $P = 4.46E-8$ ). Gene *SH3RF3* was associated with pores (gene-based  $P = 6.12E-7$ ). Novel signal rs13069665 ( $P = 4.48E-8$ ) was associated with textures. The signal tags the region of gene *RBMS3*, which has been reported to be associated with Systemic lupus erythematosus.

### Discussion

This is the largest skin genetic study in Chinese population. The identified genes could provide insights of mechanism of skin problems and aging, which could be used for developing better skin care approaches.

## PgmNr 2051: Novel metabolomic pathways associate with longitudinal declines in physical performance among middle-aged adults: Evidence from the Bogalusa Heart Study.

### Authors:

J.L. Nierenberg <sup>1</sup>; J. He <sup>1</sup>; C. Li <sup>2</sup>; X. Gu <sup>1,3</sup>; M. Shi <sup>1</sup>; A.C. Razavi <sup>1</sup>; X. Mi <sup>1</sup>; S. Li <sup>1</sup>; L.A. Bazzano <sup>1</sup>; J.M. Guralnik <sup>4</sup>; A.H. Anderson <sup>1</sup>; H. He <sup>1</sup>; W. Chen <sup>1</sup>; J.M. Kinchen <sup>5</sup>; T.N. Kelly <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) Department of Epidemiology & Biostatistics, University of Georgia College of Public Health, Athens, GA; 3) Institute of Clinical Medical Sciences, China-Japan Friendship Hospital, Beijing, China; 4) Division of Gerontology, Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD; 5) Metabolon, Inc., Morrisville, NC

**Background:** Declining physical performance predicts cognitive impairment, reductions in quality of life, disability, chronic disease exacerbation, and mortality. We conducted a metabolome-wide association study to identify novel metabolites associated with change in physical performance measures over 2.9 years of follow up in middle-aged adults.

**Methods:** Bogalusa Heart Study participants (35% African American, 58% female, ages 34-58 years) underwent non-invasive physical performance measurement at baseline and follow-up exams, using gait speed (N=1,227) and grip strength (N=1,164). Baseline serum metabolites were quantified using untargeted ultrahigh-performance liquid chromatography-tandem mass spectroscopy. Multivariate adjusted linear regression was used to assess the relationship between metabolites and baseline physical performance measures. Metabolites significant after Bonferroni correction were carried forward for assessment of association with declines in gait speed and grip strength, in a sub-sample with available data, over 2.9 years of follow up, with additional adjustment for baseline measures.

**Results:** Twenty-one metabolites associated with baseline gait speed and six with grip strength, including five metabolites that associated with both phenotypes ( $P < 4.2 \times 10^{-5}$ ). Two metabolites were associated with preservation or improvement in gait speed, including: behenoyl sphingomyelin (d18:1/22:0)\* ( $P < 0.05$ ) and sphingomyelin (d18:1/22:1, d18:2/22:0, d16:1/24:1)\* ( $P = 2.4 \times 10^{-3}$ ). Six metabolites were associated with decline in gait speed, including: gamma-glutamylphenylalanine, N2,N2-dimethylguanosine, 5,6-dihydrouridine, N-acetylalanine, N-acetylneuraminate, and N-formylmethionine (all  $P < 0.05$ ). Five of six metabolites tested for change in grip strength had consistent effect directions with the baseline analysis, however none met statistical significance thresholds. Sphingolipid Metabolism was the biological sub-pathway associated with a protective effect on gait speed. The metabolites most represented among those associated with detrimental effects on gait speed were modified nucleosides.

**Discussion:** These results add to the accumulating evidence suggesting an important role of the human metabolome in physical performance, and suggest that its underlying biological pathways may be implicated in frailty and accelerated aging phenotypes. Further research on genomic loci that predict the identified metabolites is warranted.

## PgmNr 2052: Genetic determinants of mosaic loss of the Y chromosome and their impact on age-related disease in the VA's Million Veteran Program.

### Authors:

B. Gorman<sup>1,3</sup>; S.G. Ji<sup>1,4</sup>; R.P. Igo, Jr.<sup>8</sup>; Y. Shi<sup>1</sup>; A. Sendamarai<sup>1</sup>; T.L. Assimes<sup>6,7</sup>; V. Napolioni<sup>5</sup>; P.S. Tsao<sup>6,7</sup>; N.S. Peachey<sup>9,10</sup>; S.K. Iyengar<sup>9,10</sup>; S. Pyarajan<sup>1,2</sup>; The Million Veteran Program

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) VA Boston Healthcare System, VA Cooperative Studies Program, Boston, MA; 2) Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Booz Allen Hamilton Inc., McLean, VA; 4) Seven Bridges Genomics Inc., Charlestown, MA; 5) Department of Neurology and Neurological Sciences, School of Medicine, Stanford University Stanford, CA; 6) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 7) VA Palo Alto Health Care System, Palo Alto, CA; 8) Department of Population and Quantitative Health Sciences, Case Western University, Cleveland, OH; 9) Department of Ophthalmology, Cleveland Clinic College of Medicine of Case Western Reserve University, Cleveland, OH; 10) Louis Stokes Cleveland Department of Veterans Affairs Medical Center, Cleveland, OH

Mosaic loss of Y chromosome (LOY) in peripheral blood cells is a phenomenon implicated in several aging and mortality-related disease processes in men (Forsberg *Hum. Genet.* 2017). The unique characteristics of the growing 460,000-person Million Veteran Program (MVP) cohort -- genetically diverse, majority male, and older (middle 50% age range of 54-70) -- are ideal for studying LOY. We measured LOY in men using the median log-R ratio of male-specific Y chromosome SNPs (mLRR-Y) on the Affymetrix Biobank Array. This array intensity-based metric was validated by sequencing a subset of the same samples (N=2000) and comparing intensity to the mapping frequency of Y reads.

We classified 421,326 male samples in MVP according to genetic ancestry. Overall, 25,255 of 288,388 (8.8%) of men with European ancestry (EUR), median age 66, showed evidence of LOY. A lower frequency of LOY was observed in African American (AA) men (1,909 of 73,795, or 2.6%; age 60) and Hispanic American men (1,016 of 30,321, or 3.3%; age 60). LOY frequency differences between populations were highly significant even considering the difference in age distribution ( $p=3E-142$  between EUR and AA).

We then conducted a GWAS separately amongst male EUR and AA populations on LOY, with age and smoking status as covariates. A GWAS on 236,404 EUR samples identified 70 loci with genome-wide significance, which included 18 of the 19 loci reported by an earlier UK Biobank-based study (the remaining one had  $p=6.5E-07$ ). Further, a GWAS on 65,150 AA samples identified 4 loci with genome-wide significance, of which one did not overlap with the EUR-based GWAS. In the non-overlapping locus, all 11 variants that were genome-wide significant had a MAF between 2-3% in AA samples, but less than 0.1% MAF in EUR samples. However, other variants in chr2.p13 showed significant associations in the EUR-only GWAS, suggesting different LD structure.

The GWAS results were followed up using a summary statistics-based colocalization analysis across

multiple traits associated with LOY. For regions that share a causal variant between LOY and a trait with a posterior probability greater than 0.9, statistical fine-mapping based on meta-analyzed summary statistics was applied. The list of causal variants was further narrowed by combining signals across the diverse ancestries in MVP. Thus, the large sample size and unique characteristics of MVP enabled the discovery of 52 novel loci and further elucidation of LOY in aging and disease.



## PgmNr 2053: Comparison of multiple omics aging clocks.

### Authors:

E. Macdonald-Dunlop <sup>1</sup>; P.K. Joshi <sup>1</sup>; J.F. Wilson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, United Kingdom; 2) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom

---

**Background:** Biological age, a measure of deterioration and aging that is distinct from chronological age has been found to predict disease and mortality (Fischer et al., 2014). Since the first published measure of biological age: Hannum's epigenetic clock, closely followed by Horvath's epigenetic clock, aging clocks have been built using telomere length (Zhang et al., 2014), facial morphology (Chen et al., 2015), metabolomics (Fischer et al., 2014), glycomics (Kristic et al., 2014) and proteomics (Enroth et al., 2015). However, comparison between clocks has so far been limited. The ORCADES study is a deeply annotated cross-sectional cohort with approximately 1,000 individuals, where multiple omics assays such as circulating plasma proteins, lipids, glycans and metabolomics are available, enabling consistent comparisons in a common dataset.

**Aims:** (i) To replicate and potentially improve published aging clocks, and partition aging across omics assays (ii) for the first time to compare their accuracy in predicting age in a single dataset (iii) for the first time to analyse the effect of giving up smoking on the evolution of biological clock age (iv) to test the association between age acceleration estimates and ageing and disease phenotypes.

**Methods:** Elastic net regression were used to create multiple omics clocks in a training set and their accuracy assessed in a testing set and an additional independent cohort, in univariate and bivariate analyses.

**Results:** We replicate existing clocks, (with some improvement of our proteomics clock over Enroth et al's,  $r = 0.83$ ) in predicting chronological age. We find our protein clock ( $r = 0.93$ ) outperforms our IgG glycan and metabolite clocks ( $r = 0.74$  and  $r = 0.77$  respectively). In pairwise bivariate analysis we find large but not complete overlap between clocks. We show that there is a significant difference of excess biological age (3-4 years) between current and non-smokers of the same chronological age and for the first time, the effect of having formerly been a smoker on biological age. Finally, we show that increased age acceleration estimated by clocks are consistently associated with increased BMI, cortisol and stroke across omics clocks.

**Conclusions:** Proteins more accurately predict chronological age than glycans and glycans can track the benefits of giving up smoking. Multi-omics clocks appear to track both overlapping and distinct aspects of aging and offer the prospect of identifying distinct aging pathways.

## PgmNr 2054: Tobacco smoking has no influence on buccal telomere length among Portuguese master students.

### Authors:

M. Brito <sup>1</sup>; P. Azevedo <sup>2</sup>; A. Coelho <sup>1</sup>; L. Veiga <sup>1</sup>; A. Costa-Veiga <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Health and Technology Research Center, Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Lisbon, Lisboa, Portugal; 2) Escola Superior de Tecnologia da Saúde de Lisboa, Instituto Politécnico de Lisboa, Portugal; 3) CISP, Centro de Investigação em Saúde Pública, Escola Nacional de Saúde Pública, Universidade Nova de Lisboa, Portugal

---

In Portugal, the average age of starting tobacco consumption is around 16 years, with an increase on consumption in recent years. The use of tobacco is a risk factor for age-related diseases, being one of the main causes of death and a risk factor for several chronic diseases, including diseases of the cardiovascular system, diabetes and cancer. Different studies shown a clear relationship between tobacco smoking and oxidative stress, which is one of the factors that accelerate the shortening of telomers. The aim of this study was to evaluate the effect of tobacco smoking on the relative length of telomers in master students as a biomarker of health status. A cross sectional study was performed by assessing relative Telomere Length (TL) in 131 master students. TL was evaluated from buccal swabs in 44 males and 89 females (aged between 18 and 52 years, mean 26.4) by multiplex quantitative Polymerase Chain Reaction (PCR). This technique consists in determining the relative ratio (T/S) between the telomere region copy number (T) and a single copy gene (S), albumin gene, using a relative standard curve. Mann-Whitney tests were used to detect TL differences between smokers and non-smokers and addiction levels (Light vs Moderate or Heavy). Statistical analysis were adjusted for age, sex and BMI. Age and TL was negatively correlated as expected (Spearman  $Rho = -1.66$ ,  $p = 0.080$ ). There were no statistically significant differences in TL between smokers and non-smokers, addiction levels or gender (Mann-Whitney tests,  $p > 0.10$ ). A negative association between buccal TL and tabaco smoking was not confirmed. One possible explanation can be the high turnover rate in buccal exfoliate cells relative to other cells such as lymphocytes. Although the association has not been significant with TL, the negative effect of tobacco consumption on the risk of cancer risk and respiratory alterations is largely confirmed. Future studies should include a higher number of participants. Although buccal exfoliate cells is a less invasive method, TL can be more specific using lymphocytes cells. This project was supported by a grant from Instituto Politécnico de Lisboa, ref. IPL/2018/SCITelomer/ESTeSL.

## PgmNr 2055: Modified ACMG classification criteria for mitochondrial tRNA variants.

### Authors:

T. Chen <sup>1,3</sup>; J. Wang <sup>1,4</sup>; S. Tang <sup>1,5</sup>; V. Zhang <sup>1,4</sup>; W. Craigen <sup>1,2</sup>; L. Wong <sup>1,2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TEXAS, USA; 2) Baylor Genetics Laboratory. Houston, Texas, USA; 3) Department of Endocrinology, Genetics and Metabolism, Children's Hospital of Soochow University, Suzhou, China; 4) AmCare Genomics Lab, Guangzhou, China; 5) WuXi NextCODE, Cambridge, Massachusetts, USA

---

### Background:

Standards and guidelines developed by ACMG and AMP have been widely used for the interpretation of nuclear-encoded protein coding variants. However, these guidelines need to be modified for the classification of mitochondrial (mt) DNA variants due to characteristics of the genome, including varying tissue-specific reliance on mt functions, heteroplasmy, and interactions between the nuclear and mt genomes. In particular, pathogenic mt tRNA variants are a common cause of maternally inherited mt disorders. Thus, the classification of mt tRNA variants is unique for mt disorders and requires substantial modifications to current ACMG guidelines that are primarily for nuclear mRNA variants.

### Purpose:

To establish criteria for the classification of variants in mitochondrial tRNA genes and to validate their clinical utility.

### Methods:

Based on the current ACMG guidelines, we included the degree of heteroplasmy, multi-system distribution, the genotypes of matrilineal relatives, functional and structural studies, (eg. ETC and muscle histopathology), variant location in the tRNA structure, and clinical phenotypes, all of which are crucial for mitochondrial tRNA variant interpretation. MitoTIP scores are used as alternative supporting criteria. We used known reported mt tRNA variants found in our database to test and validate the suitability of our criteria, followed by the use of the criteria to score VUS, outliers and novel variants, using clinical cases as examples.

### Results:

Among the 552 mt tRNA variants in our database, 463 have been reported. Among them, 59 were pathogenic (P), 22 likely pathogenic (LP), 116 VUS, 18 likely benign (LB), and 337 benign (B) using our modified criteria. In 80 previously reported pathogenic variants, we classified 48 as P, 9 LP, 9 VUS, 1 LB, 13 B. Among 89 novel variants, 5 were P, 11 LP, 54 VUS, 7 LB, and 12 B. Sequencing of mtDNA in different tissues and matrilineal relatives, degree of heteroplasmy, functional studies, and clinical

phenotypes are uniquely important classification criteria for mtDNA tRNA variants.

**Conclusion:**

Proper interpretation of mtDNA tRNA variants is crucial for accurate clinical diagnosis and genetic counseling. Testing of mtDNA in different types of tissue of probands and in matrilineal relatives, quantification of heteroplasmy, and functional studies are key evidence besides population and computational data in determining the clinical significance of mtDNA tRNA variants.

## PgmNr 2056: Drug metabolism anomalies co-segregating with familial prostate cancers and aortic aneurysms.

### Authors:

S. Regan <sup>1</sup>; B.R. Braddock <sup>1</sup>; K. Nanavati <sup>2</sup>; R.R. Lebel <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) SUNY Upstate Medical University, Section of Medical Genetics, Syracuse, NY; 2) SUNY Upstate Medical University, Department of Family Medicine, Syracuse, NY

An 84 year old female sought the underlying cause of her complex medical history. This included numerous adverse reactions to multiple medications since childhood. She experienced staring spells and syncope beginning at an early age and persisting into adulthood, and had adverse reactions to anti-epileptic medications. She had a diagnosis of secondary porphyria due to chronic gastrointestinal complaints, sensitivity to sunlight, and venous thrombi. She was diagnosed with an ascending aortic aneurysm at age 71.

The patient had three brothers and no sisters, three sons, and one daughter. The family history is notable for aortic aneurysm (father and all brothers), prostate cancer (all brothers and one nephew), and adverse neurologic drug reactions (father, youngest brother, eldest son). Her second son carries a diagnosis of “secondary porphyria” and his son had a severe reaction to anesthetics. Her daughter experiences medication side effects, and had seizures during puberty.

The patient and her grandson were tested for variants in 19 drug metabolism genes. The patient had 10 different homozygous mutations and 2 heterozygous mutations. Seven different homozygous and 5 heterozygous mutations were identified in the grandson, 6 and 4 of which, respectively, matched those seen in the proband. The genes affected include *CYP1A2\*1F*, *CYP3A5\*3*, *VKORC1*, *SOD2*, *GSTM1*, *GSTT1*, *NAT2*, *CYBA*, *MDR1*, and *MTHFR*. The grandson likely inherited the majority of these mutations from the proband via her son. The subsequent drug metabolism pathway alternations may account for each individual’s severe reactions to medications.

No unifying factor has been found thus far to explain the patient’s clinical phenotype. Previous sequencing identified a T118M variant in the *PMP22* gene, a nonsense mutation in *FBN1*, and frameshift mutations in *SLC2A1*, *NOTCH3*, and *FECH*. An *ND4* gene alteration does not match any of the 68 known pathogenic variants. Mutational analysis for 23 genes associated with connective tissue disorders did not identify any pathogenic variants, despite the previously reported *FBN1* mutation. Upon whole exome sequencing of the patient, with a triad including her surviving brother and one of her sons, no clinically relevant alternations were reported in the exome analysis. This family illustrates in a dramatic way the increasingly appreciated phenomenon of multiple genetic disorders presenting in a single proband.

## **PgmNr 2057: Deciphering undiagnosed conditions by NGS: An infant with 45,X/46,X der(Y)t(Y;14), whirl-pooled-type pigmentary mosaicism, and ambiguous genitalia.**

### **Authors:**

V. Romero <sup>1</sup>; J. Pozo <sup>1</sup>; H. Kazuyoshi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Medicine, Universidad San Francisco de Quito, Quito, Pichincha, Ecuador; 2) University of Kanagawa, Kanagawa, Japan

---

Trisomy 14 is one of the aneuploidies compatible with live in a mosaic state and has been reported in approximately 80 patients. The predominant phenotype of trisomy 14 mosaicism includes prenatal and postnatal growth failure, ear abnormalities, congenital heart disease, developmental delay and genitourinary abnormalities. Other cases of trisomy 14 include those of maternal and paternal UPD, having each disorder different characteristics. We evaluated a 1-year-old baby born from young non-consanguineous parents with pigmentary mosaicism in whirl-pooled hyperpigmentation along the body and ambiguous genitalia with no visible testicles. Karyotype and FISH results concluded 45,X/46,X der(Y)t(Y;14). However, the physical examination did not correlate with paternal UPD described in other cases. Using Whole Exome Sequencing, we identified the genes affected by the chromosomal rearrangements, therefore provided proper counseling for the patient's future, and analyzed the ethical implication of a patient with disorders of sexual differentiation.

## PgmNr 2058: Atypical nested (B-D) 22q11.2 deletion involving *CRKL* in a fetus with severe growth restriction and mother with severe systemic autoimmune disease.

### Authors:

I. Lin<sup>5</sup>; Y. Afshar<sup>3,5</sup>; W.W. Grody<sup>1,2,4,5</sup>; F. Quintero-Rivera<sup>2,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, UCLA, Los Angeles, CA.; 2) Pathology and Laboratory Medicine, UCLA Center for Health Sciences, Los Angeles, CA; 3) Obstetrics and Gynecology, UCLA, Los Angeles, CA; 4) UCLA Clinical Genomics Center, Los Angeles, CA; 5) David Geffen School of Medicine, UCLA, Los Angeles, CA

22q11.2 deletion syndrome (22q11.2DS) patients present with a broad phenotypic spectrum including dysmorphic features, congenital heart defects (CHDs), T-cell deficiency, hypoplastic thymus, and hypocalcemia. We report on a prenatally diagnosed case of atypical 22q11.2DS, born to a 22-year-old with SSA+ severe systemic lupus erythematosus with cutaneous and discoid components, seronegative anti-phospholipid syndrome, lupus nephritis, chronic hypertension, and frequent infections. Due to non-reassuring antenatal surveillance, the fetus was delivered at 31wk gestation with normal cardiac imaging, leukopenia, anemia, hypophosphatemia and hypercalcemia.

At 24wk gestation, the fetus was diagnosed with severe intrauterine growth restriction (IUGR) and had an amniocentesis with FISH analysis with HIRA (TUPLE1) probes, and a chromosomal microarray analysis (CMA). FISH was normal. CMA identified a heterozygous loss of the 22q11.21 region, ~737 kb size, between low copy repeats (LCR) B and D, consistent with atypical nested 22q11.2DS. Subsequently, the mother was found to have the same deletion, which includes *CRKL* and 19 other genes but excludes *TBX1*. Echocardiogram of the mother revealed moderate concentric left ventricular hypertrophy with a decreased left ventricular ejection fraction.

Most 22q11.2DS results from a typical, non-inherited deletion between LCRs A and D (~3Mb), or A and B (~1.5Mb). The atypical deletion can be missed with a FISH screen, as in this case, and often have reduced penetrance of features and an inherited mutation from a mildly affected parent.

Deletion of *TBX1* has been implicated in causing CHD. Deletion of *CRKL* has been linked to immune deficiency. Since *CRKL* has key regulatory roles in hematopoietic cell signaling, 22q11.2DS patients may have a restricted repertoire of T-cells that leads to abnormal T-cell activation and differentiation of B-cells, with a decrease in the population of memory B-cells and an increase in transitional B-cells, explaining the decreased production of immunoglobulins and abnormal specific antibodies. This may contribute to autoimmunity, as seen in the mother.

This case highlights diagnostic pitfalls, broad phenotypic spectrum and presentation of 22q11DS, and new insight into the pathophysiology and inheritance of 22q11DS through the evaluation of parents affected with autoimmune disease. The role of *CRKL* in autoimmunity can also be investigated to understand immune deficiencies and CHD in patients with 22q11DS.

## PgmNr 2059: White matter abnormalities in the human brain of sex chromosome aneuploidies.

### Authors:

S. Liu; M. Yavi; L. Clasen; J. Blumenthal; A. Raznahan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Developmental Neurogenomics Unit, National Institutes of Mental Health, Bethesda, Maryland.

---

Sex chromosome aneuploidies (SCA) are a common group of genetic disorders (1/400 individuals) characterized by variation in the number of sex chromosomes. SCA has been associated with a wide range of behavioral and cognitive deficits and previous studies identified focal gray matter volume (GMV) alterations relative to typically developing controls using T1-weighted structural magnetic resonance imaging (sMRI) data. Here we ask whether SCA further alters regional white matter volume (WMV) in the human brain by applying deformation-based morphometry (DBM) to sMRI data from 169 typically developing controls (80 XY, 86 XX) and 137 participants with SCA (28 XXX, 58 XXY, 26 XYY, 20 XXYY, 5 XXXXY) aged 5-25 years. Genetic diagnosis of SCA was confirmed by karyotype, and all participants were scanned on a 1.5T GE scanner. A well-established DBM pipeline was applied to obtain WMV over the whole brain. We used regression models to examine X and Y chromosome dosage effects on WMV changes at each voxel controlling for total white matter volume and age. The significance level of family wise error  $< 0.05$  was set to correct for multiple comparisons across voxels. We identified significant X- and Y-chromosome dosage effects on WMV in major white matter bundles. Interestingly, these local WMV effects for X- and Y-chromosome aneuploidy converged in a similar spatial pattern, consistent with prior findings for SCA effects on regional cortical anatomy. Specifically, carriage of supernumerary X- and/or Y-chromosomes was associated with (i) WMV decreases in corona radiata, internal and external capsules, cerebral and middle cerebellar peduncles connecting cortex and striatum and brain stem, basal forebrain and cortex, and cerebellum and pons within each hemisphere, and (ii) WMV increases in genu and splenium of corpus callosum bridging two hemispheres in the frontal and posterior parts of the brain. The spatial distribution of these WMV effects was also congruent with prior reports that GMV reduces in basal forebrain, insular and subcortical areas and increases in bilateral posterior parietal cortices with mounting X- and/or Y-chromosome dosage. These data provide the first evidence that X- and Y-chromosome dosage variations influence regional white matter anatomy in humans. Together with co-occurring changes in regional gray matter, these findings specify candidate neuroanatomical substrates for the known impact of sex chromosome aneuploidy on human cognition and behavior.



## PgmNr 2060: Orofacial clefts with clubfoot phenotypes in an African population: Whole exome sequencing implicates multiple syndromes and genes.

### Authors:

L. Gowans <sup>1,2</sup>; P. Donkor <sup>2</sup>; S. Obiri-Yeboah <sup>2</sup>; A. Oti <sup>2</sup>; F. Arthur <sup>1</sup>; T. Busch <sup>3</sup>; J. Murray <sup>4</sup>; N. Sobreira <sup>5</sup>; A. Butali <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology, Kumasi, Ashanti Region, Ghana; 2) Dental School, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.; 3) Oral Pathology, Radiology and Medicine, The University of Iowa, Iowa City, Iowa; 4) Pediatrics, The University of Iowa, Iowa City, Iowa; 5) Medical Genetics, John Hopkins University, Baltimore, Maryland

---

Orofacial clefts (OFCs) are congenital malformations of human face and palate, presenting as either nonsyndromic or syndromic, with syndromic forms presenting with additional congenital malformations. OFCs exhibit multifactorial inheritance with both genetic and environmental factors playing crucial roles. As part of our OFC studies in Ghana, we recruited six probands that presented with cleft lip-clubfoot, and cleft lip and palate-clubfoot phenotypes, with no clinically identifiable additional malformations at the time of recruitment. Our aim was to identify causal coding variants that will help ascertain whether these OFC-clubfoot phenotypes were syndromic or that the clubfoot phenotype was just an associated anomaly by conducting whole exome sequencing of affected probands and available relatives. Analysis of exome data sought to determine rare missense, nonsense, splicing and indel variants that exhibited either autosomal dominant or autosomal recessive homozygous and compound heterozygous inheritance patterns. We highlighted variants in genes known to be associated with phenotypes related to the one being investigated using OMIM, mouse model, GWAS, HGMD and ClinVar. Implicated variants were validated by Sanger sequencing. Multiple syndromes/genes were implicated in four out of six probands. With the exception of one proband who was autosomal recessive compound heterozygote for *DIS3L2* (p.P293H and p.E524K) - Perlman syndrome, the three other probands have multiple syndromes. One proband possibly have Vici syndrome (*EPG5*: p.H532Y and p.R350C), as well as *BARX1* (p.K141N) and *MKI67* (p.R2304K) variants. Another proband possibly have either Fraser Syndrome (*FRAS1*: p.V2882L and p.L3387P), or Treacher Collins syndrome (*TCOF1*: c.2864\_2865GG) and *MKI67* (p.R2247H) variant. The last proband possibly have either Fraser syndrome (*FRAS1*: p.I3532V, p.I3869V and p.H3969Q), or Cardiomyopathy, dilated, 1LL/Left ventricular noncompaction 8 (*PRDM16*: c.T1188C;p.C396C), or CHARGE syndrome/Hypogonadotropic hypogonadism 5 with or without anosmia (*CHD7*: c.4248delA;p.T1416fs). Our results suggest that clubfoot and OFCs are two congenital anomalies that co-segregate in some syndromes with varying genetic causes. Therefore, there is a need for deep phenotyping to delineate the additional phenotypes that distinguish these cohorts from isolated OFCs and clubfoot. Our observations are crucial for syndrome diagnosis in OFC care, genetic counseling and elucidating OFC pathogenesis.

## PgmNr 2061: A case series of a mother and two daughters with a *GLI2* gene deletion demonstrating variable expressivity.

### Authors:

C. Elward <sup>1</sup>; J. Berg <sup>1</sup>; J. Oberlin <sup>1</sup>; L. Rohena <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Brook Army Medical Center, Fort Sam Houston, TX; 2) Department of Pediatrics, Long School of Medicine, University of Texas Health, San Antonio, TX

GLI2 is a zinc-finger transcription factor involved in the Sonic Hedgehog pathway, and pathogenic mutations in the gene have been reported to be in association with variable features from holoprosencephaly (HPE) to HPE features, namely facial dysmorphisms, to the milder phenotype characterized by hypopituitarism, mainly growth hormone deficiency, and post-axial polydactyly known as Culler Jones syndrome. Here we report the case of a family, two daughters and a mother, with an identical *GLI2* gene deletion with varying presentations. We also reviewed the literature and identified seven other families of at least two members that share a mutation or deletion of the *GLI2* gene, and the phenotypes frequently included anterior pituitary lobe hypoplasia, ectopic posterior pituitary lobe identified on magnetic resonance imaging, the resulting pituitary deficiencies, post-axial polydactyly, and facial dysmorphisms, most commonly cleft lip and palate. This series adds to the body of literature of families with this mutation and supports the autosomal dominant inheritance pattern and variable expressivity observed in these patients.

Table: Ratio of common features observed in *GLI2* gene mutation or deletion in study population

Affected Region	Feature	Ratio (# with feature/total patients)
Brain/Pituitary Anomalies		
	True holoprosencephaly	2/34
	Ectopic posterior pituitary	10/34
	Anterior pituitary hypoplasia	7/34
	Panhypopituitarism	8/34
	Isolated growth hormone deficiency	3/34
Facial dysmorphisms		
	Cleft lip and palate	5/34
	Hypotelorism	5/34
Limb involvement		
	Polydactyly	17/34

## PgmNr 2062: Helping solve medical mysteries: Using machine learning and systems biology to identify gene pairs that can cause digenic disease.

### Authors:

S. Mukherjee<sup>1,2</sup>; J. Meiler<sup>2,5,6,7</sup>; J.A. Capra<sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biological Sciences, Vanderbilt University, Nashville, TN, USA; 2) Center for Structural Biology, Vanderbilt University, Nashville, TN, USA; 3) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; 4) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; 5) Department of Chemistry, Vanderbilt University, Nashville, TN, USA; 6) Department of Pharmacology, Vanderbilt University, Nashville, TN, USA; 7) Leipzig University, Leipzig, Germany

---

Previous efforts in clinical genetics, focused on effects on variants in a single gene, have led to the identification of the genetic causes for many diseases. However, the diagnosis of rare diseases has been a challenge. The Undiagnosed Diseases Network (UDN) was formed to address the challenges of diagnosing rare genetic diseases, and has had success in diagnosing many patients. The UDN patients undergo whole exome or genome sequencing and the team identifies genetic causes based on how dysfunction any mutated gene affects phenotype. Nonetheless, more than half of enrolled UDN patients remain undiagnosed.

Our central hypothesis is that oligogenic inheritance patterns must be considered in the analysis of rare disease genetics and that new tools are needed to enable these analyses. A database of known digenic disease causing gene pairs (DIDA) has been published previously, and it was found that it was observed that digenic gene pairs have unique biological characteristics, such as shared pathways, co-expression and protein-protein interactions.

In this study, we implemented a random forest machine learning classifier to identify pairs of genes that have the potential to cause digenic disease. We integrate combinations of gene and gene-pair features derived from systems biology and evolution to train the classifier on known digenic pairs derived from DIDA vs putative non-digenic pairs generated using several different strategies.

A classifier trained using six systems biology features performed well (mean ROC AUCs ~ 0.8) for all non-digenic gene pair sets. The addition of evolutionary features increased performance (mean ROC AUCs ~ 0.9), and additional gene level features, such as number of phenotypes, increased performance further (mean ROC AUCs > 0.95). Phenotypes, loss of function intolerance, protein age and selection pressure features were the highest weight features. We further validated the trained classifiers on additional digenic pairs not included in DIDA and synthetic lethal interaction pairs.

Our work reveals whether pairs of genes have sufficient synergy to lead a digenic disease upon co-mutation, and helps to uncover the mechanisms by which these genes functionally interact. To facilitate the impact of our work, we make digenic disease predictions available for all pairs of human

genes.

## PgmNr 2063: Increasing awareness and earlier testing for mucopolysaccharidoses to improve patient outcomes: Updated results from Simply Test for MPS™ Enzyme-Panel Program.

### Authors:

L. Clarke <sup>1</sup>; B. Deignan <sup>2</sup>; C. Raggio <sup>3</sup>; P. Solari <sup>4</sup>; T. Wood <sup>5</sup>; L. Pollard <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) British Columbia Children's Hospital Research Institute, Vancouver, Canada; 2) St. Joseph's Children's Hospital, Tampa, FL, USA; 3) Hospital for Special Surgery, New York, NY, USA; 4) BioMarin Pharmaceutical Inc., Novato, CA, USA; 5) Greenwood Genetics Laboratory, Greenwood, SC, USA

---

**Purpose/Background:** The subtle, non-specific signs/symptoms of mucopolysaccharidoses (MPS) often present a diagnostic challenge. Simply Test for MPS™ (ST4MPS) Enzyme-Panel Testing Program provides a no-cost MPS enzyme panel to facilitate testing for patients in the United States who present with clinical signs/symptoms consistent with MPS. The objective of this analysis is to increase awareness of non-specific signs/symptoms of MPS found through the ST4MPS program to facilitate an accurate diagnosis through earlier testing by a wide range of specialties that may encounter MPS patients.

**Methods:** Initiated in February 2017, ST4MPS provides comprehensive enzyme activity testing for all MPS disorders. Specimens are sent to Greenwood Genetics Laboratory for analysis. Results are reported.

**Results:** As of July 2018, 451 referrals for ST4MPS kits received. Since presentation of the first year program results in April 2018, there was an approximate 80% increase in the average amount of tests ordered (22 tests/month February 2017–April 2018, 40 tests/month May–July 2018) and an increased number of confirmed MPS patients through testing (n=90). Notably, of those with a positive Morquio A diagnosis (n=28), 21 were tested by a geneticist, 4 by a pediatric orthopedist, and 3 by a pediatrician. Average age at diagnosis was 16.4 years (range: 4 months–48.9 years). Of the ordering physicians that indicated signs/symptoms prompting testing (n=23; 5 missing data), all mentioned skeletal/muscular-related signs/symptoms. 46% of positive Morquio A patients were noted to have skeletal dysplasia or another skeletal condition.

**Discussion:** Increased awareness of early signs/symptoms of MPS and utilization of ST4MPS have resulted in accurate diagnoses of MPS. Broader testing by a variety of specialists that are likely to encounter MPS patients would help reduce misdiagnoses and diagnostic delays, leading to improved patient outcomes.

## PgmNr 2064: Integration of polygenic risk score with newborn metabolomics provides increased accuracy in profiling risk of infantile hypertrophic pyloric stenosis.

### Authors:

J. Fadista <sup>1</sup>; L. Skotte <sup>1</sup>; J. Courraud <sup>2</sup>; F. Geller <sup>1</sup>; S. Gørtz <sup>1</sup>; J. Wohlfahrt <sup>1</sup>; M. Melbye <sup>1,3,4</sup>; A.S. Cohen <sup>2</sup>; B. Feenstra <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Danish Center for Neonatal Screening, Clinical Mass Spectrometry, Statens Serum Institut, Copenhagen, Denmark; 3) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 4) Department of Medicine, Stanford University School of Medicine, Stanford, California, USA

---

Infantile hypertrophic pyloric stenosis (IHPS) is a disease, typically appearing between two and eight weeks after birth, caused by hypertrophy of the pyloric sphincter smooth muscle. IHPS perinatal risk factors include the use of macrolide antibiotics, male sex, being first-born, preterm birth, delivery by cesarean section, and bottle-feeding. IHPS familial aggregation of 20-fold increased risk among siblings and genome-wide SNP heritability of 30% also suggest a strong genetic component. We have previously detected an association between IHPS and lipid metabolism. In this study we aimed to (1) investigate associations between IHPS and lipid-centric targeted metabolites in dried blood spots of 267 matched pairs of IHPS cases and controls taken a few days after birth, and (2) build an IHPS risk score based on the associated metabolites and polygenic risk score (PRS) from our previous IHPS GWAS.

Using the Biocrates p400 Kit, we selected 140 metabolites with valid quantification in at least 80% of the samples. Using mixed effects linear regression model with pair as random effect and adjusting for known IHPS confounders, 12 metabolites showed significantly lower levels in cases than in controls. The top two metabolites were phosphatidylcholines PC(38:4) and PC(38:3). In adults, 23% of the variance in levels of PC(38:4) and PC(38:3) and their ratio is explained by genetic variation in *FADS1* gene, which catalyzes desaturation of PC(38:3) to PC(38:4). In our data, genetic variation in *FADS1* was not associated with IHPS, suggesting that other factors influencing PC(38:3) and PC(38:4) levels underlie the association. Plasma levels of PC(38:3) and PC(38:4) are known to be lower in bottle-fed compared to breast-fed infants, and bottle-feeding is a well-known risk factor for IHPS. Also, the fact that our cases had more diagnoses for neonatal difficulty in feeding at breast (ICD-10 code P92.5) than controls (OR=4.5,  $P = 6 \times 10^{-3}$ ), supports the hypothesis that feeding practice could underlie the metabolite associations. We also evaluated IHPS risk in an independent cohort. Individuals in the highest risk quintile of PRS plus PC(38:4) had a IHPS odds-ratio of 11.4, while with PC(38:4) or PRS alone we get 2.5 and 5.3 odds-ratios, respectively.

In conclusion, we have detected lower levels of phosphatidylcholines in IHPS, possibly reflecting different feeding patterns in the first days of life. Adding the levels of PC(38:4) to PRS increased accuracy in IHPS risk profiling.

## PgmNr 2065: Integration of genome-wide association study and protein-protein interaction network data identifies novel candidate genes associated with COPD cachexia.

### Authors:

P. Lakshman Kumar <sup>1</sup>; A.C. Wilson <sup>1,2</sup>; M.H. Cho <sup>3</sup>; E. Wan <sup>3,4</sup>; B.D. Hobbs <sup>3,5</sup>; G.R. Washko <sup>5</sup>; V.E. Ortega <sup>6</sup>; S. Christenson <sup>7</sup>; X. Li <sup>8</sup>; J.M. Wells <sup>1</sup>; S.P. Bhatt <sup>1</sup>; D. Demeo <sup>3</sup>; S. Lutz <sup>9</sup>; E.F.M. Wouters <sup>10</sup>; R. Casaburi <sup>11</sup>; S.I. Rennard <sup>12</sup>; D.A. Lomas <sup>13</sup>; W.W. Labaki <sup>14</sup>; M.K. Han <sup>14</sup>; A. Agusti <sup>15</sup>; R. Tal-Singer <sup>16</sup>; R.P. Bowler <sup>17</sup>; D.A. Meyers <sup>8</sup>; C.P. Hersh <sup>3</sup>; V.J. Thannickal <sup>1</sup>; M. Dransfeld <sup>1</sup>; E.K. Siliverman <sup>3</sup>; M.N. McDonald <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA; 2) Department of Epidemiology, University of Alabama at Birmingham, Birmingham, AL, USA; 3) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; 4) Veterans Affairs Boston Health Care System, Jamaica Plain, MA, USA; 5) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA; 6) Department of Internal Medicine, Section on Pulmonary, Critical Care, Allergy and Immunologic Diseases, Wake Forest School of Medicine, Winston-Salem, NC, USA; 7) Division of Pulmonary, Critical Care, Allergy, & Sleep Medicine, Department of Medicine, University of California San Francisco, San Francisco, CA, USA; 8) Department of Medicine, University of Arizona Health Sciences, Tucson, AZ, USA; 9) Department of Population Medicine, Harvard Medical School, Boston, MA, USA; 10) Centre of expertise for chronic organ failure, Horn, the Netherlands; 11) Rehabilitation Clinical Trials Center, Los Angeles Biomedical Research Institute at Harbor Harbor-UCLA Medical Center, Torrance, CA, USA; 12) Department of Medicine, Nebraska Medical Center, Omaha, Nebraska, USA and BioPharmaceuticals R&D, Astra Zeneca, AstraZeneca, Cambridge, UK; 13) UCL Respiratory, University College London, United Kingdom; 14) Division of Pulmonary and Critical Care Medicine, University of Michigan, Ann Arbor, MI, USA; 15) Fundació Investigació Sanitària Illes Balears (FISIB), Ciber Enfermedades Respiratorias (CIBERES), Barcelona, Catalunya, Spain Thorax Institute, Hospital Clinic, IDIBAPS, Univ. Barcelona, Barcelona, Spain; 16) GSK, Medical Innovation, Value Evidence and Outcomes, Collegeville, Pennsylvania, USA; 17) Department of Medicine, Division of Pulmonary, Critical Care & Sleep Medicine National Jewish Hospital, Denver, CO, USA

---

Chronic Obstructive Pulmonary Disease (COPD) is currently the third leading cause of death globally. COPD patients with cachexia have increased risk of death. As not every patient with COPD develops cachexia, we hypothesized genetic variation may be associated with increased risk. To test this, we performed genome-wide association study (GWAS) analyses among COPD subjects (defined using spirometry) with and without cachexia from 3 cohorts and integrated findings with publicly available protein-protein interaction (PPI) network data.

In ECLIPSE, cachexia was defined as weight-loss or low BMI in addition to 3 out of 5 of low 6-minute walking distance (6MWD), fatigue, anorexia, low muscle mass (FFMI) and abnormal biochemistry. In COPDGene and SPIROMICS, cachexia was defined as weight-loss or low BMI in conjunction with either low 6MWD, low FFMI, fatigue or anemia based on availability of collected information. In each cohort, logistic regression models were run on genotyped and imputed SNPs under an additive model

adjusting for age, sex and population substructure in PLINK in Non-Hispanic White (NHW) and African American (AA) COPD subjects. Gene-based tests were performed using MAGMA. Meta-analyses were performed in METAL and MAGMA. Further downstream analyses integrating genetic findings with PPI network data using dmGWAS.

No single variant was associated at a level of genome-wide significance. At the gene level, several genes were significantly associated with COPD cachexia. These included *TMEM30B* in the NHW COPD participants from SPIROMICS. *TMEM30B* encodes the transmembrane protein 30B and is involved in the hydrolysis of ATP for transport of phospholipids. Two genes (*POR* and *GNAQ*) were significantly associated with cachexia among AA in COPD Gene. *POR* encodes Cytochrome P450 oxidoreductase. *GNAQ* encodes a G Protein Q Polypeptide which was recently associated with body mass index in the UK Biobank. When the gene-based results were integrated with PPI network information, the top consensus network was found to include *SERPINB2* and *TKT*. *SERPINB2* is a protease inhibitor associated with stress response, thrombosis and cell migration. *TKT* encodes transketolase which is a thiamine dependent enzyme that channels sugar phosphotase to glycolysis. Collectively, the consensus network genes were enriched with GO Negative Regulation of Cell Death genes.

In summary, we have identified several new genes including *SERPINB2* and *TKT* associated with COPD cachexia.



## PgmNr 2066: Understanding the molecular pathology of sporadic inclusion body myositis using transcriptome analysis.

### Authors:

M. Johari <sup>1,2</sup>; M. Savarese <sup>1,2</sup>; A. Vihola <sup>1,2,4</sup>; J. Palmio <sup>3</sup>; M. Jokela <sup>3</sup>; P.H. Jonson <sup>1,2</sup>; P. Hackman <sup>1,2</sup>; B. Udd <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Folkhalsan Research Center, Helsinki, Finland; 2) Department of Medical Genetics, Medicum, University of Helsinki, Finland.; 3) Neuromuscular Research Center, Department of Neurology, Tampere University and University Hospital, Tampere, Finland; 4) Neuromuscular Research Center, Department of Genetics, Fimlab Laboratories, Tampere, Finland

---

Sporadic Inclusion Body Myositis (sIBM) is the most common form of acquired adult-onset muscle disorder in the world, with an age of onset after 45 years and affecting around 300 patients in Finland and 2500 in Europe. In the USA, the prevalence ranges from 10-71 per 1 million. However, due to high clinical variability, a number of cases remain largely misdiagnosed. Coexistence of inflammatory and muscle degenerative features result in a complicated and poorly understood molecular pathogenesis of sIBM. Additionally, there are no suitable treatments available with proven efficacy, resulting in an increasing burden of illness on patients and supporting healthcare systems.

Our primary goal is to understand the mechanism behind muscle degeneration in patients with sIBM. We performed whole transcriptome analysis on microdissected biopsies of 41 human skeletal muscle samples, including clinically and histopathologically defined sIBM, a myopathy control with clear genetic and histopathologic evidence and a non-muscle phenotype control. Our current results indicate an sIBM specific heavy inflammatory background with strong IFNG and STAT1 dysregulation and expand our understanding of the involvement of cytokines and chemokines in sIBM. Additionally, we observed differential expression of genes involved in calcium mobility in cells. These results will help us better interpret the cause of protein accumulation, immune dysfunction and consequential progressive muscle degeneration in sIBM.

## PgmNr 2067: Whole exome sequencing identifies a rare functional variant SHARPIN G186R associated with increased risk of late-onset Alzheimer's disease.

### Authors:

Y. Asanomi <sup>1</sup>; D. Shigemizu <sup>1,2,3</sup>; A. Miyashita <sup>4</sup>; R. Mitsumori <sup>1</sup>; T. Mori <sup>1</sup>; N. Hara <sup>4</sup>; K. Ito <sup>5</sup>; S. Niida <sup>6</sup>; T. Ikeuchi <sup>4</sup>; K. Ozaki <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division for Genomic Medicine, Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Japan; 2) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 3) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Department of Molecular Genetics, Brain Research Institute, Niigata University, Niigata, Japan; 5) Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Japan

---

Late-onset Alzheimer's disease (LOAD), the most common form of dementia, is one of common diseases and is caused by contributions of many genetic factors. In spite of the recent large scale genomic approach in mainly Caucasian population, the heritability for LOAD remains mostly unknown. Comprehensive genetic studies in Caucasian have identified several rare risk genetic variants with high odds ratio for LOAD, whereas replication studies in other race, especially in East Asian including Japanese did not show certain association for these variants and LOAD. To comprehend the LOAD pathogenesis, it is important that the identification of rare variants partly contributing to the genetic architecture for LOAD in specific race. Here, we performed whole-exome sequencing (WES) analyses of 202 LOAD individuals without the *APOE* e4 risk allele, a major genetic factor for LOAD susceptibility. We identified seven candidate risk variants thorough step-by-step selection of whole-exome variants. We then conducted a case-control association study in a large Japanese cohort consisting of 4,563 cases and 16,459 controls for these seven variants, and identified a rare nonsynonymous variant, rs572750141 (NM\_030974.3:p.Gly186Arg), in *SHARPIN* associated with increased risk of LOAD (Bonferroni-corrected  $P = 8.05 \times 10^{-5}$ , odds ratio = 6.1). We also implemented *in vitro* functional analyses of the variant to reveal possible functions associated with LOAD risk. The amino acid change in SHARPIN resulted in aberrant cellular localization of the variant protein and attenuated the activation of NF- $\kappa$ B, a central mediator of inflammatory and immune responses, indicating that the identification of a rare functional *SHARPIN* variant as a previously unknown genetic risk factor for LOAD. Modest initial sample size recruited in WES, whereas the results indicate that the effectiveness of our approach to identify rare genetic variants with large effect size, and the pipeline may be applicable to fill in the puzzle for the genetic architecture associated other common diseases.

## PgmNr 2068: Genome-wide linkage analysis of Caribbean Hispanic Puerto Rican families suggests rare risk variants in the *ELAVL2* gene.

### Authors:

G.W. Beecham <sup>1,2</sup>; F. Rajabli <sup>1</sup>; B.E. Feliciano-Astacio <sup>3</sup>; K. Celis <sup>1</sup>; K.L. Hamilton-Nelson <sup>1</sup>; L.D. Adams <sup>1</sup>; V.C. Rodriguez <sup>1</sup>; P.R. Mena <sup>1</sup>; S. Tejada <sup>1</sup>; P.L. Whitehead <sup>1</sup>; N.K. Hofmann <sup>1</sup>; P. Bussies <sup>1</sup>; M. Prough <sup>1</sup>; C.B. Sierra-Lopez <sup>3</sup>; A. Chinae <sup>3</sup>; N.I. Feliciano <sup>4</sup>; H. Acosta <sup>5</sup>; H.N. Cukier <sup>1</sup>; M.L. Cuccaro <sup>1,2</sup>; J.M. Vance <sup>1,2</sup>; M.A. Pericak-Vance <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 3) Universidad Central del Caribe, Bayamón, PR; 4) Hospital De Psiquiatría Estatal Dr. Ramón Fernández Marina-Centro Médico, San Juan, PR; 5) Clínica de la Memoria, San Juan, PR

**Background:** The ancestral genetic heterogeneity (admixture) of Caribbean Hispanics (CHI) makes studies of this population critical to the discovery of ancestry-specific genetic factors in Alzheimer disease (AD). In this study, we performed whole genome sequencing (WGS) in multiplex CHI Puerto Rican (PR) families to identify rare causal variants influencing AD through linkage and segregation-based approaches.

**Methods:** As part of the PR Alzheimer Disease Initiative, WGS data were generated for 95 individuals (69 affected) from 25 Puerto Rican families. To identify the genetic loci likely to carry risk variants, we performed a multipoint linkage scan using MERLIN software. Following the linkage analysis, we identified the consensus regions (HLOD > 2), annotated variants using Ensembl Variant Effect Predictor, and combined annotation dependent depletion score (CADD). Finally, we prioritized the variants in the consensus regions based on allele frequency (< 0.01), function (CADD > 10), and complete segregation among affected individuals.

**Results:** A locus at *9p21* produced a linkage LOD score of 3.9 in the parametric multipoint model supported by 8 families. The region overlaps with a previously reported linkage peak in an independent set of non-Hispanic White families (NHW). We identified 73 rare variants that segregated with the disease in the families that supports the linkage peak on *9p21*. Remarkably, 3 rare variants in the *ELAVL2* gene segregate with disease in two families. One of the interesting variants is rs542037226 (chr9:23,827,245 G/A). It segregates with the disease in a family with 5 AD individuals (family-specific LOD = 1.2). The variant is upstream of *ELAVL2* with a rare allele frequency in the 1kGP reference data set (MAF = 0.001) and a moderate CADD score (17.36).

**Conclusions:** Linkage analysis of CHI PR families confirmed previously reported linkage to *9p21* in NHW families. Our results suggest the presence of AD risk variants in the *ELAVL2* gene. *ELAVL2* encodes a neural-specific RNA-binding protein that plays an essential role in the post-transcriptional regulations of APP, alterations of which are associated with AD. Identified putative damaging rare variants in multiplex families indicates the critical role of rare variation in AD etiology.

## PgmNr 2069: Bivariate genome-wide association scans detect novel pleiotropic loci for Alzheimer's disease and cardiometabolic traits.

### Authors:

W.P. Bone<sup>1</sup>; K.M. Siewert<sup>1</sup>; D. Klarin<sup>2,3,4</sup>; S.M. Damrauer<sup>5,6</sup>; K. Chang<sup>5,7</sup>; P.S. Tsao<sup>8,9</sup>; T.L. Assimes<sup>8,9</sup>; M.D. Ritchie<sup>10,11,12</sup>; B.F. Voight<sup>5,10,11,13,14</sup>; the VA Million Veteran Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Boston VA Healthcare System, Boston, MA, USA; 3) Center for Genomic Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, USA; 6) Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 7) Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 8) VA Palo Alto Health Care System, Palo Alto, CA, USA; 9) Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA; 10) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 11) Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 12) Center for Precision Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 13) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 14) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Epidemiological and genetic studies have consistently found a positive correlation between Alzheimer's disease (AD) and cardiometabolic diseases, yet the biological mechanisms behind this correlation remain elusive. One avenue to investigate the genetic etiology of this relationship is through multi-trait association studies, designed to detect single nucleotide polymorphisms (SNPs) that are associated with AD and cardiometabolic traits. We performed our multi-trait association analyses via a pairwise, bivariate genome-wide association study. Each bivariate scan was performed using summary statistics from the largest single-trait scans for AD (Jensen et al. 2019, involving 71,880 cases and 383,378 controls) paired with one cardiometabolic trait including: coronary heart disease, type 2 diabetes, blood pressure metrics, anthropometric traits, and lipid traits. In total, we performed bivariate analyses between AD and eleven cardiometabolic traits and identified three novel genome-wide significant loci, with each of the three colocating (COLOC PP4>0.85) with an expression quantitative trait locus (eQTL) for a nearby gene. One of these three loci was detected in two bivariate scans: AD with high-density lipoprotein (HDL) levels and diastolic blood pressure (DBP). The lead SNPs at this locus from both scans colocated with an eQTL for the gene *SPPL2A*, with lower expression associated with increased risk for AD and elevated HDL and DBP. In addition, we identified a second signal in the AD and DBP bivariate scan at near-genome-wide significant association in the gene *DOCK4* ( $P=8.34 \times 10^{-8}$ ). This signal implicates a low-frequency, coding variant near a 3' splice junction of *DOCK4* that is predicted to alter splicing *in silico* and warrants further functional validation. This provides a direct hypothesis for the causal variant and genetic mechanism underlying this signal.

These results emphasize the utility of multi-trait GWAS methods for (1) characterizing pleiotropy across complex human traits, (2) generating credible leads for biological and genetic mechanisms for these loci, and (3) predicting the expected phenotypes given a genetic perturbation at a locus.

## PgmNr 2070: Dissecting the local ancestry effect on *APOE* $\epsilon$ 4 and Alzheimer association in an admixed Peruvian population.

### Authors:

M. Cornejo-Olivas <sup>1,7</sup>; F. Rajabli <sup>2</sup>; V. Marca <sup>1</sup>; P. Whitehead <sup>2</sup>; N. Hofmann <sup>2</sup>; O. Ortega <sup>1</sup>; K. Milla-Neyra <sup>1</sup>; D. Veliz-Otani <sup>1</sup>; N. Custodio <sup>4</sup>; R. Montesinos <sup>4</sup>; S. Castro-Suarez <sup>5,6</sup>; M. Meza <sup>5,6</sup>; L. Adams <sup>2</sup>; P. Mena <sup>2</sup>; J. Vance <sup>2,3</sup>; R. Isasi <sup>2,3</sup>; M. Cuccaro <sup>2,3</sup>; G. Beecham <sup>2,3</sup>; P. Mazzetti <sup>1,6</sup>; M. Pericak-Vance <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurogenetics Research Center, Instituto Nacional de Ciencias Neurológicas, Lima, Peru; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA; 3) Dr. John Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA; 4) Instituto Peruano de Neurociencias, Lima, Peru; 5) Centro Básico de investigación en Demencias y enfermedades desmielinizantes del sistema nervioso, Instituto Nacional de Ciencias Neurológicas, Lima, Peru; 6) School of Medicine, Universidad Nacional Mayor de San Marcos, Lima, Peru.; 7) Center for Global Health, Universidad Peruana Cayetano Heredia, Lima, Peru

Alzheimer disease (AD) is the leading cause of dementia in the elderly and occurs in all ethnic and racial groups. Multiple AD-associated loci have been identified, mainly in participants from European (EU) backgrounds. Studies in diverse populations suggest that while there is some overlap in genetic architecture, there are also significant differences between various ethnic groups for genetic risk and protective effects.

The Peruvian population results from admixture of several Amerindian ethnic groups (~80%) and European populations (~15%). Establishing a Peruvian AD cohort will enhance discovery of novel risk/protective factors. To date, we have ascertained 72 unrelated individuals (37 AD cases; 35 cognitively intact controls) with additional ascertainment ongoing. All cases were assessed by trained neurologists following NINDS-ADRDA criteria for probable AD. Cognitively intact controls were screened using the MMSE, the Clock drawing test, and the Pfeffer functional activities questionnaire. Samples were matched for sex and age (75.8(5.5) vs 76.3(6.5), case vs control). Genotyping was performed using the Illumina GSA array; APOE genotyping was also performed. PC-AIR and model-based ADMIXTURE approach were used to infer population structure. To assess local ancestry, phasing was performed using SHAPEIT (with 1kGP reference) followed by RFMix (HGDP reference panels). Association between affection status and gene dose of the APOE  $\epsilon$ 4 allele was analyzed using logistic regression, adjusting for age, gender, PC1, and PC2.

The average genome-wide ancestry proportions in these data are estimated as 0.56, 0.37, 0.05, and 0.02 for Native American (AI), European (EU), West African, and East Asian, respectively. Logistic regression results showed a significant association between APOE  $\epsilon$ 4 allele and AD (OR=5.1, CI:1.6-22.8, p-value < 0.01). By stratifying the dataset according to the local ancestry and APOE genotypes, we observed unexpected deviation in local ancestry proportions. While individuals without the  $\epsilon$ 4 allele have local ancestry proportions (AI:0.66, EU:0.32, AF:0.02) similar to the global ancestry one, individuals with the  $\epsilon$ 4 allele showed unusual variation in ancestry proportions (AI:0.29, EU:0.56, AF:0.15). Results indicate a high frequency of EU ancestry among the  $\epsilon$ 4 risk allele carriers. This suggests that association between the  $\epsilon$ 4 allele and AD is derived through EU genetic ancestry local

to the APOE gene.

## PgmNr 2071: Utilizing CRISPR to investigate an ethnic specific deletion in *ABCA7* associated with Alzheimer's disease in African Americans.

### Authors:

H.N. Cukier<sup>1,2,3</sup>; K. Nuytemans<sup>1,3</sup>; J. Laverde-Paz<sup>1</sup>; T.A. Strong<sup>1</sup>; R.A. Eldabbas<sup>1</sup>; M.A. Pericak-Vance<sup>1,2,3</sup>; D.M. Dykxhoorn<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida.; 2) Department of Neurology, University of Miami, Miami, Florida.; 3) The Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, Florida.

---

Diverse populations are particularly vulnerable to Alzheimer's disease (AD); Hispanics and African Americans (AA) are 1.5 and 2 times as likely to develop AD, respectively, compared to non-Hispanic whites. We identified an ethnic specific deletion in the coding portion of *ABCA7* which significantly increases AD risk among AA ( $p=1.414 \times 10^{-5}$ , OR=1.81 [95% CI:1.38-2.37], Cukier, et al, 2016). The deletion is enriched in AA cases (15.2%) compared to AA controls (9.74%). By generating isogenic cell lines that carry the 44 base pair deletion, we aim to determine the cellular and molecular consequences on *ABCA7* function and how it may contribute to AD risk. CRISPR genome editing was performed in human embryonic kidney (HEK) cells (HEK) to remove the precise 44 base pairs in *ABCA7* in exon 14 (rs142076058) and mimic the deletion. Four guide RNAs (gRNAs) were designed that recognize protospacer adjacent motif (PAM) sequences in the region flanking the deletion through the publically available ZiFiT Targeter software (<http://zifit.partners.org/ZiFiT/>) and CRISPR Design (<http://crispr.mit.edu/>) programs. All four of the gRNAs were cloned into the pCAG-eCas9-GFP-U6-gRNA vector (Addgene), a plasmid that carries the Cas9 enzyme, gRNA and GFP tag all in one. Three of the gRNAs were verified to cut the region near the deletion in the correct manner via the Surveyor Mutation Detection kit (IDT). We transformed one of the gRNAs into HEK cells, performed fluorescence-activated cell sorting (FACS), and screened 119 unique HEK clones. Sanger sequencing confirmed that four HEK clones were generated that were heterozygous for the deletion, as well as two more clones that were homozygous for the deletion. Furthermore, RNA extracted from the HEK clones was found to be expressing *ABCA7* from both the wild type and the allele with the deletion. This demonstrates that these cell lines are an appropriate model to evaluate how the deletion may interfere with *ABCA7* function, such as cellular transport and amyloid beta processing. Utilizing CRISPR genome editing, isogenic HEK cell lines heterozygous and homozygous for the *ABCA7* deletion were created. These cell lines will be used to determine the consequences that result from this ethnic specific deletion.



## PgmNr 2072: Shared mechanisms between type 2 diabetes and Alzheimer's disease via integrative genomics analysis.

### Authors:

J. Ding <sup>1</sup>; Y. Zhao <sup>1</sup>; L. Shu <sup>1</sup>; K. Wei <sup>1</sup>; X. Yang <sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Integrative Biology and Physiology, UCLA, Los Angeles, CA.; 2) Institute for Quantitative and Computational Biosciences, UCLA, Los Angeles, CA.; 3) Molecular Biology Institute, UCLA, Los Angeles, CA.; 4) Brain Research Institute, UCLA, Los Angeles, CA.

---

Epidemiological evidence suggests that Alzheimer's disease (AD) and type 2 diabetes (T2D) are tightly connected, and molecular pathways such as insulin signaling, growth factor pathways, protein misfolding, and inflammation are involved in both diseases. The exact causal molecular connections between AD and T2D, however, remain unclear. We hypothesize T2D may induce perturbations in biological processes that promote AD and/or the two diseases share genetically-driven pathogenic pathways. We conducted a multi-omics systems biology study that integrates human genetic association studies of AD and T2D, gene expression profiling studies of disease-relevant tissues, biological pathways, and tissue-specific gene networks to investigate the potential mechanistic links between the two diseases. Incorporation of genetic association information of both AD and T2D informed on shared tissue-specific gene subnetworks relating to immune regulation, insulin signaling, lysosome, proteostasis, cell cycle, cell adhesion, spliceosome, lipid metabolism, and extracellular matrix. AD-specific pathways were numerous and included oxidative phosphorylation, Golgi-associated vesicle biogenesis, cell-cell communication, Notch signaling, steroid hormones, and infection. For T2D, unique processes included incretin secretion, beta cell development, obesity, glycerolipid metabolism, PPAR signaling, and water homeostasis. Our systems-wide integrative genomics approach offers comprehensive molecular insights into the molecular networks connecting metabolic and neurodegenerative diseases.

## PgmNr 2073: Stratified case-control association mapping of Alzheimer's disease using family history from UK Biobank.

### Authors:

S. Esmaeeli; X. Zheng; J.W. Davis; J.F. Waring

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genomics Research Center, Abbvie Inc, Chicago, Illinois.

---

Understanding the role that genetics has in phenotypic and disease variation, and its potential interactions with other factors, is crucial for a better understanding of human biology. And it is hoped that this will lead to more successful drug development. UKBB (UK Biobank) is a Prospective cohort of 502,620 participants with deep genetic and phenotypic data. We are interested in pathways and regulators of late onset Alzheimer's disease. During the study ascertainment period at enrollment, UKBB recorded only a few hundred cases with Alzheimer disease (AD).

In this study we have conducted a genome-wide association study by proxy (GWAX) using imputed genotypes on SAIGE (Scalable and Accurate Implementation of Generalized mixed model) platform, utilizing the family history of AD from UKBB. We thus defined individuals who are not adopted and reported having at least one parent or sibling with AD as proxy cases, while the proxy controls include subjects 70 years or older who reported no parental history of AD, depression or Parkinson's disease. Also we have excluded from the control set anyone that reporting any neurological or behavioral disorders. We stratified the cases and controls based on their APOE status, by excluding all carriers of at least one *APOE4* allele from either the case or control set. We performed the GWAX analysis on 33,425 cases versus 91,259 controls. We have identified the well-known loci associated with AD, e.g., *BIN1*, *TREM2*, *INPP5D*, *CTNN3* and as well as novel loci. Many of the genes within these loci map to pathways involved in neuronal development. Currently we are validating these novel findings in genetic studies from diagnosed AD patients, such as ADNI. This study furthers substantiates the use of GWAX analysis for identifying underlying genetics of disease. These results could further serve as a valuable resource for selection of promising genes for functional follow-up experiments and may identify targets for drug development and stratification approaches.

## PgmNr 2074: A rare variant in *SYT7* and common variants in *APOE* and *LRP2* are associated with urate level in exomes of 150,000 subjects from the UK Biobank.

### Authors:

J. Freudenberg; K. Praveen; K. Devalaraja; M. Ferreira; L. Gurski; A. Li; J. Backman; C. O'Dushlaine; D. Sun; A. Marcketta; J. Horowitz; E. Maxwell; L. Barnard; L. Habegger; Y. Hahn; A. Lopez; N. Banerjee; A. Yadav; S. Khalid; S. Balasubramanian; W. Salerno; P. Tamargo; M. Jones; A. Shuldiner; J. Overton; J. Reid; G. Abecasis; M. Cantor; A. Baras; G. Coppola

[View Session](#) [Add to Schedule](#)

**Affiliation:** Regeneron Genetics Center, Tarrytown, New York.

---

Elevated uric acid (UA) is the cause of gout, a common inflammatory arthropathy. In turn, UA is also hypothesized to exert a protective role in neurodegeneration, possibly due to its anti-oxidant properties. To better understand the genetic determinants of UA levels, we studied 150,000 individuals from the UK Biobank (UKB) with whole-exome sequencing data. We find 21 protein-altering index variants associated with UA at genome-wide significance ( $p < 5 \times 10^{-8}$ ). Among the three rare ( $MAF < 5\%$ ) index variants, the most significant one (rs747186203) is located in exon 4 of *SYT7*, changing Arg to Gly at position 132 of the protein ( $p = 5.50 \times 10^{-17}$ ,  $\beta = -0.71$ ,  $95\%CI = [-0.88, -0.55]$ ,  $MAF = 0.0005$ ). *SYT7* encodes synaptotagmin 7, a calcium sensor involved in the exocytosis of vesicles. Among common variants we interestingly find an association of UA with *APOE*. More specifically, we find an UA lowering effect for *APOE*-4 (rs429358,  $p = 4.5 \times 10^{-10}$ ,  $\beta = -0.032$ ,  $95\%CI = [-0.042, -0.022]$ ) and a weak UA increasing effect for *APOE*-2 (rs7412,  $p = 2.3 \times 10^{-2}$ ,  $\beta = 0.015$ ,  $95\%CI = [0.0021, 0.029]$ ). Because *APOE*-4 is the major risk factor for Alzheimer's disease, this association, and its molecular mechanism, may influence the correlation of UA with neurodegeneration. We further find UA being associated with a common missense variant in *LRP2* (rs2075252, Lys4094Glu,  $p = 1.9 \times 10^{-14}$ ,  $\beta = -0.033$ ,  $95\%CI = [-0.041, -0.024]$ ,  $MAF = 0.24$ ). *LRP2* (LDL receptor related protein 2) functions in reuptake of a range of ligands in the kidney, and is implicated in different forms of kidney disease. *LRP2* also is hypothesized to function in *APOE*-mediated uptake of cholesterol into neurons, as well as APOJ-mediated transport of amyloid-beta through the blood brain barrier. These studies contribute to understanding the relationship between UA levels, kidney function, and neurodegeneration, and may lead to novel therapeutic principles.

## PgmNr 2075: Using linkage analysis to identify novel gene-gene interactions in Alzheimer's disease.

### Authors:

M. Grunin <sup>1,2</sup>; N.R. Wheeler <sup>1</sup>; W.S. Bush <sup>1,2</sup>; J.L. Haines <sup>1,2</sup>; Alzheimer's Disease Sequencing Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 2) Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH

---

Linkage analysis was a workhorse of disease gene mapping until the advent of more refined genomic scans through genome wide association studies (GWAS). However, GWAS' ability to detect gene-gene interactions is hampered by the required huge sample sizes. Whole Genome Sequencing (WGS) captures nearly all allelic variation enabling direct measurement of co-segregation with disease for each allele in a family through linkage analysis without the need to consider recombination fractions. We developed an approach integrating co-segregation data with extant biological interaction data to identify interacting genes involved in disease using WGS of multiplex Alzheimer's Disease (AD) families. Two point linkage (MERLIN) was performed on WGS of 46 AD Sequencing Project (ADSP) families. Per family, an empirical maximum LOD score (MaxLOD) was calculated. Interactions between proteins were identified by querying multiple functional/pathway databases. Variants reaching a family-specific MaxLOD were filtered by high/moderate impact (missense+stop/loss) and functional interaction. Genes of interest were defined as genes of known interaction where at least 2 biologically interacting loci per family reached a family-specific MaxLOD and were found in >1 family, creating a single shared transchromosomal locus. The transchromosomal MaxLODs were combined across multiple families sharing the same biologically interacting loci. Multiple families had biologically interacting genes with individual variants attaining a MaxLOD>1.2. One interaction of shared gene pairs/family had a transchromosomal MaxLOD across 3 families of 4.05, linking variants in SLC10A6, DHCR7, SOAT1, ABCA1 and SREBF2 in the cholesterol metabolism and lipid processing pathways. An additional 5 families had MaxLOD variants in at least one of these genes. 4 families had interacting pairs with a transchromosomal MaxLOD score >2.4 unique to the family. This approach utilizes biological interactions to identify potential disease loci despite genomic distance, thus identifying possible epistatic effects not easily observable in case-control datasets. It is easily applicable to any complex disorder having multiplex families.

## PgmNr 2076: Genetic associations with cognitive performance inform Alzheimer's disease etiology.

**Authors:**

L. Gurski; K. Praveen; G. Coppola; G. Abecasis; M. Ferreira; Regeneron Genetics Center

[View Session](#) [Add to Schedule](#)

**Affiliation:** Regeneron Genetic Center, Tarrytown, New York.

---

Alzheimer's disease (AD) is a leading cause of dementia and affects over 40 million people worldwide. As AD typically develops late in life, it is challenging to perform genetic association studies with very large numbers of affected individuals. To help identify genetic risk factors for AD we analyzed a quantitative phenotype that is genetically correlated with AD case-control status ( $r_g=0.23$ ,  $P=6 \times 10^{-7}$ ) and was measured in participants of the UK Biobank study: time to complete the touchscreen questionnaire (TCQ). We analyzed 9 million common variants ( $MAF > 1\%$ ) available in 462,958 individuals, and 3.8 million rare variants ( $MAF < 1\%$ , including gene burden tests) obtained from exome sequencing of a subset of 142,360 individuals. We identified 222 variants independently associated with TCQ ( $p < 5 \times 10^{-8}$ ), 29 of which also showed at least a nominally significant association ( $p < 0.05$ ) with AD risk. Sixteen variants were in high linkage disequilibrium with a sentinel expression quantitative trait locus for a nearby gene. Overall, our results suggest that the study of phenotypes related to cognitive performance can help identify novel genetic risk factors for AD.

## PgmNr 2077: Genome-wide analyses identify novel sex-specific candidate loci for Alzheimer disease.

### Authors:

B.W. Kunkle <sup>1</sup>; K.L. Hamilton-Nelson <sup>1</sup>; L. Gomez <sup>1</sup>; M. Jean-Francois <sup>1</sup>; A.C. Naj <sup>2</sup>; L. Wang <sup>1</sup>; A.B. Kuzma <sup>2</sup>; G.W. Beecham <sup>1</sup>; W.S. Bush <sup>3</sup>; T. Hohman <sup>4</sup>; B.N. Vardarajan <sup>5</sup>; L.S. Wang <sup>2</sup>; R. Mayeux <sup>4</sup>; J.L. Haines <sup>5</sup>; L.A. Farrer <sup>6</sup>; G.D. Schellenberg <sup>2</sup>; M.A. Pericak-Vance <sup>1</sup>; E.R. Martin <sup>1</sup>; Alzheimer's Disease Genetics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Miami, Miller School of Medicine, Miami, FL, USA; 2) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 3) Case Western University School of Medicine, Cleveland, OH, USA; 4) Vanderbilt University Medical Center, Nashville, TN, USA; 5) Columbia University Medical Center, New York, NY, USA; 6) Boston University School of Medicine, Boston, MA, USA

**Background:** Women comprise nearly two-thirds of all Alzheimer disease (AD) cases, suggesting sex-specific risk and protective factors. For example, a number of studies have established that apolipoprotein E (*APOE*) genotype contributes to risk of AD differently in men and women. To identify additional sex-specific genetic associations with AD, we meta-analyzed imputed genome-wide datasets using two complementary statistical analyses.

**Methods:** Sex-interaction and sex-stratified analyses were performed in the Alzheimer's Disease Genetics Consortium (ADGC) haplotype reference consortium (HRC) -imputed dataset (N=9,135 cases, 60% female; 9,677 controls, 60% female). Study-specific logistic regression analyses were conducted on 21 studies using Plink (sex-interaction) and SNPTest (sex-stratified). Results were meta-analyzed using inverse variance based analyses in METAL.

**Results:** We found evidence of sex-interaction effects (interaction  $P < 10^{-5}$ ) for 131 variants and 15 loci, including strong associations with opposing effect directions for intronic *BBS9* (interaction- $P = 2.51 \times 10^{-7}$ ) and *LIP1* variants (interaction- $P = 4.03 \times 10^{-7}$ ). Sex-stratified analyses identified 86 variants and 34 loci significant in one sex ( $P < 10^{-5}$ ) and not the other ( $P > 0.05$ ). Sixteen variants in the *APOE* region, the strongest AD risk locus, were associated with AD in females (top  $P = 8.95 \times 10^{-10}$ ) but not in males ( $P = 0.19$ ). Two variants in another known AD locus, *BIN1*, showed association in females (top  $P = 7.89 \times 10^{-7}$ ) but not in males ( $P = 0.08$ ). Additional associated variants in females include intronic variants in *ATE1* (female- $P = 1.19 \times 10^{-7}$ ; male- $P = 0.60$ ), *HYDIN* (female- $P = 3.32 \times 10^{-7}$ ; male- $P = 0.28$ ), and *ECE1* (female- $P = 1.37 \times 10^{-6}$ ; male- $P = 0.41$ ); and an intergenic variant near *SORD* (female- $P = 9.12 \times 10^{-7}$ ; male- $P = 0.23$ ). Male-specific variants include intronic variants in *SMOX* (male- $P = 1.43 \times 10^{-7}$ ; female- $P = 0.48$ ), *PLD1* (male- $P = 4.37 \times 10^{-7}$ ; female- $P = 0.45$ ), and *RAB10* (male- $P = 1.89 \times 10^{-6}$ ; female- $P = 0.55$ ).

**Conclusions:** We identified several sex-specific associations with AD in genes with potentially AD-relevant function including *BBS9* (hypertension, heart disease and obesity phenotypes), *LIP1* (lipoprotein metabolism), *ATE1* ( $\beta$ ?amyloid,  $\alpha$ ?synuclein, and TDP43 processing), *PLD1* (enzyme elevated in AD hippocampus), and *RAB10* (AD resilience locus). Understanding the nature of these

associations will help explain differential risk and progression for AD between the sexes.

## PgmNr 2078: Analysis and validation of tissue-specific genetic regulated expression in late-onset Alzheimer's disease in ~160,000 individuals identifies novel genes and function-based fine mapping of known regions.

### Authors:

B. Lee<sup>1</sup>; H.H. Chen<sup>1</sup>; . ADGC<sup>2</sup>; Y. Zhao<sup>3</sup>; S. van der Lee<sup>4</sup>; K.L. Hamilton-Nelson<sup>5</sup>; L. Petty<sup>1</sup>; B. Kunkle<sup>5</sup>; A. Kuzma<sup>3</sup>; O. Valladares<sup>3</sup>; C. Reitz<sup>6</sup>; G.W. Beecham<sup>5</sup>; E.R. Martin<sup>5</sup>; L.S. Wang<sup>3</sup>; J.L. Haines<sup>7</sup>; R. Mayeux<sup>6</sup>; L.A. Farrer<sup>8,9</sup>; M.A. Pericak-Vance<sup>5,10</sup>; G.D. Schellenberg<sup>3</sup>; W.S. Bush<sup>7</sup>; A.C. Naj<sup>3,11</sup>; J.E. Below<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, Tennessee.; 2) the Alzheimer's Disease Genetics Consortium; 3) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 4) Alzheimer Center, Vanderbilt University Medical Center, Amsterdam, the Netherlands; 5) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 6) The Taub Institute for Research on Alzheimer's Disease and the Aging Brain and The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA; 7) Department of Population & Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 8) Department of Biostatistics, School of Public Health, Boston University, Boston, MA, USA; 9) Department of Neurology, School of Medicine, Boston University, Boston, MA, USA; 10) Dr. John T. Macdonald Foundation, Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 11) Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

---

Late-Onset Alzheimer Disease (LOAD) is the most common neurodegenerative disease, occurring in >35% of individuals age 85 years and older, and is the 6<sup>th</sup> most common cause of death in the US. Heritability of LOAD is estimated to be 40-80%, yet the numerous genome-wide studies (GWAS) conducted to date have identified risk variants that explain only 8% phenotypic variance. Many genetic mechanisms underlying LOAD remain unclear, and identifying functional genetic risk factors will likely illuminate the biological processes underlying the disease. To improve understanding of the functional genetic etiology of LOAD we used 25 GWAS datasets including 13,042 cases and 13,674 controls from the Alzheimer's Disease Genetics Consortium. These data were imputed to ~40M variants with the HRC r1.1 reference panel using Minimac3, and high quality variants with MAF>0.05 and imputation  $R^2>0.8$  were used to predict gene expression in 45 tissues using PrediXcan, an approach that infers the genetically regulated portion of gene expression from common expression quantitative trait loci identified in the Genotype-Tissue Expression Project, along with additional brain and immune cell studies. Meta-analysis across all studies was performed using METAL and identified 47 unique genes reaching Bonferroni-corrected genome-wide significance ( $p<2.5\times10^{-6}$ ) across all tissues. Top hits included APOC1 ( $p=2.3\times10^{-61}$ ), GEMIN7 ( $p=1.0\times10^{-32}$ ), and AMMECR1L ( $p=2.1\times10^{-19}$ ). In a cross-tissue model, 3 genes had significant effect, including *MRE11A* ( $p=1.1\times10^{-11}$ ), *CDHR5* ( $p=2.9\times10^{-11}$ ), and *GATAD2B* ( $p=2.7\times10^{-7}$ ). We further adjusted models using nearby previously-reported sentinel risk variants for AD. Across all tissues, 3 unique genes reached genome-wide



significance after correction for known signals: *COLCA1*, *GABPB1*, and *P2RY14*. Replication analyses and additional fine mapping were performed by MetaXcan analysis of summary statistics from logistic regression in an additional independent 11424 cases and 26277 controls. Further independent validation of identified genes was performed in 95,000 individuals in a large, electronic health record-linked DNA databank. Genes identified in our discovery analysis were enriched for associations with cerebral degeneration ( $p=1 \times 10^{-5}$ ) and memory loss ( $p=3 \times 10^{-5}$ ). In conclusion, leveraging predicted expression in ~160,000 individuals illustrates the utility of gene regulatory models in the discovery of new loci and function-based fine mapping of known regions for LOAD.

## PgmNr 2079: Next generation sequencing in dementia subgroup in Thailand.

### Authors:

P. Lertwilaiwittaya<sup>1</sup>; W. Jinda<sup>2</sup>; V. Senanarong<sup>1</sup>; J. Assanasen<sup>1</sup>; W. Muangpaisal<sup>3</sup>; C. Rattanabannakit<sup>1</sup>; C. Limwongse<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine Department, Faculty of Medicine, Siriraj hospital, Mahidol University, Bangkok, Thailand;  
 2) Research Department, Faculty of Medicine, Siriraj hospital, Mahidol University, Bangkok, Thailand;  
 3) Preventive and Social Medicine Department, Faculty of Medicine, Siriraj hospital, Mahidol University, Bangkok, Thailand

---

Dementia is a neurodegenerative disease with high genetic predisposition. Prior studies revealed causative and susceptible genes for both early onset (EOD) and late onset dementia (LOD). Stepwise DNA testing has been traditionally employed. This pilot study is aimed to investigate the usefulness of next generation sequencing (NGS) for dementia testing in Thailand in aspect of the diagnostic yield, correlation with clinical diagnosis, and concern in variants interpretation. Genes included in the NGS panel are those responsible for Alzheimer's disease (AD) (*APP ABCA7 ITM2B PSEN1 PSEN2 SORL1 TOMM40 TREM2*), Frontotemporal Dementia (FTD) (*ALS2 ANG C9orf72 CHMP2B DCTN1 FUS GRN MAPT MATR3 OPTN PFN1 SETX SOD1 SPG11 SQSTM1 TARDBP TBK1 UBQLN2 VAPB VCP*), Dementia with Lewy bodies (DLB) (*SNCA SNCB*) and other dementia conditions (*CSF1R NOTCH3 SERPINI1 TFG TYROBP UBE3A VPS35*). NGS was performed on an Ion S5XL sequencer. Variant annotation and filtering were done using Ion Reporter™, wANNOVAR server, and Variant Effect Predictor. Selected variants with minimum read depth of 30x and minor allele frequency lower than 0.01 were interpreted with ACMG2015 guideline. Samples were from 45 *PSEN1* and *PSEN2* negative AD, 10 FTD, 3 DLB, and 5 unclassified dementia patients diagnosed by neurologists or geriatricians. These were also classified into EOD or LOD, and sporadic or familial (having at least one 1st degree relative with dementia) groups. In AD group, among 103 identified variants, 6 pathogenic/likely pathogenic (p/lp) variants and 28 variants of uncertain significance (VUS) were found (34/103 = 33%). Detection rate of p/lp variant were 13% and 21% in sporadic and familial early onset AD. For FTD group, 28 variants were found included 2 p/lp variants and 3 VUS (5/28 = 18%). Detection rate of p/lp variant were 25% and 33% in sporadic and familial early onset FTD. No p/lp variants were found in DLB group. In unclassified dementia, 13 variants were found included 1 p/lp variant and 6 VUS (7/13 = 54%). Detection rate of p/lp variant was 20% in unclassified dementia. Surprisingly, 2 out of 6 p/lp variants in AD group were indeed causative for FTD and DLB. Overall detection rate of p/lp variant or VUS in Thai *PSEN1* and *PSEN2* negative dementia patient was 68% (43/63 probands). The magnitude of clinical and molecular diagnostic discrepancy should be further explored. This study demonstrated usefulness of mutation screening with NGS over traditional stepwise testing in EOD patient.

## PgmNr 2080: Genome wide meta- and mega-analysis for identifying genes modifying age-at-onset of Alzheimer disease in the ADGC cohorts.

### Authors:

Y.J. Li <sup>1,2</sup>; S.H. Slifer <sup>3,4</sup>; S. Sun <sup>1</sup>; M.A. Schmidt <sup>3,4</sup>; A.C. Naj <sup>5</sup>; X.R. Gao <sup>6</sup>; E.R. Martin <sup>3,4</sup>; Alzheimer Disease Genetics Consortium (ADGC)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Duke Molecular Physiology Institute, Duke Univ Med Ctr, Durham, NC; 2) Department of Biostatistics and Bioinformatics, Duke Univ Med Ctr, Durham, NC; 3) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 4) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 5) Department of Biostatistics, Epidemiology, and Informatics; University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 6) Departments of Ophthalmology and Visual Science and Biomedical Informatics, Division of Human Genetics, The Ohio State University, Columbus, OH

---

Genetic predispositions of age-at-onset (AAO) of Alzheimer disease (AD) was first demonstrated by linkage studies (Daw et al. 1999, Li et al. 2002). In the era of genome wide association study (GWAS), the Alzheimer Disease Genetics Consortium (ADGC) has previously conducted a meta-analysis of 14 cohorts for AAO of AD (Naj et al. 2014), and reported several variants in known risk genes as well as novel AAO variants. Here we re-examine AAO in the ADGC dataset with additional cohorts and improved genome-wide imputation data by both meta- and mega-analysis. A total 6642 unrelated AD cases with AAO data from 20 ADGC cohorts were investigated. Imputation was performed for each dataset based on the Haplotype Reference Consortium reference panel through the Michigan Imputation Server. Consistent quality control procedures were applied to each dataset. Variants with allele counts < 10 in each dataset or overall MAF < 0.01 were excluded. We tested allelic dosage association with AAO by multivariable linear regression models. For meta-analysis, we performed association tests for each dataset with covariate adjustment of sex and three principal components (PCs), where PCs were computed for each dataset. Meta-analysis was followed using the METAL program. For mega-analysis, we used QCTOOL to merge genotype data of 20 cohorts. PCs were generated from the merged data. Covariate adjustment includes sex, 4 PCs, and cohort. Variants meeting  $p < 10^{-5}$  in either methods were examined. The average AAO ranged from 72.2 to 86.1 years among AD cases in the 20 datasets. Meta-analysis revealed 11 regions (meta-regions) associated with AAO with  $p < 10^{-5}$ , but only the chromosome 19 APOE region reached genome wide significant ( $p < 5 \times 10^{-8}$ ). On the other hand, mega-analysis identified 14 regions (mega-regions) with 10 regions reaching  $p < 5 \times 10^{-8}$ . APOE and chr1p36.31 intra-genic regions were identified by both methods with. In addition, 7 meta-regions still maintain nominal significance in mega-analysis except with smaller effect sizes. However, eight mega-regions were not detected by meta-analysis ( $p > 0.05$ ). While more detailed investigation is underway, our results imply that meta-analysis are likely influenced by within-cohort sample size and data normality, which may explain some of the discrepancy. When feasible, mega-analysis would be preferable. By using both approaches, we identified several genetic loci for AAO of AD.

## PgmNr 2081: The genetic and phenotypic overlap between human herpesvirus and Alzheimer's disease risk in the UK Biobank.

### Authors:

J. Liu; T. Vosburg; G. Zheng; K. Li; W. Chen

[View Session](#) [Add to Schedule](#)

**Affiliation:** Translational Biology, Biogen, Cambridge, Massachusetts.

---

Human herpesvirus (HHV) is a family of nine viruses known to cause a range of common diseases including mononucleosis, chickenpox, and shingles. It has been suggested that HHV infection in the brain may increase the risk of developing late onset Alzheimer's disease (AD). While these studies highlight the involvement of viral activity in the brain and amyloid beta, there is a lack of understanding on the role of genetics in susceptibility to viral infection and AD.

Using data from the UK Biobank, we sought to examine whether there is an association between HHV-related phenotypes and AD risk, and the extent to which genetic factors explain such associations. After constructing a general HHV phenotype using data from medications, ICD9, ICD10, and self-reported categories, we assessed the phenotypic overlap between individuals with HHV and AD, as well as AD family history. We performed a GWAS using BOLT-LMM on a general HHV cohort (10,559 cases, 397,606 controls). We used DEPICT and DAPPLE to identify pathways and PPI networks between our HHV and an AD risk GWAS. Finally, we generated polygenic risk scores (PRS) from the general HHV GWAS summary statistics and tested whether the PRS was as associated with AD/AD family history.

We found that individuals with a general HHV phenotype are 2.6 times more likely to have been diagnosed with AD ( $P=5.95e-26$ ) and 1.4 times more likely to have a family history of AD ( $P=7.08e-25$ ) than those without an HHV phenotype. We identified one genome-wide significant locus for the general HHV GWAS, rs3903160 ( $P=6.2e-9$ , OR=1.09), which falls within the HLA region on chromosome 6. Pathway and PPI analysis identified common gene sets that are nominally enriched in HHV and AD GWAS, including immune-related, lipid binding, apoptotic and cell-cell adhesion, as well as pathways previously reported in AD, such as INPPL1 and PLCG2 protein-protein interaction subnetworks. Finally, the HHV susceptibility PRS was significantly associated with AD ( $P=2.94e-4$ ) and AD family history ( $P=8.92e-8$ ), suggesting some shared genetic basis between HHV and AD risk. While the HHV phenotype is not the most well-defined and is difficult to interpret, our findings are consistent with other literature on HHV infection and AD and suggest a genetic overlap between the two. Future studies with better defined HHV phenotypes could further elucidate this relationship, and could potentially fuel drug development decisions and resulting therapies for AD.

## PgmNr 2082: Genome wide association study for dementias in a Japanese population.

### Authors:

R. Mitsumori <sup>1</sup>; D. Shigemizu <sup>1,2,3</sup>; S. Akiyama <sup>1</sup>; Y. Asanomi <sup>1</sup>; T. Mori <sup>1</sup>; S. Niida <sup>4</sup>; K. Ozaki <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division for Genomic Medicine, Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Aichi, Japan; 2) Department of Medical Science Mathematics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; 3) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Medical Genome Center, NCGG, Obu, Japan; 5) Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

---

Dementia is a brain disorder caused by the aging which obstructs social life and interpersonal relationships due to cognitive dysfunction such as memory impairment, consciousness impairment, and slowed ability to judgment. In Japan, about 30% of elderly people aged 65 over were reported to have some types of dementia or mild cognitive impairment (MCI), and the number of affected people is expected to increase as the population aging. The most common form of dementia is late-onset Alzheimer's disease (LOAD), whereas vascular dementia (VaD), dementia with Lewy bodies (DLB), frontotemporal dementia (FTD) and normal pressure hydrocephalus (NPH) are also known as certain types of dementia. The pathogenesis of dementias is affected by environmental and genetic factors and is often misdiagnosed among dementias because of similar symptoms. Dementias are highly heritable, especially in LOAD (58~79%), therefore identification of the genetic factors for dementias contributes to elucidate the pathogenesis and eventually lead to developing the novel diagnosis, prevention, and medicine. Recently, multiple genetic loci associated with dementia have identified in the Genome Wide Association Study (GWAS) in Caucasian subjects. Here, we conducted an ethnicity specific GWASs for LOAD, DLB, VaD, FTD and NPH to clarify the genetic architecture for dementias using an ethnicity specific whole genome genotyping array, Japonica (Affymetrix), which constructed based on whole-genome sequences of 1,070 Japanese individuals, with approximately 7,100 Japanese population recruited at National Center for Geriatrics and Gerontology (NCGG) biobank. We replicated the *APOE* locus (rs429358,  $P$ -value  $\leq 5.0 \times 10^{-8}$ ) for LOAD and DLB and found convincing associations for 4 known LOAD loci, *CLU*, *PICALM*, *ABCA7*, *SORL1*. We further identified new candidate risk loci for these types of dementia. We have also investigated the differences in the susceptible loci identified in the GWASs among dementias.

## PgmNr 2083: Utilizing Massively Parallel Reporter Assays to identify protective functional variants in the *ApoE* region.

### Authors:

K. Nuytemans<sup>1,2</sup>; D.J. Van Booven<sup>1</sup>; N.K. Hofmann<sup>1</sup>; F. Rajabli<sup>1</sup>; A.J. Griswold<sup>1,2</sup>; C.D. Brown<sup>3</sup>; M.A. Pericak-Vance<sup>1,2</sup>; J.M. Vance<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, Miami, Florida.; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, Florida; 3) University of Pennsylvania, Philadelphia, Pennsylvania

---

The risk for late-onset Alzheimer disease (AD) in *ApoEε4* carriers differs between ethnic groups with risk decreasing from Japanese over Europeans to Africans. Ancestry analyses in *ApoEε4* carriers of African-American and Puerto Rican admixed populations showed significant interaction of risk with local ancestry (LA); i.e. ancestry of genomic region surrounding *ApoE* (Rajabli et al, 2019). Reduced risk was observed for African LA *ApoEε4* carriers versus European LA *ApoEε4* carriers. Identification of the protective variant on the African LA haplotype will lead to greater insights into the mechanism with which *ApoEε4* leads to AD.

1000 Genomes variants with significant frequency difference between African and Japanese or European LA haplotypes in the topologically associated domain surrounding *ApoE* (chr19:45375500-45431600) were selected for functional analysis. All these variants are noncoding. Massively parallel reporter assays (MPRA) were used to assess, first, the regulatory potential of these variants and their surrounding regions to present with enhancer activity and, second, the effect size of these variants on enhancer activity, in immortalized SHY-SY5Y neuronal cells, U-118 microglial and HMC3 astrocytes.

We selected variants with significant frequency differences ( $p < 1 \times 10^{-5}$ ) between Japanese or European and African *ApoE* LA ancestry for MPRA. We generated ~900bp PCR fragments surrounding these variants, representing candidate regulatory elements (CRE), in heterozygous carriers to ensure equal representation of alleles. We pooled the amplicons and cloned them into a library of enhancer reporter vectors (pGL4.24) tagged by over 15M different barcodes in the 3'UTR. Subassembly linking APOE fragments to unique barcodes has been completed and has identified >250k unique barcode:fragment links. The vector library was transfected in all three cell lines in duplicate. Analyses of differential expression driven by opposite alleles at the positions of interest are ongoing.

## PgmNr 2084: Integrative genome and transcriptome analysis reveals link between complement pathway and APOE e2 protective mechanism for Alzheimer's disease.

### Authors:

R. Panitch <sup>1,2</sup>; C. Zhu <sup>2</sup>; J. Hu <sup>2</sup>; J. Chung <sup>2</sup>; L.A. Farrer <sup>2</sup>; G.R. Jun <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics Graduate Program, Boston University; 2) Biomedical Genetics Section, Department of Medicine, Boston University School of Medicine

---

The  $\epsilon 4$  allele of apolipoprotein E (*APOE*) is the major genetic risk factor for late onset Alzheimer disease (AD), while the  $\epsilon 2$  allele is known to mitigate disease risk. Mechanisms underlying the protective effect of  $\epsilon 2$  are largely unknown. To investigate the genomic basis of the influence of  $\epsilon 2$  on AD risk, we evaluated gene expression data obtained from the dorsolateral *prefrontal cortex* area of 545 human autopsied brains (including 32 AD and 39 controls with the e23 genotype) that were donated to the Religious Orders Study and Memory and Aging Project (ROSMAP). Expression data obtained from 153 autopsied brains (including 4 AD and 12 controls with the e23 genotype) from the Mayo Clinic Study of Aging (MAYO) were used for replication. Expression levels for each gene were compared between cases and controls within *APOE* genotype groups in the ROSMAP sample. Results obtained from the ROSMAP for significantly differentially expressed genes (DEGs) were combined with results from MAYO by meta analysis. We also constructed co-expressed gene-networks separately in cases and controls for each *APOE* genotype subgroup using Weighted Co-expression Network Analysis (WGCNA). Preserved networks in both datasets ( $Z_{summary} > 10$ ) were further validated using integrative gene enrichment analysis that included the top-ranked DEGs and genes previously implicated in AD by GWAS. Among *APOE* e23 subjects, significant DEGs ( $P < 10^{-5}$ ) were identified including *GFAP* ( $Z_{score} = 4.55$ ,  $P = 6 \times 10^{-6}$ ) and *C4A* ( $Z_{score} = 4.45$ , and  $P = 9 \times 10^{-6}$ ). These DEGs were transcriptome-wide significant in all subjects combined ( $P < 5 \times 10^{-8}$ ). Other complement pathway genes, *C6* ( $Z_{score} = 2.90$ ,  $P = 0.004$ ) and *CR1* ( $Z_{score} = 2.61$ ,  $P = 0.009$ ), were nominally significant DEGs in the e23 subgroup. A network module containing these and AD risk genes was significantly enriched ( $P = 5 \times 10^{-4}$ ). Our findings demonstrate a biological link between the *APOE*  $\epsilon 2$  protective mechanism(s) and complement pathway in AD. Together with a recent report of protein interaction between *APOE* and C1q, our study also suggests that further research aimed at understanding how complement pathway genes interact with *APOE*  $\epsilon 2$  may lead to a potential novel therapeutic target for AD.



## PgmNr 2085: Blood-based transcriptional risk scores predict neuroimaging biomarkers for Alzheimer's disease.

### Authors:

Y.H. Park <sup>1,2</sup>; A. Hodges <sup>3</sup>; A. Simmons <sup>3</sup>; S. Lovestone <sup>4</sup>; M.W. Weiner <sup>5</sup>; A.J. Saykin <sup>1,6</sup>; K. Nho <sup>1,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN; 2) Department of Neurology, Seoul National University Bundang Hospital, Seongnam, Korea; 3) Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; 4) Department of Psychiatry, University of Oxford, Oxford, UK; 5) Departments of Radiology, Medicine, and Psychiatry, University of California-San Francisco, San Francisco, CA; 6) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 7) Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN

Dysregulated gene expression associated with genetic loci influencing Alzheimer's disease (AD) risk could possibly reflect the pathophysiological status of AD. Here we calculated blood-based transcriptional risk scores (TRS) by integrating genome-wide association study (GWAS) summary statistics and expression quantitative trait locus (eQTL) data to predict MRI-based biomarkers for AD. First, we selected 188 independent SNPs with  $p < 1 \times 10^{-5}$  from a large-scale GWAS (Jansen *et al.* 2019) and then identified 31 candidate genes significantly associated with these SNPs in blood-based eQTL database (Westra *et al.* 2013) (FDR-corrected  $p < 0.05$ ). To identify pathogenic genes among the 31 genes, we performed GWAS summary-based Mendelian randomization (SMR) ( $p < 8.4 \times 10^{-6}$ ) and a Bayesian test for colocalization ( $H_4 > 80\%$ ) (Marigorta *et al.* 2017) using summary-level GWAS and eQTL data, which yielded the identification of 3 pathogenic genes (*CD33*, *CTSH*, *PILRA*). Finally, we calculated TRS by summing z scores of the 3 pathogenic genes' expression levels using peripheral-blood microarray gene expression data from two independent cohorts (661 and 674 participants from the Alzheimer's Disease Neuroimaging Initiative (ADNI) and the AddNeuroMed cohorts as discovery and replication samples, respectively). We evaluated whether TRS was associated with diagnosis (AD vs. cognitively normal control) and MRI-based biomarkers (hippocampal volume and entorhinal cortical thickness). In ADNI, TRS was significantly associated with diagnosis ( $p = 1.69 \times 10^{-3}$ ; Odds ratio (OR) = 1.21), hippocampal volume ( $p = 2.11 \times 10^{-3}$ ;  $t$  value = -3.09), and entorhinal cortical thickness ( $p = 2.47 \times 10^{-3}$ ;  $t$  value = -3.04). In AddNeuroMed, TRS was significantly associated with diagnosis ( $p = 1.28 \times 10^{-4}$ ; OR = 1.19) and entorhinal cortical thickness ( $p = 2.07 \times 10^{-2}$ ;  $t$  value = -2.33). The expression of *CD33* on immune cells is known to be increased in AD and to be associated with inhibition of phagocytosis of A $\beta$ . Cathepsin H, encoded by *CTSH*, is one of the lysosomal cysteine proteases which clear accumulated materials including A $\beta$  by proteolytic processing or autophagy. Paired immunoglobulin-like type 2 receptor alpha, encoded by *PILRA*, is a cell surface inhibitory receptor and is known to be a negative regulator of microglia and neutrophils. Blood-based TRS predicted neuroimaging biomarkers for AD with biological plausibility and replication in an independent cohort.



## PgmNr 2086: Sex-specific analysis of Alzheimer's disease in family-based association studies: An application to two whole-genome sequencing datasets.

### Authors:

D. Prokopenko <sup>1</sup>; J. Hecker <sup>2</sup>; R. Kirchner <sup>2</sup>; K. Mullin <sup>1</sup>; L. Bertram <sup>3</sup>; N. Laird <sup>2</sup>; D.L. Demeo <sup>4</sup>; C. Lange <sup>2</sup>; R.E. Tanzi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology, Massachusetts General Hospital, HMS, Charlestown, Massachusetts, United States; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States; 3) Lübeck Interdisciplinary Platform for Genome Analytics, Institutes of Neurogenetics and Cardiogenetics, University of Lübeck, Lübeck, Germany; 4) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, United States.

---

Late onset Alzheimer's disease (AD) is caused by a complex polygenic and environmental mechanism. It is the 6<sup>th</sup> leading cause of death in the United States. Women are at twofold risk for AD with a more rapid neurodegeneration.

With the arrival of whole genome-sequencing studies, family-based designs enable sex-specific analysis approaches that can be applied to only affected individuals and are completely robust against the effects of population substructure. These advantages make the family-based association tests (FBATs) especially suited for the analysis of diseases such as Alzheimer's Disease (AD). However, the application of FBATs to assess sex-specific effects can require additional filtering steps, as sensitivity of FBATs to sequencing errors is amplified in this type of analysis.

Here, we illustrate the implementation of robust analysis approaches and additional filtering steps that can minimize the chances of false positive-findings due to sex-specific sequencing errors. FBAT framework is used to model sex-specific effects separately and perform a joint analysis of the main genetic effect and the sex-specific effect. We apply this approach to two family-based datasets consisting of 558 multiplex AD families. We identify three novel genetic loci showing sex-specific association with AD risk. Following strict quality control filtering, the strongest candidate was *ZBTB7C*, conferring increased risk for AD in females and protection in males. *ZBTB7C* encodes the Zinc Finger and BTB Domain Containing 7C, a transcriptional repressor of membrane metalloproteases (MMP), including members of this family implicated in AD neuropathology.

## PgmNr 2087: Using the lens of local ancestry to focus risk in Alzheimer disease.

### Authors:

F. Rajabli <sup>1</sup>; B.E. Feliciano <sup>2</sup>; K. Celis <sup>1</sup>; K.L. Hamilton-Nelson <sup>1</sup>; P.L. Whitehead <sup>1</sup>; L.D. Adams <sup>1</sup>; P.L. Bussies <sup>1</sup>; C.P. Manrique <sup>1</sup>; V. Rodriguez <sup>1</sup>; T. Starks <sup>3</sup>; G.E. Byfield <sup>3</sup>; C.B. Sierra Lopez <sup>2</sup>; J.L. McCauley <sup>1</sup>; H. Acosta <sup>4</sup>; A. Chinae <sup>2</sup>; B.W. Kunkle <sup>1</sup>; C. Reitz <sup>5</sup>; L.A. Farrer <sup>6</sup>; G.D. Schellenberg <sup>7</sup>; B.N. Vardarajan <sup>5</sup>; J.M. Vance <sup>1,8</sup>; M.L. Cuccaro <sup>1,8</sup>; E.R. Martin <sup>1,8</sup>; J.L. Haines <sup>9</sup>; G.S. Byrd <sup>3</sup>; M.A. Pericak-Vance <sup>1,8</sup>; G.W. Beecham <sup>1,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL; 2) Universidad Central del Caribe, Bayamón, PR; 3) Maya Angelou Center for Health Equity, Wake Forest University, Winston-Salem, NC; 4) Clinica de la Memoria, San Juan, PR; 5) Gertrude H. Sergievsky Center, Taub Institute for Research on the Aging Brain, Departments of Neurology, Psychiatry, and Epidemiology, College of Physicians and Surgeons, Columbia University, New York, NY; 6) Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, Epidemiology, and Biostatistics, Boston University Schools of Medicine and Public Health, Boston, MA; 7) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PE; 8) Dr. John T. MacDonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL; 9) Department of Population & Quantitative Health Sciences, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, OH

### Background:

Alzheimer disease occurs in all ethnic and racial groups. Genetic studies across populations have shown differences in both risk effect size (e.g., *APOE*, *ABCA7*) and risk variants (e.g., *ABCA7*). While the role of genetic “local” ancestral effects for *APOE* is rapidly unfolding, the role of local ancestry at other AD genetic risk/protective factors has not been assessed. This study examines the effect of genetic ancestral backgrounds in admixed African-American (AA) and Caribbean Hispanic Puerto Rican (PR) populations for AD genes identified in non-Hispanic Whites.

### Methods:

Genome-wide genotyping was performed in 5,506 AA (1,776 cases, 3,730 controls) and 1,132 PR (255 cases, 877 controls) individuals. Quality control analyses were done using software PLINK v.2. Global ancestry for each population was estimated by performing Principal Components Analysis (using GENESIS). Local ancestry (LA) was calculated using SHAPEIT (with 1kGP reference) followed by RFMix (using HGDP reference panels). We performed genotype-based logistic regression tests, testing for the interaction between genetic ancestral background and known AD risk/protective variants.

### Results:

Results showed that two genes (*CD33* (rs3865444g): p-value=0.03, *CASS4* (rs7274581): p-value=0.03) in AAs and one gene (*EPHA1* (rs11771145): p-value=0.02) in PRs were significantly associated with AD in the presence of LA. All variants were significantly associated with the European (EU) background around the genes. The main effect for *CD33* and *CASS4* in AAs were not significant, whereas the *EPHA1* main effect was significant for the variants in PRs. The global proportion of European (EU) ancestral background in PRs (~67%) and AAs (~20%) likely explains the difference in

the significance of the main effects among the populations.

**Conclusion:**

These results demonstrate that individuals with an EU genetic region at the *CD33*, and *CASS4* genes in AAs and *EPHA1* gene in PRs have different risk of AD from those who inherited an African local ancestral background. This suggests that genetic risk/protective factors in admixed populations are correlated with the local genetic ancestry emphasizing the need to consider LA when determining risk in admixed groups. Identification of population-specific variation could lead to the development of ancestry-specific AD therapies. This would improve treatments and help reduce health disparities.

## PgmNr 2088: Association of short tandem repeats with neuropathological features in late-onset of Alzheimer's disease brains.

### Authors:

S. Saria<sup>1,2</sup>; J.A. Schneider<sup>4,5,6</sup>; D.A. Bennett<sup>4,5</sup>; P.L. De Jager<sup>1,2,3,7</sup>; B.N. Vardarajan<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Taub Institute for Research on Alzheimer's Disease and the Aging Brain Columbia University, New York, NY, USA; 2) The Gertrude H. Sergievsky Center Columbia University, New York, NY, USA; 3) The Department of Neurology, College of Physicians and Surgeons Columbia University, New York, NY, USA; 4) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois; 5) Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois; 6) Department of Pathology, Rush University Medical Center, Chicago, Illinois; 7) Center for Translational and Computational Neuroimmunology, Columbia University Medical Center, New York, New York

**Background:** Late-Onset Alzheimer's Disease (LOAD) is characterized by genetic heterogeneity and there is no single model explaining the genetic mode of inheritance. Short tandem repeats (STRs), which are hyper-mutable sequences in the human genome could explain some of the missing heritability in LOAD. STRs are involved in several neural-degenerative disorders such as Huntington disease. We systematically evaluated the impact of STRs on neuropathological LOAD features.

**Methods:** We used whole-genome sequencing (WGS) data in for 1,134 unrelated individuals of European ancestry from Religious Orders Study (ROS) and Rush Memory and Aging project (MAP) cohorts to identify genome-wide STRs. WGS was generated from DNA extracted from blood and brain tissues. We used hipSTR to perform STR calling and genotyping with human genome build 19 as reference. For, quality control, we excluded monomorphic STRs, STRs with depth<10 or missing rate>20%. We tested the association of STRs with a) neuropathological LOAD status, b) beta-amyloid levels, c) neurofibrillary tangle (NFT) burden and d) global measure of pathology based on the scaled scores of 5 brain regions. Regression models adjusting for age, sex and first three principal components were used to test association and Bonferroni correction was used to establish significance.

**Results:** Post QC filtering, we identified 917,298 STRs genome-wide. Two STRs within *APOE* gene (19:45423986 and 19:45412779) were associated with global pathology score ( $p=4.5E-17$ ,  $p=3.9E-12$ ), beta-amyloid levels ( $p=6.3E-08$ ,  $3.6E-07$ ) and pathological AD status ( $1.7E-07$  and  $8.2E-07$ ) but not with tangle counts. We found a novel association for a STR (18:7901364) in the *PTPRM* gene with tangle counts ( $p=6.9E-08$ ). The STR has "T" as the repeat motif and an average length of 17. Additionally, *rs6506544* in *PTPRM* which is 0.1 MB from the associated STR, is an expression quantitative trait loci ( $p=9.73E-05$ ).

**Conclusions:** *PTPRM*, tyrosine phosphatase receptor-type mu, is a signaling molecule that regulates various cellular processes such as cell growth and differentiation. *PTPRM* was identified as a 5-Hydroxymethylation-associated epigenetic modifier of Alzheimer's disease that modulates Tau-induced neurotoxicity. The proximity of eQTLs near the associated STR suggests a regulatory role.

*PTPRD*, from the same gene family, has been associated with NFT burden in the same cohort. The results indicate STRs could explain some of the missing heritability in LOAD.

## PgmNr 2089: Genome-wide association study of rate of cognitive decline in Alzheimer's disease patients identifies novel genes and pathways.

### Authors:

R. Sherva<sup>1</sup>; A. Gross<sup>2</sup>; S. Mukherjee<sup>3</sup>; R. Koesterer<sup>4</sup>; P. Amouyel<sup>5</sup>; C. Bellenguez<sup>5</sup>; C. Dufouil<sup>6</sup>; D.A. Bennett<sup>7</sup>; L. Chibnik<sup>8</sup>; C. Cruchaga<sup>9</sup>; J. del-Aguila<sup>9</sup>; L.A. Farrer<sup>1,10</sup>; R. Mayeux<sup>11</sup>; L. Munsie<sup>12</sup>; S.J. Newhouse<sup>13</sup>; A.J. Saykin<sup>14</sup>; P.K. Crane<sup>3</sup>; R.C. Green<sup>15</sup>; The Alzheimer's Disease Genetics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) aDepartment of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA; 2) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Department of Medicine, University of Washington, Seattle, WA; 4) Programs in Metabolism and Medical & Population Genetics, Broad Institute, Cambridge, MA; 5) Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 Facteurs de risque et déterminants moléculaires des maladies liées au vieillissement, F-59000 Lille, Institut Pasteur de Lille, University of Lille, Lille France; 6) hInserm Unit 1219 Bordeaux Population Health, CIC 1401-EC (Clinical Epidemiology), University of Bordeaux, ISPED (Bordeaux School of Public Health), Bordeaux University Hospital, Bordeaux, France; 7) iRush Alzheimer's Disease Center, Rush University Medical Center, Department of Neurological Sciences, Rush University Medical Center, Chicago, Illinois, USA; 8) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 9) Hope Center for Neurological Disorders, Department of Psychiatry, Knight Alzheimer's Disease Research Center, NeuroGenomics and Informatics, Washington University School of Medicine, St. Louis, MO; 10) Departments of Bioinformatics, Neurology, Biostatistics, Ophthalmology, Epidemiology, Boston University Schools of Medicine and Public Health, Boston, MA; 11) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, The Gertrude H. Sergievsky Center, Neurology, Vagelos College of Physicians and Surgeons, Columbia University, and Columbia University Medical Center, New York, NY; 12) Eli Lilly and Company, Indianapolis, IN; 13) zDepartment of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London; 14) Indiana Alzheimer Disease Center and Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN; 15) ffDivision of Genetics, Department of Medicine, Brigham and Women's Hospital; Harvard Medical School, Partners HealthCare Personalized Medicine, Boston, MA

---

Substantial inter-individual variability exists in the disease trajectories of Alzheimer's disease (AD) patients, some of which is likely due to genetics. We describe a genome wide association study to examine rate of cognitive decline in a sample of AD patients with longitudinal measures of cognition. We tested for interactions between genetic variants and time since AD diagnosis to predict a composite measure of global cognitive performance in a sample of 3,946 AD cases from eleven cohorts and performed pathway analysis on the top genes identified. Suggestive associations ( $p < 1.0 \times 10^{-6}$ ) were observed in a large region on chromosome 15 that includes DNA polymerase- $\gamma$  (*POLG*) (rs3176205,  $p = 1.11 \times 10^{-7}$ ), on chromosome 7 (rs60465337,  $p = 4.06 \times 10^{-7}$ ) in contactin-associated protein 2 (*CNTNAP2*), in lincRNA RP11-384F7.1 on chromosome 3 (rs28853947,  $p = 5.93 \times 10^{-7}$ ), and family with sequence similarity 214 member A (*FAM214A*) on chromosome 15

(rs2899492,  $p=5.94 \times 10^{-7}$ ); as well as intergenic regions on chromosomes 16 (rs4949142,  $p_{GEE}=4.02 \times 10^{-7}$ ), and 4 (rs1304013,  $p=7.73 \times 10^{-7}$ ). One variant (rs1476679,  $p_{AD}=5.6 \times 10^{-10}$ ) in previously established AD risk gene *ZCWPW1*, was significantly associated with rate of decline ( $p_{decline}=3.07 \times 10^{-6}$ ). Ten genes (*APBA1*, *BANK1*, *KCNH5*, *NEGR1*, *PDE4D*, *PTPRN2*, *RBFOX1*, *SGCZ*, *SLC17A3*, and *ZCCHC4*) significantly associated with intelligence were also among the top results. Several significant pathways were identified including ones related to neuronal development and function (Gαq signaling, ephrin signaling, synaptic long term depression, axonal guidance signaling), neuronal apoptosis (Huntington's disease signaling, phospholipase C signaling), memory (CREB signaling in neurons, protein kinase A signaling), and inflammation and immunity (CXCR4 signaling, thrombin signaling). These results suggest that pathways related to AD, intelligence, and neurological function all determine rate of cognitive decline, while the previously identified AD risk variants, including ApoE, do not have a major impact.

## PgmNr 2090: Multiomic data integration in Alzheimer's disease using deep learning.

**Authors:**

P. Singhal; J.E. Miller; A. Verma; S.S. Verma; S.M. Dudek; M.D. Ritchie

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics, University of Pennsylvania, Philadelphia, PA

---

Despite intense investigation into preclinical Alzheimer's Disease (AD) models, there are currently no approved disease-modifying drugs. Genome-wide association studies have proposed a number of genes significantly associated with AD across different pathways, however these genes explain a fraction of disease heritability. The prevailing view of AD as a singular disease with one driving genetic mechanism is oversimplified and does not consider the extensive phenotypic heterogeneity observed. While almost all patients present the hallmark b-amyloid plaque and neurofibrillary tangle neuropathology, there is significant variability in severity of cognitive symptoms, age of onset, and neurophysiology across patients. This suggests that there are likely multiple different genetic mechanisms and pathways in the patient population driving disease to the commonly observed neurodegeneration. High-throughput studies have revealed that AD is a result of complex interactions within and between the genome, transcriptome, epigenome, and proteome. Given that univariate association methods are unable to detect such nonlinear interactions, in this study we interrogate the complex genetic architecture of AD by using a deep learning framework to integrate multiomics data for AD case-control individuals. Using prior biological knowledgebases, we mapped gene-level multiomics data to pathways and aggregated the data to generate "pathway scores". Using a neural network approach, we combined pathway scores from each omic data type to determine models of candidate AD pathways that best explain disease status. Preliminary analysis from this study uses multiomics AD data from Religious Orders Study and Memory and Aging Project. We integrated RNA-sequencing and methylation data pathway scores to identify pathway models with accuracy of 0.71 and area under the curve of 0.69. Pathways known to be associated with AD from the literature such as amino acid synthesis, vesicular transport, and DNA damage repair were identified in our models, in addition to novel innate immunity pathways not previously associated with AD such as NOD-like receptor signaling and leishmaniasis infection. These results reveal there is remaining complex AD biology that can be identified only through simultaneous evaluation of multiomics data. Our study addresses the larger challenges in data dimensionality reduction and data integration, and contributes to the identification of novel disease pathway models.



## PgmNr 2091: Circadian-relevant gene *PERIOD3* is related to autism spectrum disorder and has function neuronal development.

### Authors:

A. Miyauchi <sup>1</sup>; M. Noda <sup>2</sup>; A. Matsumoto <sup>1</sup>; M. Goto <sup>1</sup>; K. Kojima <sup>1</sup>; H. Osaka <sup>1</sup>; E. Jimbo <sup>1</sup>; K. Nagata <sup>2</sup>; T. Yamagata <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Jichi Medical University, Tochigi, Japan; 2) Department of Molecular Neurobiology, Institute for Developmental Research, Aichi Developmental Disability Center, Kasugai, Japan

---

Synaptic dysfunction is considered to be a part of pathology for autism spectrum disorder (ASD) since many genes associated with synaptic function were detected as candidates for ASD. On the other hand, 44-83% of children with ASD have sleep problems and some circadian-relevant genes such as *PER1* and *MTNR1A/B* have been reported to be associated with ASD. Previously, we detected mutations in several circadian-relevant genes and revealed that *NR1D1* contributed to ASD and had the role on neural network formation. This time, we focused on *PERIOD3* (*PER3*) to analyze for the mutations in ASD patients and also the contribution to neuronal development. *PER3* that is located on 1p36.23 with 25 exons belongs to the period family. However, exact function is not well understood. (Methods) We screened Japanese ASD patients for *PER3* mutation by Sanger sequencing analysis. After obtained informed consent from parents, genomic DNA was extracted from lymphocytes, amplified each exon and its neighboring introns by PCR, and subjected to sequence. We also analyzed the role of *Per3* on the brain during embryogenesis by using in utero electroporation in the mouse. In utero electroporation, *Per3* was suppressed using RNAi on embryonic day (E) 14.5 and the brain was analyzed after birth.

(Results) We analyzed only up to 100 patients until now, but detected 13 kinds of miss sense base changes. Among them, six mutations were judged as damaging in polyPhen-2 analysis, and four of them, p.V134M, p.A264D, p.A554T and p.Q1085K, were not reported as SNPs previously. Acute knockdown of *Per3* with in utero electroporation caused abnormal positioning of cortical neurons, which was rescued by RNAi-resistant *Per3*. *Per3*-deficient cells showed abnormal migration phenotypes, impaired axon extension and dendritic arbor formation.

(Discussion) High incidence of rare mutations affecting the gene function suggested the relation of *PER3* to ASD. And also, *Per3* was found to play a pivotal role in corticogenesis via regulation of excitatory neuron migration and synaptic network formation. Addition to our previous results of *NR1D1*, it is suggested that the circadian related genes closely related to ASD, and had function on neuronal development and synaptic network formation besides the circadian rhythm formation.

## PgmNr 2092: Assessment of the ancestral origin of multiple sclerosis risk within the major histocompatibility complex in a large Hispanic patient cohort.

### Authors:

A.H. Beecham<sup>1,2</sup>; L. Amezcua<sup>3</sup>; A. Chinae<sup>4</sup>; G.W. Beecham<sup>1,2</sup>; S.R. Delgado<sup>5</sup>; N. Patsopoulos<sup>6</sup>; J.R. Oksenberg<sup>7</sup>; J.L. McCauley<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 3) Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 4) San Juan MS Center, Guaynabo, Puerto Rico, USA; 5) Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA; 6) Department of Neurology and Ann Romney Center for Neurological Diseases, Brigham and Women's Hospital, Boston, MA, USA; 7) Department of Neurology, University of California at San Francisco, San Francisco, CA, USA

---

The Major Histocompatibility Complex (MHC) demonstrates the strongest genetic association with multiple sclerosis (MS), with additive and dominant effects at 32 independent alleles explaining ~20% of the estimated heritability within European populations. Despite the importance of the region for understanding the etiology of MS, the complex genomic structure has made it difficult to comprehensively identify risk alleles across populations. We therefore sought to characterize the MHC within a Hispanic cohort consisting of 1298 MS cases and 1354 controls. Dense genotype data across the MHC were available from a custom Illumina array, termed the MS Chip. Classical HLA alleles were imputed using HIBAG, and logistic regression was used to assess the association of each SNP and classical allele with MS after adjustment for global ancestry. A forward stepwise approach was used to identify independent signals. In order to assess the ancestral origin of risk, several approaches were utilized. First, admixture mapping identified regions within the MHC of differential local ancestry between MS cases and controls. Second, risk allele and haplotype frequencies for MS cases and controls were compared to those frequencies observed in the 1000 Genomes populations. Lastly, principal components analysis was used to compare the Identity by Descent distances between the Hispanics and 1000 Genomes samples; first using markers within the MHC and then using ancestry informative markers (AIMs). Greater variability was observed with the MHC than with the AIMs components, due to regionally high levels of selection within the MHC. Differential clustering of MS cases and controls in the MHC only analysis indicated ancestral specific risk haplotypes. Collectively these analyses provided insight into the ancestral origins of MHC alleles conferring MS risk in our Hispanic cohort.

## PgmNr 2093: Genome-wide association of myasthenia gravis.

### Authors:

R. Chia<sup>1</sup>; N. Murphy<sup>1</sup>; C. Blauwendraat<sup>2</sup>; C. Provenzano<sup>3</sup>; D. Drachman<sup>1,4</sup>; B. Traynor<sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neuromuscular Disease Research Section, Laboratory of Neurogenetics, NIH, Bethesda, MD; 2) Molecular Genetics Section, Laboratory of Neurogenetics, NIH, Bethesda, MD; 3) Institute of General Pathology, Catholic University, Rome, Italy; 4) Department of Neurology, Johns Hopkins University, Baltimore, MD

---

Myasthenia gravis is a chronic disease of the neuromuscular junction causing fluctuating weakness of the skeletal muscles and, in more serious cases, the respiratory muscles. It is an underdiagnosed autoimmune condition affecting 14 to 20 persons per 100,000 in the general population and can occur at any age. Epidemiological studies show a bimodal pattern of incidence where early onset cases are predominantly females under the age of 40 years, but appears to affect both sexes equally after the age of 50 years. Due to aging of the world's population, it is expected that the number of individuals living with this disease will also increase in the coming years.

Although the disease is mediated by circulating auto-antibodies against proteins at the neuromuscular junction, the exact pathogenesis triggering disease onset is unclear. To understand the mechanism underlying this autoimmune condition, we have undertaken a genome-wide approach using a cohort of ~2100 cases and ~46000 healthy controls of European descent to identify genetic variants that alter the risk of myasthenia gravis or drive onset- and sex differences. We replicated previous association signals at *PTPN22* (chr1), *HLA-DQA1* (chr6), and *TNFRS11A* (chr18). We also identified seven new loci on chr2, chr5, chr7, chr8, chr13, and chr14. Additionally, we identified several novel loci that were distinct in male versus female or early versus late onset cases, suggesting that the mechanisms underlying age at onset and risk based on gender may be different. Having the knowledge of which genes or mutations cause or modify MG will facilitate disease modeling and enable the design and testing of new targeted therapeutics tailored to the gender or age at onset of the patient. Thus, the identification of new genes that cause MG is of great significance.

## PgmNr 2094: Genome-wide association analysis of short-tandem repeats in Parkinson's disease.

### Authors:

B.I. Bustos<sup>1</sup>; V. Escott-Price<sup>2</sup>; N.M. Williams<sup>2</sup>; M.A. Nalls<sup>3</sup>; A. Brice<sup>4</sup>; T. Gasser<sup>5,6</sup>; J. Hardy<sup>7,8</sup>; P. Heutink<sup>5,6</sup>; N.W. Wood<sup>7,9</sup>; A.B. Singleton<sup>3</sup>; H.R. Morris<sup>10</sup>; A. Pfaff<sup>11</sup>; S. Koks<sup>12</sup>; J. Quinn<sup>13</sup>; S.J. Lubbe<sup>1</sup>; The International Parkinson's Disease Genomics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology and Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.; 2) Institute of Psychological Medicine and Clinical Neurosciences, Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Department of Psychological Medicine and Neurology, Cardiff University School of Medicine, Cardiff, UK; 3) Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland, USA.; 4) Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France.; 5) German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany.; 6) Department of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany.; 7) Department of Molecular Neuroscience, Institute of Neurology, University College London, London, UK; 8) Reta Lila Weston Institute, University College London Institute of Neurology, Queen Square, London, UK; 9) UCL Genetics Institute, University College London, London, UK.; 10) Department of Clinical Neuroscience, Institute of Neurology, University College London, London, UK.; 11) Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, WA, Australia.; 12) The Perron Institute for Neurological and Translational Science, Nedlands, WA, Australia; 13) Department of Molecular and Clinical Pharmacology, Institute of Translational Medicine, University of Liverpool, Liverpool, UK.

---

Parkinson's disease (PD) is a complex neurodegenerative disorder with a strong genetic component, where most disease associated genetic elements are single nucleotide variants (SNVs) and Indels. The role of more complex and larger genomic events, such as short tandem repeats (STRs) have only been studied for candidate genes, therefore their genome-wide contribution to the disease remains unknown. Here we present a genome-wide association analysis (GWAS) of STRs for PD. We have used a GWAS dataset from the International Parkinson's Disease Genomic Consortium (IPDGC) composed of 15,368 individuals (5,333 PD cases and 10,035 controls) that were recruited from different academic health centers across the US and Europe. The dataset includes 547,364 genotyped SNPs, that served as proxies for STR imputation using a recently published STR reference panel built on the 1000 genome project. Imputation was performed using BEAGLE v.5.0. Case control association analysis was done using logistic regression in PLINK, controlling for gender and population structure. We have successfully imputed 17,465 good quality autosomic STRs ( $R^2 > 0.3$ ,  $MAF > 1\%$ ). The association analysis with PD revealed 28 STRs surpassing the genome-wide significance threshold of  $5 \times 10^{-8}$ , with the strongest signal located in the known PD locus of *MAPT* ( $p=2.43 \times 10^{-20}$ ,  $OR=1.35$ , C.I. 95%= 1.27-1.43), depicted as 5 repetitions of a GAAT motif. We also observed significant STRs in other known PD genes: *KANSL1* ( $p=2.88 \times 10^{-18}$ ,  $OR=1.33$ , C.I. 95%=1.41-1.25), *GAK* ( $p=1.04 \times 10^{-12}$ ,  $OR=1.58$ , C.I. 95%=1.39-1.78) and *IFNAR1* ( $p=3.34 \times 10^{-10}$ ,  $OR=1.20$ , C.I. 95%=1.14-1.28). We identified STRs mapping to genes not previously related to PD: *GNL1* ( $p=3.79 \times 10^{-19}$ ,  $OR=1.38$ , C.I.

95%=1.29-1.48), *CNTLN* ( $p=6.78 \times 10^{-10}$ , OR=1.23, C.I. 95%=1.15-1.32), *TEX14* ( $p=9.52 \times 10^{-10}$ , OR=1.19, C.I. 95%=1.13-1.23), *MSH3* ( $p=1.33 \times 10^{-9}$ , OR=1.22, C.I. 95%=1.14-1.30), *SRSF12* ( $p=1.08 \times 10^{-8}$ , OR=1.21, C.I. 95%=1.13-1.29) and *ZNF723* ( $p=1.42 \times 10^{-8}$ , OR=1.31, C.I. 95%=1.19-1.44). Interestingly, *GNL1*, *CNTLN* and *MSH3* have previously been associated to neurodegenerative disorders, linking at their possible functional roles in PD. These results show that STRs in known and novel associated genes could be involved in the pathogenesis of the disease. Study of multiallelic events, gene-sets and pathway enrichments, heritability contribution and replication in independent populations are currently underway with the aim to expand and confirm our findings.

## PgmNr 2095: Human leukocyte antigen class I variants increase risk of clinically diagnosed migraine.

### Authors:

C. Huang<sup>1,2</sup>; Y.H. Huang<sup>1</sup>; Y.F. Wang<sup>1,2,3</sup>; M.H. Lee<sup>1</sup>; S.P. Chen<sup>1,2,3</sup>; S.J. Wang<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan; 2) Faculty of Medicine, National Yang-Ming University, Taipei, Taiwan; 3) Department of Neurology, Neurological Institute, Taipei Veterans General Hospital, Taipei, Taiwan

### Background

Migraine is a complex neurological disorder with high heritability. Associations between human leukocyte antigen (*HLA*) genes and migraine remained uncertain. We aimed to evaluate the association of *HLA* variants with migraine or headache in clinic and community-based settings.

### Methods

The study included 772 clinic-based migraine patients in a tertiary medical center and 10,019 population-based participants in Taiwan Biobank (TWB). All study subjects were aged between 30-70 and were examined by Axiom Genome-Wide SNP Arrays. The diagnosis of clinic-based cases was ascertained by headache specialists and the individuals in TWB were interviewed by a structured questionnaire including their history of headache and migraine. Single nucleotide polymorphism-based imputation for *HLA-A*, *B*, *C*, *DPA1*, *DPB1*, *DQA1*, *DQB1* and *DRB1* were performed by R package HIBAG with an Asian reference group. After quality control, 605 clinic-based cases remained eligible for subsequent analyses, while 2,394 self-reported headache cases, including 911 with self-reported migraine, were identified in TWB. 6,055 TWB subjects free of self-reported headache or migraine were the control group. We used likelihood ratio tests to examine *HLA* associations with migraine and logistic regressions to estimate odds ratios (OR) and 95% confidence intervals (CI) for *HLA* alleles.

### Results

*HLA-B* and *C* were significantly associated with migraine in clinic-based patients (FDR q-value < 0.05). We found *HLA-B\*39:01*, *HLA-B\*51:01*, *HLA-B\*58:01* and *HLA-C\*03:02* to be significantly associated with migraine, with age and sex-adjusted OR (95% CI) of 1.80 (1.28-2.53), 1.50 (1.15-1.97), 1.36 (1.14-1.62) and 1.36 (1.14-1.62), correspondingly. Among the clinic-based patients, the odds of chronic migraine (CM) with medication-overuse headache (MOH) but not that of CM without MOH were significantly ( $p < 0.05$ ) higher than episodic migraine if patients carried *HLA-A\*33:03*, *HLA-B\*58:01* or *HLA-C\*03:02*. *HLA-A\*33:03*-*HLA-B\*58:01*-*HLA-C\*03:02* was the most common haplotype, found 9.74% in the control population. Individuals carrying the haplotype had 1.44 (1.20-1.73) likelihood of clinical migraine. The attributable fraction of the haplotype was 8%. However, no *HLA* genes were associated with self-reported headache or migraine in community-based cases.

### Conclusion

*HLA* class I genetic variants confer risk of migraine in clinic but not population-based cases and may contribute to medication overuse-related migraine chronification.

## PgmNr 2096: Nanopore sequencing of the glucocerebrosidase (*GBA*) gene in a New Zealand Parkinson's disease cohort.

### Authors:

M.A. Kennedy <sup>1</sup>; O.E.E. Graham <sup>1</sup>; Y. Liao <sup>1</sup>; A.L. Miller <sup>1</sup>; T.L. Pitcher <sup>2,3</sup>; T.J. Anderson <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology and Biomedical Science, University of Otago, Christchurch, New Zealand; 2) Department of Medicine, University of Otago, Christchurch, Christchurch, New Zealand; 3) New Zealand Brain Research Institute, Christchurch, New Zealand

---

**Introduction:** Mutations in the gene for glucocerebrosidase (*GBA*) cause Gaucher disease, an autosomal recessive lysosomal storage disorder. *GBA* mutation carriers are at greatly elevated risk of Parkinson's disease (PD) or dementia with Lewy bodies (DLB) compared with non-carriers. Large studies have shown that ~ 7% of PD cases will carry *GBA* mutations, and these patients may also experience an earlier age of onset and more aggressive disease.

**Methods:** Comprehensive analysis of *GBA* mutations is technically challenging, due to the presence of a closely related pseudogene, and complex structural mutations that can occur at this locus. Using a targeted long PCR step can largely overcome the pseudogene problem, and recently a nanopore sequencing method for complete analysis of this long PCR amplicon was published (Leija-Salazar et al. 2019). We examined *GBA* variants in a large, longitudinal cohort of PD, which includes extensive neuropsychological and clinical assessment. We established the nanopore sequencing method for *GBA* amplicons and analysed samples on a GridION platform (Oxford Nanopore Technology, UK). Sequencing was carried out on R9.4 flow cells, with barcoding and pooling of samples in each run. Guppy base caller was used, and alignment was carried out with Minimap2 and NGMLR software.

**Results:** *GBA* amplicon sequencing revealed 19 mutations in 17 PD cases. All variants were confirmed by Sanger sequencing. Mutations known to be associated with risk of PD included p.E365K (rs2230288) in 10 cases, p.N409S (rs76763715) in one case and R78C (rs146774384) in one case. Mutations possibly associated with PD included D179H (rs147138516) in one case, who had E365K on the same allele, and p.T408M (rs75548401) in six cases, one of whom had p.E365K on the other allele. One potentially novel mutation, p.L335= (C>T), was also detected, which is predicted to impact splicing of *GBA* transcripts.

### Conclusions:

We have further confirmed the utility of nanopore sequencing of *GBA* as a means to identify variants in multiple patient samples. In a large New Zealand PD cohort we found multiple known *GBA* mutations, and several variants of unclear significance. This work may contribute to a better understanding of the effects of these mutations on disease presentation or progression, and potentially to the development of more targeted treatments.

**Reference:** Leija-Salazar et al. 2019. Mol Genet Genomic Med 7(3):e564.



## PgmNr 2097: Identification of candidate amyotrophic lateral sclerosis risk loci using pedigree based analyses.

### Authors:

K.L. Russell <sup>1</sup>; J.M. Downie <sup>1</sup>; S.B. Gibson <sup>2</sup>; K.P. Figueroa <sup>2</sup>; M.B. Bromberg <sup>2</sup>; S.M. Pulst <sup>2</sup>; L.B. Jorde <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Univ of Utah, Salt Lake City, Utah.; 2) Neurology, Univ of Utah, Salt Lake City, Utah.

---

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease that is characterized by upper and lower motor neuron loss. This neuronal degeneration leads to loss of voluntary muscle movement and eventual death 3-5 years after symptom onset. The molecular mechanisms that cause neuronal demise in ALS are not well understood. ALS can be classified into familial and sporadic forms. Familial ALS (FALS), which accounts for 10% of ALS cases, is defined by a case having a first or second-degree relative that is also affected. Sporadic ALS (SALS) accounts for the remaining 90% of cases and has a heritability estimate of ~0.60. However, only 17% of SALS cases have a pathogenic variant in a known ALS-causing gene, suggesting that additional causal genes remain to be elucidated. To investigate the potential role of *de novo* mutations in SALS causation, we used ALS family trios to detect potential ALS risk variants present in probands, but absent in both unaffected parents. Two trios were selected from a large cohort of ALS cases on whom whole-genome sequencing (WGS) was carried out at an average coverage of 60x. BWA and GATK (v3.3) were used to perform read alignment and variant calling. To identify candidate ALS risk genes, the pedigree-Variant Annotation Analysis and Search Tool (pVAAST) software was used. The software first calculates logarithm of odds (LOD) scores to estimate linkage between the identified variants and disease. It then combines LOD scores with VAAST scores, which identify genes that are more burdened with deleterious, rare (MAF<0.01) variants in case genomes compared to 96 WGS controls. After removal of false positives, the pVAAST analysis identified *PRPF40B* as a top candidate risk gene with a proband having the deletion chr12:50026795\_50026796delAG. Interestingly, this mutation is not present in the gnomAD database and is predicted to disrupt transcript splicing. *PRPF40B* codes for pre-mRNA-processing factor 40 homolog B protein, which functions in mRNA splicing. We find this intriguing because other known-ALS genes encode for proteins that function in RNA binding. Additionally, *PRPF40B* knockdown increases cell apoptosis. Future experiments include determining if the deletion renders the protein nonfunctional. These results indicate that ALS trio analysis can help to identify novel candidate ALS risk loci.



## PgmNr 2098: A unified framework for global and local genetic correlation analysis.

### Authors:

Y. Zhang <sup>1</sup>; Q. Lu <sup>2</sup>; H. Zhao <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, Yale School of Public Health, New Haven, CT; 2) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI; 3) Department of Genetics, Yale School of Medicine, New Haven, CT

---

Genetic correlation analysis based on summary statistics from genome-wide association studies (GWASs) has quickly gained popularity in the past few years and provided fundamental new insights into the shared genetic architecture of many diseases and traits. However, most applications only report the global genetic correlation, a metric that quantifies the overall genetic sharing between two phenotypes. Although it is of great interest to identify and understand specific genetic regions contributing the most to the overall genetic similarity, existing analytic tools face several major technical challenges including complex linkage disequilibrium (LD) in local regions and pervasive sample overlap across GWASs. Due to these issues, existing methods provide biased estimates for local genetic correlations with inflated rates of false findings. Here, we introduce SUPERGNOVA, a unified framework for diverse types of genetic correlation analyses. Through extensive simulations and analyses of real data, we demonstrate that SUPERGNOVA provides statistically rigorous and computationally efficient inference for both global and local genetic correlations. Additionally, our approach only uses GWAS summary statistics as input, and is robust against overlapping GWAS samples when the shared sample size is unknown. Compared to existing methods, SUPERGNOVA provides comparable estimates on global genetic correlations but substantially outperforms all methods when applied to local genetic regions. We applied SUPERGNOVA to investigate the shared genetic basis of two neurodevelopmental disorders, attention deficit/hyperactivity disorder (ADHD;  $n=53,293$ ) and autism spectrum disorder (ASD;  $n=46,351$ ). Consistent with the literature, we confirmed a positive genetic correlation between ADHD and ASD ( $\text{cor}=0.27$ ,  $p=5.5\text{e-}4$ ). Further, we identified six specific regions showing significant local genetic correlations after correcting for multiple testing, including chromosomes 20p11.22 ( $p=1.5\text{e-}22$ ; *KIZ*), 17q21.31 ( $p=5.6\text{e-}9$ ; *MAPT*), 10q25.1 ( $p=2.1\text{e-}7$ ; *SORCS3*), 4q24 ( $p=5.8\text{e-}7$ ; *MANBA*), 5q21.2 ( $p=1.4\text{e-}6$ ; intergenic), and 1p21.3 ( $p=7.0\text{e-}6$ ; intergenic). Among these regions, 1p21.3 is the only locus implicated in GWASs for both ASD and ADHD. 10q25.1 was identified for ADHD and 20p11.22 is a known ASD locus. The other three loci are novel findings for ASD and ADHD. Taken together, SUPERGNOVA is a major advancement in genetic correlation analysis and has wide applications in complex trait research.

## PgmNr 2099: Prioritization of ALS and PD causal genes from GWAS using protein-protein network analysis.

### Authors:

M. Lalioti <sup>1</sup>; G. Kuthethur Gururaj <sup>2</sup>; S. Artavanis-Tsakonas <sup>3</sup>; M. Cokol <sup>1</sup>; J. Guo <sup>1</sup>; K. Li <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Translational Genome Sciences, Biogen, Cambridge, Massachusetts.; 2) Neuromuscular Disorders Research Unit, Biogen, Cambridge, Massachusetts.; 3) Cell Biology, Harvard Medical School, Boston, MA; 4) Human Target Validation Hub, Biogen, Cambridge, Massachusetts.

---

**Background:** Amyotrophic Lateral Sclerosis (ALS) and Parkinson's disease (PD) are neurodegenerative disorders with complex genetic architecture. Mutations in genes, including genes coding for RNA-binding proteins (RBPs), account for less than 20% of the cases. Genome-wide association studies (GWAS) have identified many genomic loci with strong associations with ALS and PD. However, the functional consequences of the causal SNPs are yet to be determined, which limits the ability to translate the GWAS hits into actionable therapeutic hypothesis.

**Methods:** The Disease Association Protein-Protein Link Evaluator (DAPPLE) is a tool developed to prioritize genes in the linkage disequilibrium (LD) block containing the causal SNP, using protein-protein interaction (PPI) networks. A PPI network was constructed by expressing 160 RBPs in HEK293T cells, followed by immunoprecipitation and mass-spec. All baits were selected because they cause genetic forms of neurodegeneration in humans. The network was constructed using CompPASS developed by Harvard. InWeb (combined network of inWeb3 and inWeb\_IM, licensed from Intomics) network was also used. The RBP network is dense, composed of 1,843 proteins and 5,433 interactions.

**Results:** DAPPLE based on GWAS SNPs only was unable to prioritize any genes. For this reason, eQTLs and known disease genes were spiked in, to create eDAPPLE (enhanced DAPPLE) with markedly improved results. For example, for ALS, eDAPPLE identified 13 direct interactions compared to 0 identified by DAPPLE. Known genes had the biggest effect on eDAPPLE predictions, while eQTLs alone had variable effect. Combination of RBP network with InWeb network identified significantly more interactions than InWeb or RBP only network, for both ALS and PD subnetworks. Common interactors between ALS and PD were also identified by using this combined network.

**Conclusion:** The curated RBP PPI network, although considerably smaller than the inWeb, significantly improved eDAPPLE performance, presumably because it consists of disease-specific interactions. The addition of eQTLs and known ALS and PD genes was necessary to allow DAPPLE to identify new disease modules and prioritize genes for further studies. Functional validation of interactions will be done by network perturbation in animal models and/or iPSCs. Common interactors between diseases indicated that the shared pathways may be affected and supporting the hypothesis that a common drug may ameliorate symptoms of both.

## PgmNr 2100: Rare genomic variants impacting neuronal functions modify the Dup7q11.23 phenotype.

### Authors:

F. Qaiser <sup>1,2</sup>; Y. Yin <sup>2</sup>; C.B. Mervis <sup>3</sup>; C.A. Morris <sup>4</sup>; B.P. Klein-Tasman <sup>5</sup>; L.R. Osborne <sup>1,6</sup>; R.K.C. Yuen <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular Genetics, University of Toronto; 2) Genetics & Genome Biology Program, The Hospital for Sick Children; 3) Department of Psychological and Brain Sciences, University of Louisville, KY; 4) Department of Pediatrics, University of Nevada, NV; 5) Department of Psychology, University of Wisconsin-Milwaukee; 6) Department of Medicine, University of Toronto

---

**Background:** Duplication of the 7q11.23 region (Dup7) results in a complex neurodevelopmental disorder. Common characteristics include distinctive craniofacial features, abnormal neurological findings, low-average intellectual ability and Social Anxiety Disorder. Interestingly, Dup7 is one of the most frequent (0.2%) recurrent copy number variants (CNVs) in individuals with Autism Spectrum Disorder (ASD). However, only 19% of Dup7 carriers have ASD, suggesting that additional genetic factors may be necessary to manifest the ASD phenotype.

**Methods:** To assess the contribution of additional rare variants to the phenotype of individuals with Dup7, we conducted systematic whole-genome sequencing (WGS) analysis of 24 Dup7 carriers: 12 with ASD (Dup7-ASD) and 12 without ASD (Dup7-non-ASD). We analyzed microarray and WGS data through burden, functional and pathway enrichment analyses, and correlated the results with categorical and quantitative phenotypic measures.

**Results:** We identified three variants of clinical significance: a pathogenic loss-of-function (LoF) frameshift deletion variant in *EP400* in one Dup7-ASD individual, and two rare deletions which disrupt *IMMP2L* at the 7q31.1 locus in one Dup7-ASD and one Dup7-non-ASD individual. No findings differentiated the Dup7-ASD group from the Dup7 Non-ASD group. When analyzing the Dup7 cohort as a whole, we found that adaptive behaviour scores (measured by SIB-R) were negatively correlated with an enrichment of rare likely damaging missense variants in genes expressed in the prenatal human brain ( $p=0.004$ ). In addition, intellectual ability (measured on DAS-II) was positively correlated with an enrichment for rare CNVs (non-Dup7) impacting genes with a low to medium mutation intolerance ( $p=0.002$ ) and in structural molecular activity pathways ( $p=0.0001$ ). Lastly, intellectual ability was negatively correlated with an enrichment for rare likely damaging LoF variants in nervous system development ( $p=0.002$ ) and membrane component pathways ( $p=0.001$ ).

**Conclusion:** Overall, WGS characterization of Dup7 carriers suggests that in the presence of the same primary pathogenic Dup7 variant, there are additional genetic variants which tend to have low penetrance and act additively to the overall Dup7-related phenotype. This study highlights the power of pathway analyses and the correlation of quantitative traits towards understanding how additional rare variants in the genetic background modulate neurodevelopmental phenotypes.

## PgmNr 2101: Creating a phenome risk score to identify developmental stuttering in electronic health records.

### Authors:

D. Pruett <sup>1</sup>; D. Shaw <sup>2</sup>; L. Petty <sup>2</sup>; H. Chen <sup>2</sup>; S. Kraft <sup>3</sup>; R. Jones <sup>1</sup>; J. Below <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN.; 2) Dept of Medicine, Division of Genetic Medicine, Vanderbilt University, Nashville, TN; 3) Dept of Communication Sciences and Disorders, Wayne State University, Detroit, MI

---

**Introduction:** Developmental stuttering is a speech disorder that typically begins around age 3 and is characterized by syllable repetitions and prolongations. Despite population prevalence of 1%, no population-based genetic studies of stuttering have been conducted and few, if any, studies have utilized DNA database-linked electronic health records (EHRs). Given stuttering is diagnosed by speech pathologists, not physicians, information regarding diagnosis is not consistently noted in EHRs. Rather than relying on ICD 9/10 codes, medications, or other typical approaches to defining phenotypes in EHRs, text-based keyword searches prove more helpful in defining behavioral phenotypes like stuttering. However, manually reviewing files following keyword search is impractical due to the number of records returned. Therefore, the purpose of this project is to develop a phenome risk score to identify likely cases of developmental stuttering in EHRs; enabling use of large-scale DNA biobank resources for well-powered GWAS and epidemiological studies of stuttering.

**Method/Results:** A keyword text search for stuttering terms within the Synthetic Derivative (SD), a de-identified EHR database at Vanderbilt University Medical Center, returned ~14,000 individuals with at least one keyword mention in their records. Files were manually reviewed by an expert in speech pathology who identified 50 confirmed cases of developmental stuttering and an additional 50 files were confirmed as *not* cases of developmental stuttering. A text mining word cloud algorithm was developed to identify stuttering cases and found that the frequency of mentions of *stuttering*, *mother*, *school*, and *childhood* distinguished the stuttering cases from the exemptions. Five percent of the 14,000 files were manually reviewed to test the PPV of the text mining algorithm. We then examined ICD 9/10 codes, labs, and other values associated with developmental stuttering in the SD. Results show enrichment for a variety of comorbid disorders; some were expected based on previous population studies (anxiety, ADHD), others were previously unidentified (constipation, anemias). From this data, we created a phenome risk score by extracting weights for each comorbidity to identify likely cases of stuttering. We applied the phenome risk score in BioVU, VUMC's biorepository of DNA linked to de-identified medical records, and report results from the first GWAS study of quantitative risk scores for developmental stuttering.

## PgmNr 2102: The role of copy number variants in the genetic architecture of common, complex epilepsies.

### Authors:

E.P. Almanza Fuerte; H.C. Mefford; Epi4K Consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Pediatrics, University of Washington, Seattle, Washington.

---

**Background:** Epilepsy is a group of neurological disorders characterized by recurrent, unprovoked seizures. Major classes of epilepsy include focal epilepsy, generalized epilepsy and epileptic encephalopathies. While the encephalopathies are largely due to de novo genetic and genomic mutations, the more common focal and generalized epilepsies appear to be due to complex genetic inheritance. Copy number variants (CNVs) are known to contribute to the genetic etiology of epilepsy and can act as de novo pathogenic variants or inherited risk variants. Here we investigate the role of CNVs in multiplex families with focal or generalized epilepsy. **Methods:** We performed CNV analysis using exome data from 303 multiplex families with at least 3 individuals with non-acquired seizures, at least 2 of which had a diagnosis of one or more of the common epilepsy syndromes. Detailed phenotyping was previously performed for each affected individual, and exome sequence data was generated for at least two affected individuals in each family. To identify potential CNVs, we analyzed exome data with CoNIFER; recurrent CNVs and CNVs involving epilepsy-related genes were prioritized for follow-up studies. High-priority CNVs were validated using array comparative genomic hybridization (aCGH). Segregation studies were performed when possible for all validated CNVs. **Results:** We validated 11 CNVs in 9 (3%) families. In two cases, the CNVs are likely pathogenic deletions of known epilepsy genes: de novo deletion of HNRNPU in one subject from a family, and deletion of the DEPDC5 gene in affected siblings and an unaffected parent of another family. In five families, we identified recurrent deletions known to be associated with increased risk of epilepsy and developmental disorders: 15q11.2 deletion (3x), 15q13 deletion, and 16p13 deletion. Four CNVs are variants of unknown significance: 15q13.2 duplication, 18q deletion, CHRNA7 duplication, and a duplication of the entire SCN1A gene. Notably, the duplication of SCN1A, a known epilepsy gene, was inherited from an unaffected parent. None of the CNVs we identified segregated perfectly with disease in all affected individuals of a family. **Conclusion:** CNVs contribute to the complex genetic architecture of the common epilepsies. Analysis of exome data for contributing single nucleotide variants in these same families is ongoing.

## PgmNr 2103: Assessment of focal epilepsy subtypes using polygenic risk scores.

### Authors:

M. Gramm<sup>1,2</sup>; C. Leu<sup>1,3</sup>; L. Ferguson<sup>4,5,6</sup>; L. Jehi<sup>4,6</sup>; M.J. Daly<sup>3,7,8</sup>; I.M. Najm<sup>4,6</sup>; R.M. Busch<sup>4,5,6</sup>; D. Lal<sup>1,2,3,4,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, US; 2) Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, DE; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT & Harvard, Cambridge, MA 02142, US; 4) Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, US; 5) Department of Psychiatry & Psychology, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, US; 6) Department of Neurology, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, US; 7) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, US; 8) Institute of Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, FI

---

Focal epilepsies (FE) comprise a clinically and genetically heterogeneous group of disorders. An array of comorbid neurological and psychiatric disorders contribute to the clinical heterogeneity of FE. The cause of FE can be acquired or non-acquired (idiopathic). Common causes for acquired forms of FE include stroke, traumatic brain injuries, brain tumors, central nervous system infections, or vascular insults. Recently, it has been shown that common genetic variants can confer risk for focal epilepsy. These variants, combined into FE-Polygenic risk scores (FE-PRS) can identify FE patients with a higher genetic burden for FE compared to population controls. Currently, large research studies explore potential clinical applications for FE-PRS - for example, assessing the risk of developing FE after a first seizure. The goal of this study was to determine if the predictive power of FE-PRS differ across subtypes of FE. We generated genotype data for 430 patients with FE ascertained at the Cleveland Clinic. We subgrouped FE patients and compared the FE-PRS burden against >20k ethnically matched controls for which genetic data were available. FE patient groups were selected based on individual clinical variables (acquired/non-acquired epilepsy, early/late seizure onset, occurrence/no occurrence of psychiatric comorbidities) or combined using a computational clinical data clustering approach. We observed that individuals with non-acquired FE and/or an early seizure onset (<21 years) showed higher FE-PRS compared to controls. In contrast, individuals with acquired FE, defined by the presence of a possible cause (stroke, traumatic brain injuries, brain tumors, central nervous system infections, or vascular insults) are not different from population controls. The highest mean FE-PRS was observed in a group defined by the clustering approach, enriched for individuals with non-acquired FE and early seizure onset. In summary, we observe a heterogeneous burden of PRS in different focal epilepsy subtypes defined by deep clinical phenotyping and clustering analysis. Particularly, the current available FE-PRS have the most predictive power in young FE patients of unknown etiology.



## PgmNr 2104: Characterization of focal and generalized epilepsy polygenic burden in 630,603 individuals.

### Authors:

C. Leu<sup>1,2,3</sup>; R. Stevelink<sup>4</sup>; A. Smith<sup>2</sup>; T.G. Richardson<sup>5</sup>; S.B. Goleva<sup>6,7</sup>; M. Kanai<sup>2,8,9,10,11</sup>; L. Ferguson<sup>12,13,14</sup>; C. Campbell<sup>15,16</sup>; Y. Kamatani<sup>11,17</sup>; Y. Okada<sup>11,18,19</sup>; S.M. Sisodiya<sup>3,20</sup>; G.L. Cavalleri<sup>15,16</sup>; B.P.C. Koeleman<sup>4</sup>; H. Lerche<sup>21</sup>; G. Davey Smith<sup>5</sup>; L. Jehi<sup>12,14</sup>; L.K. Davis<sup>6,7</sup>; I.M. Najm<sup>12,14</sup>; A. Palotie<sup>2,22</sup>; M.J. Daly<sup>2,8</sup>; R.M. Busch<sup>12,13,14</sup>; Epi. 25 Consortium<sup>23</sup>; D. Lal<sup>1,2,12,24</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, US; 2) Stanley Center for Psychiatric Research, Broad Institute of Harvard and M.I.T, Cambridge, MA 02142, US; 3) Department of Clinical and Experimental Epilepsy, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK; 4) Department of Genetics, University Medical Center Utrecht, Utrecht, NL; 5) MRC Integrative Epidemiology Unit (IEU), Population Health Sciences, Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK; 6) Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, US; 7) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, US; 8) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, US; 9) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, US; 10) Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, US; 11) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, JP; 12) Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, US; 13) Department of Psychiatry & Psychology, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, US; 14) Department of Neurology, Neurological Institute, Cleveland Clinic, Cleveland, OH 44195, US; 15) Department of Molecular and Cellular Therapeutics, The Royal College of Surgeons in Ireland, Dublin 2, IE; 16) The FutureNeuro Research Centre, Dublin 2, IE; 17) Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, JP; 18) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita 565-0871, JP; 19) Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita 565-0871, JP; 20) Chalfont Centre for Epilepsy, Chalfont-St-Peter, Buckinghamshire SL9 0RJ, UK; 21) Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, DE; 22) Institute of Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, FI; 23) <http://epi-25.org/>; 24) Cologne Center for Genomics (CCG), University of Cologne, 50931 Cologne, DE

**Purpose:** Rare genetic variants can cause epilepsy, and genetic screening has been widely adopted for severe, pediatric-onset epilepsies. The phenotypic consequences of common genetic risk burden for epilepsies and their potential future clinical applications have not yet been determined.

**Method:** Using polygenic risk scores (PRS) from a European-ancestry genome-wide association study in generalized (GE) and focal epilepsy (FE), we quantified common genetic burden in people with GE or FE from two independent non-Finnish European-ancestry cohorts (*Epi25* Consortium, N=5,705; *Cleveland* Clinic Epilepsy Center, N=620; both compared to 20,435 controls). One Finnish-ancestry population isolate (*Finnish-ancestry Epi25*, N=449; compared to 1,559 controls), two European-

ancestry biobanks (*UK Biobank*, N=383,656; *Vanderbilt biorepository*, N=49,494), and one Japanese-ancestry biobank (*BioBank Japan*, N=168,680) were available for additional explorations.

**Results:** Across 630,603 individuals, we found significantly higher GE-PRS in people with GE compared to people with FE (*Epi25*:  $P=1.64 \times 10^{-15}$ ; *Cleveland*:  $P=2.85 \times 10^{-4}$ ; *Finnish-ancestry Epi25*:  $P=1.80 \times 10^{-4}$ ) or population controls (*Epi25*:  $P=2.35 \times 10^{-70}$ ; *Cleveland*:  $P=1.43 \times 10^{-7}$ ; *Finnish-ancestry Epi25*:  $P=3.11 \times 10^{-4}$ ; *UK Biobank* and *Vanderbilt biorepository* meta-analysis:  $P=7.99 \times 10^{-4}$ ). FE-PRS were significantly higher in patients with FE compared to controls in the non-Finnish, non-biobank cohorts (*Epi25*:  $P=5.74 \times 10^{-19}$ ; *Cleveland*:  $P=1.69 \times 10^{-6}$ ). European-ancestry derived PRS did not predict GE or FE status in Japanese-ancestry people. We observed significant 4.63-fold and 4.47-fold enrichments of people with GE compared to controls in the top 0.5% highest GE-PRS of the two non-Finnish European-ancestry cohorts, respectively. Finally, in 383,197 European-ancestry individuals from the UK Biobank without epilepsy, high GE-PRS or FE-PRS were associated with low educational attainment and several personality/mood traits.

**Conclusion:** Focal and generalized epilepsy common variant burden is associated with specific epilepsy types and neuropsychiatric phenotypes. As sample sizes and PRS accuracy continues to increase with further common variant discovery, PRS could complement established clinical biomarkers and become valuable tools for patient classification, comorbidity research, and potentially targeted treatment.



## PgmNr 2105: Whole exome sequencing in patients with *gba1* mutations may identify pathways implicated in Parkinson disease.

### Authors:

P. Sharma; A. Steward; G. Lopez; N. Tayebi; E. Sidransky

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medical Genetics Branch, NHGRI, Section on Molecular Neurogenetics, National Institute of Health, USA, Bethesda, Maryland.

---

While it is known that mutations in *GBA1*, the gene mutated in the lysosomal storage disorder Gaucher disease, are an important risk factor for the development of Parkinsonism, the majority of patients with Gaucher disease (GD) and *GBA1* carriers never develop Parkinson disease (PD). To explore other factors contributing to Parkinson pathogenesis in these individuals, Whole Exome Sequencing (WES) was performed on 125 subjects evaluated at the NIH Clinical Center with informed consent. The cohort included 35 patients with Parkinsonism (29 with GD and 6 carriers) and 40 without PD, but with a strong family history of Parkinsonism (25 with GD, 9 carriers and 4 non-carrier family members). WES data was analyzed using three different approaches. Rare probands with both GD and PD where parental DNA was available were evaluated as trios. In addition, sequence from nine sets of GD sib pairs where only one sib had PD were compared. Lastly, candidate genes implicated in lysosomal pathways and known PD genes were explored in patients with and without PD. Thus far, few interesting leads have uncovered in exonic and canonical splice variants in genes related to endolysosomal pathways, and these are currently being validated and will be discussed in meeting. However, few are present in multiple families, suggesting that many different PD risk allele or genetic modifiers will be involved. These subjects provide a rich cohort for the identification of further alleles that enhance or lessen the risk of developing Parkinsonism.

## PgmNr 2106: Association of expression quantitative trait loci in blood with ischemic stroke.

### Authors:

H. Amini<sup>1</sup>; N. Shroff<sup>1</sup>; P. Sitorus<sup>1</sup>; P. Carmona-Mora<sup>1</sup>; X. Zhan<sup>1</sup>; B. Stamova<sup>1</sup>; G. Jickling<sup>1,2</sup>; F. Sharp<sup>1</sup>; B. Ander<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology, School of Medicine, University of California at Davis, Sacramento, California.; 2) Department of Medicine, University of Alberta, Edmonton, Canada.

---

Several single nucleotide polymorphisms (SNPs) have been associated with ischemic stroke (IS). It is also known that expression of many genes changes after ischemic stroke. However, SNPs that are associated with changes of gene expression in blood after ischemic stroke are largely unknown. Thus, we studied the association of common genetic variants with changes in mRNA expression levels (i.e. expression quantitative trait loci; eQTL) in blood after ischemic stroke.

RNA and DNA were isolated from blood samples collected from 137 ischemic stroke patients and 138 vascular risk factor controls (VRFC). Gene expression of all protein-coding transcripts was quantified by Affymetrix HTA 2.0 microarrays and SNP variants assessed by Axiom Biobank Genotyping microarrays. In order to identify ischemic stroke diagnosis-dependent eQTL, a linear model with a genotype (SNP) × diagnosis (IS and VRFC) interaction term was fit for each SNP-gene pair.

We found several SNP-gene pairs associated with the magnitude of gene expression change after IS. Some of the strongest cis-eQTLs were rs56348411 related to expression of neurogranin (NRGN), rs78046578 related to expression of chemokine (C-X-C motif) ligand 10 (CXCL10) and rs975903 related to expression of SMAD family member 4 (SMAD4). CXCL10 is involved in blood-brain barrier (BBB) breakdown following stroke and SMAD4 is a cause of hereditary hemorrhagic telangiectasia syndrome and modulates N-cadherin expression in endothelial cells to stabilize the BBB. Regarding trans-eQTL, we found SNP rs148791848 significantly affected the expression of anosmin-1 (ANOS1) in a disease-specific manner. ANOS1 is key in developmental processes including neural cell adhesion and axonal migration and might play a similar role after stroke.

This study describes the importance of genetic variants in the expression of genes that relate to post-ischemic stroke injury and/or recovery. Understanding the mechanisms used by these eQTLs illustrates patient heterogeneity and could lead to the development of a novel strategy to treat stroke. Further work examining these relationships will help with establishing personalized treatment strategies to improve stroke outcome.

## PgmNr 2107: A new case of nonepileptic neurodevelopmental disorder due to a novel GRIN2A mutation.

### Authors:

P.S. Buonuomo <sup>1</sup>; G. Mastrogiorgio <sup>1</sup>; P. Alfieri <sup>2</sup>; A. Terracciano <sup>3</sup>; I. Rana <sup>1</sup>; M. Macchiaiolo <sup>1</sup>; M.V. Gonfiantini <sup>1</sup>; R. Capolino <sup>1</sup>; A. Bartuli <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Rare and Genetic Diseases, Bambino Gesù Children Hospital, Rome, Italy; 2) Pediatric Neuropsychiatry, Bambino Gesù Children's Hospital Rome, Italy; 3) Laboratory of Medical Genetics, Bambino Gesù Children's Hospital Rome, Italy

---

The GRIN2A gene codes for the GluN2A subunit of N-Methyl-D-aspartate receptors (NMDARs). Inherited and de novo GRIN2A mutations are usually associated with different forms of epilepsy from moderate to severe (i.e. Landau-Kleffner syndrome, continuous spike and wave during slow-wave sleep syndrome, atypical rolandic epilepsy), and speech impairment. Recently, some authors suggested that the phenotypic spectrum of GRIN2A-related disorders may also include neurodevelopmental and movement disorders without seizures.

Here we report the case of a 10-year-old girl coming from southern Italy born to non consanguineous, healthy parents after full-term and uneventful pregnancy by vaginal delivery. Her birth weight and length were 3980 g and 50 cm, respectively, and her Apgar score was 9/10. She displayed retarded psychomotor development, with delayed acquisition of language and motor milestones (sitting at 19 months, walking at 30 months, 2-word phrases at 4 years, and 3-word phrases at 6-7 years). The patient exhibited insomnia with frequent awakenings and stereotypic movements of upper limbs (right predominance) with mild problems in fine motor coordination.

However, she has never experienced a seizure, and repeated EEG recordings have revealed an absence of epileptiform abnormalities.

At 8 years of age, to establish a diagnosis, targeted next-generation sequencing analysis of a panel of genes associated with neuropediatric diseases was performed. The heterozygous variant c.2539 C>T, which promotes the amino acid change p.Arg847Ter, was identified in the GRIN2A gene. The variant was not present in either parents.

Our findings reinforce the association between GRIN2A mutations and neurodevelopmental and movement disorders but not necessarily associated with seizures and emphasize the importance of molecular analyses in undiagnosed patients.

## PgmNr 2108: Genetic basis for blood pressure associations with migraine.

### Authors:

Y. Guo<sup>1</sup>; P.M. Rist<sup>1,2</sup>; I. Daghlas<sup>2</sup>; F. Giulianini<sup>1</sup>; T. Kurth<sup>1,3</sup>; D.I. Chasman<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Preventive Medicine, Brigham and Women's Hospital, Boston, MA 02215; 2) Harvard Medical School, Boston, MA 02115; 3) Charité – Universitätsmedizin Berlin, Institute of Public Health, Charitéplatz 1, 10117 Berlin

**Background:** Blood pressure (BP) has been inconsistently associated with migraine and the biological mechanism of BP-lowering medications in migraine prophylaxis is not clear. We aimed to gain insight into mechanistic links between BP and migraine using genetic approaches.

**Methods:** We leveraged the most recent large-scale GWAS summary-statistics for migraine ( $N_{\text{cases}}/N_{\text{controls}}=59,674/316,078$ ) and BP ( $N=757,601$ ) and genetic methods to i) estimate global and local genetic correlation, ii) identify shared genetic components, iii) discover new migraine loci, and iv) infer the potential causality and directionality between BP and migraine.

**Findings:** There were positive overall genetic correlations of migraine with diastolic BP (DBP) ( $r_g=0.11$ ,  $P=3.56 \times 10^{-06}$ ) and systolic BP (SBP) ( $r_g=0.06$ ,  $P=0.01$ ), but not pulse pressure (PP) ( $r_g=-0.01$ ,  $P=0.75$ ). Local genetic correlations were experiment-wide significant at three known migraine loci (*FHL5*, *C7orf10*, and *PLCE1*) ( $P<0.05/1703$ ). Cross-trait meta-analysis revealed 14 shared loci novel to migraine ( $P_{\text{CPASSOC}} \leq 5 \times 10^{-08}$ ), nine of which were replicated in the UK Biobank ( $P<0.05$ ). Transcriptome-wide association studies (TWAS) identified 12 novel genes for migraine shared with one or more of the BP traits in expression data from artery, nerve, skin, esophagus mucosa, and whole blood. SNPs at five of these shared genes (*ITGB5*, *SMG6*, *ADRA2B*, *ANKDD1B*, and *KIAA0040*) were also significant in the cross-trait meta-analysis revealing important roles of vascular development, endothelial function, and calcium homeostasis as biological mechanisms shared by migraine and BP regulation. Specifically, *ADRA2B* associations highlight effects of glutamate, serotonin (5-HT), and calcitonin gene-related peptide (CGRP). Additionally, we found stronger genetic instrumental estimates (per 10mm Hg) of elevated DBP (OR [95%CI]=1.20 [1.15-1.25];  $P=5.57 \times 10^{-25}$ ) on migraine than elevated SBP (1.05 [1.03-1.07];  $P=2.60 \times 10^{-07}$ ) or decreased PP (1.09 [1.05-1.14];  $P=3.65 \times 10^{-07}$ ). Conditional instrumental analysis suggested that effects of elevated SBP on migraine may be explained by effects of DBP. By contrast gene-based influences of SBP were dominant to DBP for migraine comorbidities CAD and stroke in parallel analyses.

**Interpretation:** The findings prioritize a more prominent role of DBP than SBP in migraine susceptibility. Shared genetic components provide insights into biological mechanisms linking migraine and BP, and potentially into migraine treatment.

## PgmNr 2109: Genetic profile of a multiplex Hirschsprung disease family.

### Authors:

C.S. Tang <sup>1,2</sup>; M. Karim <sup>1</sup>; X. Zhuang <sup>1</sup>; M.T. So <sup>1</sup>; P.C. Sham <sup>3,4</sup>; M.M. Garcia-Barcelo <sup>1</sup>; P.K.H. Tam <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Surgery, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 2) The Dr Li Dak-Sum Research Centre, The University of Hong Kong – Karolinska Institutet Collaboration in Regenerative Medicine; 3) Center for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong; 4) Department of Psychiatry, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong

---

Hirschsprung disease (HSCR), also known as congenital megacolon, is characterized by the absence of enteric ganglia in the hindgut. HSCR is a rare multigenic neurocristopathy with receptor tyrosine kinase, *RET*, being the major HSCR gene. Increasing evidence from genetic studies suggests that a sensitized genetic background of reduced *RET* expression conferred by the regulatory HSCR-associated common variant is critical for the phenotypic expression of HSCR. We performed a whole-genome sequencing study of a multiplex HSCR family with two affected siblings, one with more severe long-segment HSCR (LHSCR) while another with the common form of short-segment HSCR (SHSCR). A truncating *de novo* mutation of *KIF26A* was identified in the LHSCR patient where knockout of *kif26a* in mice was shown to present with hyperganglionosis and megacolon. We hypothesized that *KIF26A de novo* mutation acts as a genetic modifier contributing to the increased severity and there exists other causal variants or genetic modifiers shared between the affected siblings for the phenotypic expression. Combined with the approach of polygenic risk score, we delineated the relationship between the sensitized genetic background of common variants and constellations of rare DNA changes, including copy number variations, with respect to risk of HSCR. Latest results of the genetic profile of the family will be presented.

## PgmNr 2110: A genome-wide association screen of quantitative reading and language traits in more than 22,000 people in the GenLang consortium.

### Authors:

E. Eising<sup>1</sup>; A. Allegrini<sup>2</sup>; C. Barr<sup>3</sup>; E. van Bergen<sup>4</sup>; D.I. Boomsma<sup>4</sup>; M. van Donkelaar<sup>1</sup>; M.C.J.P. Franken<sup>5</sup>; J. Gruen<sup>6</sup>; P.R. Jansen<sup>7</sup>; P.F. de Jong<sup>8</sup>; M. Luciano<sup>9</sup>; D. Newbury<sup>10</sup>; M.M. Nöthen<sup>11</sup>; S. Paracchini<sup>12</sup>; K. Price<sup>3</sup>; K. Rimfeld<sup>2</sup>; G. Schulte-Körne<sup>13</sup>; C.Y. Shapland<sup>1</sup>; N. Simpson<sup>14</sup>; S. Smith<sup>15</sup>; N. Timpson<sup>16</sup>; J.B. Tomblin<sup>17</sup>; D.T. Truong<sup>6</sup>; C. Wang<sup>18</sup>; A. Whitehouse<sup>19</sup>; E.L. de Zeeuw<sup>20</sup>; G. Zhu<sup>21</sup>; B. St Pourcain<sup>1,16,22</sup>; C. Francks<sup>1,22</sup>; S.E. Fisher<sup>1,22</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Language and Genetics, Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; 2) Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, London, UK; 3) Genetics and Development Division, Krembil Research Institute, The Hospital for Sick Children, Toronto, Canada; 4) Department of Biological Psychology, VU University, Amsterdam, The Netherlands; 5) Department of Otorhinolaryngology, Sophia Children's Hospital, Erasmus University Medical Center, Rotterdam, the Netherlands; 6) Departments of Pediatrics, Genetics, and the Investigative Medicine Program, Yale School of Medicine, New Haven, USA; 7) Department of Complex Trait Genetics, CNCR, VU University, Amsterdam, the Netherlands; 8) Research Institute of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands; 9) School of Philosophy, Psychology and Language Sciences, University of Edinburgh, Edinburgh, UK; 10) Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK; 11) Institute of Human Genetics, University of Bonn, Bonn, Germany; 12) School of Medicine, University of St Andrews, St Andrews, UK; 13) Department of Child and Adolescent Psychiatry, Psychosomatic, and Psychotherapy, Ludwig-Maximilians University, Munich, Germany; 14) Department of Experimental Psychology, University of Oxford, Oxford, UK; 15) Pediatrics, University of Nebraska Medical Center, Omaha, Nebraska, USA; 16) Medical Research Council Integrative Epidemiology Unit, Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.; 17) College of Liberal Arts and Sciences, The University of Iowa, Iowa City, USA; 18) School of Medicine and Public Health, Faculty of Medicine and Health, University of Newcastle, Newcastle, Australia; 19) Telethon Kids Institute, University of Western Australia, Subiaco, Australia; 20) Amsterdam Public Health research institute, VU University, Amsterdam, the Netherlands; 21) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 22) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

---

**Background:** The processing and production of complex spoken and written language are traits that are unique to humans. Decades of family and twin studies established that these traits are highly heritable. Nonetheless, the relevant genetic architecture is complex, heterogeneous, and multifactorial, and yet to be explored in well-powered studies. The GenLang consortium is facilitating larger studies on the genetic underpinnings of spoken and written language.

**Methods:** Here, we performed a genome-wide association study (GWAS) for quantitative reading- and language-relevant traits: word reading, spelling, nonword reading, phoneme awareness, nonword

repetition and rapid automatized naming. The phenotype data were collected across different languages (primarily English, and also Dutch, Spanish, German, Finnish, French and Hungarian) and ages (range 5 to 18 years) in 17 GenLang cohorts. Uniform pipelines were established for consistent quality control and association analysis across cohorts. Univariate, inverse variance-weighted meta-analyses for each trait were performed in METAL, and multivariate analyses with N-weighted GWAMA. Functional mapping was performed in FUMA, and genetic correlation analyses with LDSC.

**Results:** We reached sample sizes of more than 6,000 to 22,000 individuals. A single genome-wide significant locus was discovered for word reading, spelling and phoneme awareness. A multivariate meta-analysis of the four highly correlated traits: word reading, nonword reading, spelling and phoneme awareness, identified further significant loci. Initial functional enrichment analyses revealed an enrichment in brain-expressed genes.

**Discussion:** These efforts by GenLang discovered multiple genome-wide significant loci for reading and language traits, and will open up novel avenues for deciphering the biological underpinnings of spoken and written language. Language-related phenotypes pose special challenges for genetic analysis, due to the labour-intensive nature of assessment and additional heterogeneity caused by different assessment tools and languages. Therefore they have yet to fully benefit from advances in large-scale genomic screens.

## PgmNr 2111: Febrile seizures GWAS identifies 7 new loci implicating genes involved in fever response and neuronal excitability.

### Authors:

B. Feenstra<sup>1</sup>; L. Skotte<sup>1</sup>; J. Fadista<sup>1</sup>; J. Bybjerg-Grauholm<sup>2,3</sup>; V. Appadurai<sup>3,4</sup>; M.S. Hildebrand<sup>5</sup>; T.F. Hansen<sup>3,4,6</sup>; K. Banasik<sup>7</sup>; X. Liu<sup>1</sup>; B. Pasternak<sup>1,8</sup>; T.H. Pers<sup>1,9</sup>; J.W. Dreier<sup>10</sup>; J. Christensen<sup>10,11</sup>; D.M. Hougaard<sup>2,3</sup>; A. Buil<sup>3,4</sup>; F. Geller<sup>1</sup>; A. Hviid<sup>1</sup>; H. Ullum<sup>12</sup>; S.F. Berkovic<sup>5</sup>; T.M. Werge<sup>3,4</sup>; M. Melbye<sup>1,13</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Danish Centre for Neonatal Screening, Department of Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 3) iPSYCH, Lundbeck Foundation Initiative for Integrative Psychiatric Research, Denmark; 4) Mental Health Center Sct. Hans, Mental Health Services, Capital Region Denmark, Roskilde, Denmark; 5) Epilepsy Research Centre, Department of Medicine, University of Melbourne (Austin Health), Victoria, Australia; 6) Danish Headache Center, Department of Neurology, Rigshospitalet-Glostrup, Denmark; 7) Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark; 8) Clinical Epidemiology Division, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden; 9) Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen, Denmark; 10) National Centre for Register-Based Research, Department of Economics and Business Economics, Aarhus BSS, Aarhus University, Denmark; 11) Department of Neurology, Aarhus University Hospital, Aarhus, Denmark; 12) Department of Clinical Immunology, Copenhagen University Hospital, Rigshospitalet, Denmark; 13) Department of Medicine, Stanford University School of Medicine, Stanford, California, USA

**Introduction:** Febrile seizures affect 2-5% of children in western countries including Denmark with an incidence peak around 16 months of age. While usually benign, in a minority of cases febrile seizures precede development of epilepsy.

**Methods:** We conducted a GWAS meta-analysis of febrile seizures with a total of 4,304 cases and 48,290 controls in the discovery stage, all sampled from the Danish National Biobank. The analysis was based on 8.5 million variants imputed from the 1000 genomes project phase III reference panel, using logistic regression on imputed SNP dosages for association testing. Replication testing was carried out in independent Danish and Australian cohorts (3,235 cases, 40,519 controls).

**Results:** Seven novel robustly associated loci (replication  $P < 0.05$ , combined  $P < 5 \times 10^{-8}$ ) were detected harboring genes related to fever response (*PTGER3*, OR=1.14,  $P=2.2 \times 10^{-11}$ ; *IL10*, OR=1.14,  $P=3.1 \times 10^{-8}$ ), GABA-signaling (*GABRG2*, OR=1.28,  $P=1.9 \times 10^{-15}$ ), neurotransmitter release (*BSN*, OR=1.15,  $P=4.9 \times 10^{-12}$ ; *ERC2*, OR=1.30,  $P=1.8 \times 10^{-12}$ ; *HERC1*, OR=1.20,  $P=3.3 \times 10^{-12}$ ), and MAP kinase signaling (*MAP3K9*, OR=1.13,  $P=1.5 \times 10^{-10}$ ). Fine-mapping and analysis of exome sequencing data at these loci is ongoing. The four previously known febrile seizures loci (*SCN1A*, *SCN2A*, *ANO3*, 12q21.33) were also confirmed, and collectively the 11 loci explained 3.3% of the variation in liability to febrile seizures. Subsequent genome-wide enrichment analysis using MAGMA suggested that presynaptic cytomatrix proteins are implicated in seizure etiology ( $P = 0.039$ ). Looking at variants



with a minor allele frequency  $>1\%$  across the genome and comparing with GWAS summary statistics from the International League Against Epilepsy Consortium on Complex Epilepsies, we found positive genetic correlations with focal epilepsy ( $r_g = 0.44$ ,  $P = 0.04$ ), genetic generalized epilepsy ( $r_g = 0.17$ ,  $P = 0.03$ ), and all epilepsies combined ( $r_g = 0.31$ ,  $P = 0.002$ ).

**Conclusion:** Our study identified novel genetic associations with febrile seizures implicating genes involved in neuronal excitability and fever response. We also found positive genetic correlations between febrile seizures and complex epilepsies.

## **PgmNr 2112: A genome-wide meta-analysis of migraine with over 102,000 cases identifies 124 risk loci and provides first genetic insights to new migraine therapeutics targeting CGRP pathway.**

### **Authors:**

H. Hautakangas<sup>1</sup>; P. Gormley<sup>2,3</sup>; C. Benner<sup>1</sup>; B.S. Winsvold<sup>4,5</sup>; A. Auton<sup>6</sup>; N. Litterman<sup>6</sup>; S.E. Ruotsalainen<sup>1</sup>; E. Widén<sup>1</sup>; S. Ripatti<sup>1,7,8</sup>; J.A. Zwart<sup>4,5,9</sup>; A. Palotie<sup>1,8,10,11</sup>; M. Pirinen<sup>1,7,12</sup>; the HUNT Study, the 23andMe Research Team and the International Headache Genetics Consortium (IHGC)

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute for Molecular Medicine Finland FIMM, HiLIFE, University of Helsinki, Helsinki, Finland; 2) Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) The Stanley Center for Psychiatric Research, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Department of Research, Innovation and Education, Division of Clinical Neuroscience, Oslo University Hospital, Oslo, Norway; 5) K. G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 6) 23andMe, Inc., Mountain View, CA, USA; 7) Department of Public Health, University of Helsinki, Helsinki, Finland; 8) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 9) Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Norway; 10) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 11) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 12) Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

---

Migraine is a common brain disorder typically characterized by disabling episodes of severe headache accompanied by nausea and hypersensitivity to light, sound and smell. Its lifetime prevalence is 15-20% worldwide and family and twin studies estimate a heritability of about 40%. To date, 41 migraine susceptibility loci have been identified from the genome-wide association studies (GWAS), and these loci are suggested to be enriched for genes highly expressed in vascular and smooth muscle tissues.

In this meta-analysis, we increase the largest published migraine meta-analysis (Gormley et al. Nat Gen 2016) sample size by 71% by adding to it 22,644 new cases from 23andMe, 10,881 cases from the UK Biobank, 7,801 cases from the HUNT Study and 1,084 cases from the GeneRISK Study. This takes our total sample size to 102,084 cases and 771,257 controls.

In a fixed-effects meta-analysis, we identify 124 independent ( $r^2 < 0.1$ ) genome-wide significant ( $P < 5 \times 10^{-8}$ ) autosomal single-nucleotide polymorphisms (SNPs) that are at least 250kb apart from each other. Of these 124 loci, 91 are reported here for the first time. We report enrichment of GWAS signal in 5 central nervous system and 3 cardiovascular cell types, and in single cell types of digestive system, musculoskeletal/connective tissue and ovary (FDR 5% from LDSC-SEG).

Among the 91 new loci, we report the first evidence from GWAS at *CALCA-CALCB* locus (lead SNP

rs1003194,  $P=2.43 \times 10^{-10}$ ) encoding calcitonin gene-related peptide (CGRP). Recently approved monoclonal antibodies (mAbs) targeting CGRP or its receptor are a first major breakthrough for the migraine specific treatments since the development of the triptans over two decades ago. This indicates that also other drugs targeting genes from our GWAS could be excellent candidates for new ways to treat migraine. While also a mAb (erenumab) targeting the CGRP receptor has proven efficient against migraine, we report that none of the central genes coding the receptor proteins (*CALCRL*, *RAMP1* or *RCP*) show a clear association to migraine ( $P > 10^{-4}$ ).

To conclude, we report the largest migraine GWAS to date with 91 new susceptibility loci. The enrichment of association signal in both central nervous system and cardiovascular system reflects the heterogeneous genetic background of migraine. The first genetic evidence for CGRP association in migraine provides a new route to further develop the recent breakthroughs in migraine therapeutics that target the CGRP pathway.

## PgmNr 2113: Whole genome sequencing analysis of gastroparesis.

### Authors:

M. Keefe; J. Carlin; S. Smieszek; C. Xiao; C. Polymeropoulos; G. Birznieks; M. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals Inc., 2200 Pennsylvania Ave, Suite 300E, Washington, D.C. 20002

Gastroparesis is a disorder characterized by delayed gastric emptying of solid food in the absence of a mechanical obstruction of the stomach, resulting in the cardinal symptoms of early satiety, postprandial fullness, nausea, vomiting, belching and bloating. To ascertain the genetic risk factors for gastroparesis we conducted the first to date whole genome sequencing (WGS) study of gastroparesis cohort. We investigated the frequency and effect of rare loss-of-function (LOF) variants in patients with idiopathic and diabetic gastroparesis enrolled in a clinical study, VLY686-2301. The dataset consisted of 119 WGS samples.

Among rare LOF variants, we report an increased frequency of a frameshift mutation within Motilin Receptor (MLNR) gene, variant rs562138828. Motilin is a 22 amino acid peptide hormone expressed throughout the gastrointestinal (GI) tract. The protein encoded by this gene is a motilin receptor which is a member of the G-protein coupled receptor 1 family. We have shown an increased frequency of a frameshift mutation with MAF 0.01 as compared 0.0009 AF in GNOMAD ( $P$ -value = 0.01) with Odds Ratio of 21.9. We detect 4/119 Gastroparesis patients carry the variant of interest that results in p.Leu202ArgfsTer105. Noteworthy is the fact that the 4 cases were equally split among idiopathic and diabetic, so this possible gastroparesis risk factor appears to be agnostic as to condition. The finding may be of direct relevance to treatment as individuals with the identified mutation may respond differently to gastroparesis treatments especially those targeting MLNR. Among other rare LOF, we identified a case of CHD7 discussed in literature in gastro-related context and a case of CFTR duodenal stenosis pathogenic variant. The CFTR variant has been seen in pancreatitis and is likely causative of Cystic Fibrosis Gut which is characterized by increased mucous viscosity and development of intestinal inflammation, dysbiosis and dysmotility.

Whole genome sequencing of Gastroparesis patient samples showed enrichment for rare variants in the MTLR in cases compared with controls. The identified LOF variants within the region can serve as risk factor for disease as well as inform treatments, especially given the knowledge of different response to treatment.

## PgmNr 2114: Automated genetic and functional fine-mapping of Parkinson's disease loci.

### Authors:

B. Schilder; T. Raj

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Neuroscience, Department of Genetics & Genomic Sciences, Ronald M Loeb Center for Alzheimer's disease Icahn School of Medicine at Mount Sinai 1 Gustave L. Levy Pl, New York, NY 10029

---

A recent large-scale GWAS of Parkinson's Disease (PD) by Nalls et al. (2019, BioRxiv) identified 78 independent loci containing PD-associated SNPs. However, for any given locus the SNP with the smallest p-value may merely be a proxy for the true causal SNP (or set of causal SNPs) due to linkage disequilibrium (LD). Fine-mapping aims to identify the causal SNP(s), which can greatly reduce the number of false-positive genetic associations and thus accelerate the identification of mechanisms causal for the disease. We fine-mapped all 78 loci by developing echolocatoR, an open-access R package that enables automated end-to-end statistical and functional fine-mapping in a single line of code. By integrating multiple fine-mapping tools (FINEMAP and SuSiE), we reduced the average number of candidate SNPs from ~5,000 to 8 per locus. In 56% of loci, multi-tool consensus converged on 1-4 SNPs (i.e. consensus SNPs) that had the highest posterior probability of being causal. In 71% of loci the lead SNP was within the credible set, suggesting this approach does more than simply recapitulate the SNP with the lowest GWAS p-value. Specifically, within the *LRRK2* locus, rs7294619 was the most probable causal SNP for PD. Using eQTL data from monocytes, we found that rs7294619 has cis-QTLs in the *IFN* ( $\beta=0.917$ , FDR<1.34e-05) and *LPS* 24-hour ( $\beta=0.983$ , FDR<1.34e-05) stimulated data, but not in the *CD14* or *LPS* 2-hour, suggesting context-dependent regulatory effects. Within the *BST1* locus, 6,227 SNPs were refined to a credible set of 9, and a single consensus SNP (rs4613561). Both consensus SNPs in the *BST1* and *LRRK2* loci are located within monocyte enhancers (HaploReg) and have histone marks in blood and brain (ROADMAP). As both *LRRK2* and *BST1* are involved in innate immune function, these results further support the hypothesis that genetic dysregulation of the immune system is an important causal factor in PD risk. We also identified potentially causal missense variants at in the *TRIM40* (rs1265096, rs2233952), *RPS12* (rs41286192), and *CD19* (rs7140) loci. We are now validating these findings through Bayesian epigenomic data integration and functional experiments (genome editing, allele-specific protein binding and differential enhancer assays). Overall, these results demonstrate that echolocatoR robustly and rapidly identifies functionally relevant variants that have a statistically higher probability of being causal for PD and thus targets of therapeutics development.

## PgmNr 2115: Characterization to the genomic landscape in SMS with whole genome sequencing.

### Authors:

M.A. Fisher; E.L. Czeisler; S.P. Smieszek; C. Xiao; C.M. Polymeropoulos; G. Birznieks; M.H. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals Inc., Washington, DC

---

Smith-Magenis syndrome (SMS; OMIM #182290) is a rare complex genomic disorder resulting from haploinsufficiency of *RAI1*. It is associated with decreased total nighttime sleep for age, frequent and prolonged nocturnal awakenings, early morning sleep offset, and excessive daytime sleepiness.

We conducted whole genome sequencing analysis to explore the genomic landscape of the SMS population with confirmed *RAI1* aberration. We investigated 41 individuals with confirmed diagnosis of SMS. We detect 1x stopgain (novel) 727723, *RAI1*:NM\_030665:exon3:c.C2878T:p.R960X. 39 patients have a haploinsufficiency detected as stretches of ROH across whole coding sequence of *RAI1*, missense patient.

*RAI1* is a dosage-sensitive gene expressed in many tissues and highly conserved among species. Many studies have demonstrated that *RAI1* and its homologs acts as a transcriptional factor implicated in embryonic neurodevelopment, neuronal differentiation as well as behavioral functions and circadian activity. Patients with *RAI1* pathogenic variants show some phenotypic differences when compared to those carrying the typical deletion. This deletion encompasses *RAI1*, whose haploinsufficiency is considered the primary cause for most of the SMS features including aberrant clock activity via its effect on *CLOCK* (ChIP-Chip and luciferase data suggest that *RAI1* binds, directly or in a complex, to the 1st intron of *CLOCK* enhancing transcriptional activity).

Interestingly since *COPS3* is located within the common deletion interval, most individuals with SMS are haploinsufficient for *COPS3*. Within the core SMS region on chromosome 17, the gene *COPS3* encodes subunit 3 of the COP9 related to the 26S proteasome regulatory complex. The latter has been associated with control of the rate-limiting step in melatonin metabolism by N-acetyltransferase. Altogether, 34 cases have *COP3* haploinsufficiency, 2 have LOF variants, and 2 do not. It may be that the joint *COPS3* haploinsufficiency is more tightly linked to the clinical phenotype of disrupted sleep in SMS patients.

In addition, we have conducted a GWAS on SMS patients and a set of matched controls. MDGA2 MDGA2:NM\_001113498:exon4:c.C502A:p.R168R, variants on chromosome 17, HDAC7 and PLEKHA7 (10-8) though these may rather be risk factors, given the incidence of the disorder. Investigation of the landscape not only reveals the fundamental cause and the pathophysiology but helps delineate risk factors and comorbidities in this already complex disorder.

## PgmNr 2116: Integration of genomic and gene expression profiles from normal appearing white matter tissue in multiple sclerosis reveals novel candidate genes and inferences of pathogenesis.

### Authors:

A.M. Manuel <sup>1</sup>; Y. Dai <sup>1</sup>; L.A. Freeman <sup>2</sup>; P. Jia <sup>1</sup>; Z. Zhao <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, Texas, USA; 2) Department of Neurology, Dell Medical School, The University of Texas at Austin, Austin, Texas, USA; 3) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA; 4) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA

---

Multiple sclerosis (MS) is a neuroinflammatory disorder characterized by demyelination and tissue degeneration of the central nervous system. Pathological changes taking place in the normal appearing white matter (NAWM) remain poorly understood. Genome-wide association studies (GWAS) have detected genetic loci associated to MS. Disease-related tissue expression data can aid in better interpreting the nature of disease-associated variants. In this study, a recent GWAS of 68,379 MS cases and controls was integrated with gene expression profiles derived from MS NAWM post-mortem brain tissue. Tissue samples collectively included 38 MS NAWM cases and matched controls from two datasets: RNA sequencing reads of NAWM microglia and microarray profiles of active perilesional NAWM. These datasets were independently studied in network-based analyses considering the human protein interactome as the reference network. We integrated MS GWAS summary statistics (node weight) with MS NAWM tissue co-expression (edge weight) and searched signal-enriched modules with our in-house tool: Edge Weighted Dense Module Search of GWAS (EW\_dmGWAS). Among the top 100 modules we identified, surprisingly, we found several MS FDA drug target genes: *HDAC1*, *IFNRA1*, *KEAP1* and *RELA*, which were not reported by original GWAS or gene expression studies. A subsequent dual evaluation anchored by the MS GWAS considered microglia NAWM profiles as the discovery set and perilesional NAWM profiles as the evaluation set. It yielded 19 resulting modules containing 55 non-redundant genes. This gene set was enriched in the Gene Ontology Biological Process terms “Fc-gamma receptor signaling pathway involved in phagocytosis” (adjusted p-value =  $6.87 \times 10^{-3}$ ) and “intracellular transport of virus” (adjusted p-value = 0.02), among other relevant functional terms. We pinpointed the genes *CCNB2* (p-value =  $4.08 \times 10^{-3}$ ), *CDK3* (p-value =  $6.56 \times 10^{-4}$ ), *HSPA1L* (p-value =  $1.58 \times 10^{-14}$ ), *MORF4L1* (p-value =  $5.10 \times 10^{-5}$ ), *TEC* (p-value =  $2.11 \times 10^{-6}$ ) and *VAV1* (p-value =  $7.21 \times 10^{-4}$ ) as novel MS candidate genes for their presence in top MS NAWM gene networks. Genes enriched for phagocytic activity give insight to MS pathogenesis in NAWM. Furthermore, functional terms related to virus transport pathways supported the involvement of viral infection in MS etiology. In summary, the results from our novel network analysis provided better understanding of the mechanisms underlying MS and several novel genes that warrant further investigation.

## PgmNr 2117: Responder analysis to melatonin agonist, in a whole genome sequencing study of major depression disorder.

### Authors:

M. Mohrman; S. Smieszek; D. Xiao; C. Polymeropoulos; G. Birznieks; M. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals, Washington, DC.

---

Major Depression Disorder (MDD) can severely impact patients in terms of excess morbidity and mortality, disability and diminished functioning. Circadian rhythm disruption may be involved in the etiology of MDD. Interestingly, differences were previously observed in the expression of circadian timing system between African Americans (AA) and Caucasians. Melatonin agonist Tasimelteon 20 mg was shown to improve symptoms of depression in African American patients with MDD. This may further reiterate a circadian component in the etiology and treatment of Major Depression.

We have conducted a large whole genome sequencing study of MDD to examine the genetic underpinnings of those differences. Participants included patients between the ages of 18-65 with primary diagnosis of MDD. The cohort included 166 AA patients assessed with Hamilton Depression Scale. Timing of the onset of melatonin was measured (the pineal melatonin rhythm is considered the most reliable marker of the circadian timing system). To ascertain the genetic underpinnings of response to melatonin agonist, we conducted whole genome sequencing association analysis using responders and non-responders in the AA population.

We detect a significant association with a NTF3 variant rs11063714 (MAF 0.10) 19/2 and OR of 15.6. The allele is a marker of higher response (AA). The protein encoded by this gene is a member of the neurotrophin family that controls survival and differentiation of mammalian neurons. NTF3 variants have been previously associated with insomnia in a meta-analysis study of AA insomnia. The variant is located downstream in a highly regulatory region of the gene as assessed with Haploreg (ENCODE). Of potential interest, we identify an association with variants in ANO2, which is immediately adjacent to NTF3 that harbors an insomnia associated SNP among AA subjects.

Clinically, at 8 weeks AA MDD patients treated with tasimelteon improved by 9.9 points on the HAMD scale as compared to 6.9 points for the placebo treated patients (p-value =0.018).

In a responder analysis, 59% of tasimelteon treated AA patients improved as compared to 30% of placebo treated AA patients (p-value =0.0019). Interestingly previous data have shown that NTF3 rapidly modulates the activity of NPO neurons involved in REM sleep and that cholinergic neurons in the LDT and PPT contain NTF3. Taken together these results support the hypothesis that NTF3 may be involved in the control of naturally occurring REM sleep.



## PgmNr 2118: Longitudinal positron emission tomography of dopamine synthesis in subjects with *GBA1* mutations.

### Authors:

E. Sidransky <sup>1</sup>; G. Lopez <sup>1</sup>; M. Gregory <sup>2</sup>; D. Eisenberg <sup>2</sup>; J. Kim <sup>1</sup>; C. Groden <sup>1</sup>; G. Monestime <sup>1</sup>; K. Berman <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Neurogenetics Section, NHGRI, NIH Bethesda MD; 2) Clinical and Translational Neuroscience Branch, NIMH, NIH, Bethesda, MD

Mutations in *GBA1*, the gene mutated in Gaucher disease (GD), are the most common known genetic risk factor for Parkinson disease (PD) and dementia with Lewy bodies. However, the majority of *GBA1* mutation carriers never develop parkinsonism. To test for early signs of PD in this at-risk population, we performed <sup>18</sup>F-fluorodopa PET imaging on *GBA1* mutation carriers, with and without PD, to identify patients with evolving loss of dopamine prior to the development of overt symptoms. While we previously published PET results in a smaller cohort, we have extended the study to include more participants, many of whom were studied longitudinally, representing 117 scans performed on subjects with *GBA1* mutations, including 41 patients with GD and 20 *GBA1* heterozygotes, as well as 98 controls, all seen at the Clinical Center of the National Institutes of Health between 2006-2016.

PET studies were performed on a GE Advance 3D scanner, and presynaptic dopamine synthesis was measured with <sup>18</sup>F-fluorodopa. Drugs that might impact the results were tapered appropriately. Sixty minutes before scanning, participants received carbidopa (200 mg) to reduce peripheral <sup>18</sup>F-fluorodopa metabolism and increase tracer availability in the brain. A bolus dose of <sup>18</sup>F-fluorodopa of 15±2 mCi was administered intravenously and 27 frames collected over 90 minutes. Longitudinal studies were performed on 33 subjects with *GBA1* mutations and 15 controls.

Seventeen patients with clinical evidence of parkinsonism showed diminished striatal dopamine synthesis, particularly in the putamen, as expected. Six had 2-3 subsequent scans. The study also followed 43 patients (27 with GD and 16 *GBA1* carriers) who did not show clinical evidence of parkinsonism at their first evaluation, including 38 with a strong family history of PD. Sixteen had one scan, but 27 were followed longitudinally. Twelve had two scans, 12 completed three scans and 3 completed four scans, with an interval between 1-7 years between scans, and clinical follow-up has continued. Only one of the 43 has subsequently been diagnosed with PD, and his sole PET study, performed 12 months prior to symptom onset appeared normal.

The analysis revealed that over time, as a whole, the PET studies in patients without PD did not show a marked decline in striatal dopamine synthesis capacity and were comparable to the results seen in controls. Thus, close clinical monitoring may be as effective as costly <sup>18</sup>F-fluorodopa imaging for this at-risk group

## PgmNr 2119: Detection of somatic CNVs in SNCA and genome-wide in the brains of synucleinopathies (multiple system atrophy and Parkinson's disease).

### Authors:

C. Proukakis<sup>1</sup>; M. Kalyva<sup>1</sup>; M. Leija-Salazar<sup>1</sup>; T. Lashley<sup>1</sup>; J. Holton<sup>1</sup>; S. Gentleman<sup>2</sup>; A. Schapira<sup>1</sup>; Z. Jaunmuktane<sup>1</sup>; D. Perez-Rodriguez<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Queen Square Institute of Neurology, University College London, London, England, United Kingdom;  
2) Imperial College of Medicine, London, England, United Kingdom

---

Parkinson's disease (PD) and multiple system atrophy (MSA) are synucleinopathies, progressive neurodegenerative disorders of partly unexplained aetiology. Somatic mutations can be found in neurons. We recently used FISH to detect somatic *SNCA* (alpha-synuclein) CNVs in substantia nigra, where they were commoner in dopaminergic (DA) neurons in disease. To further investigate somatic CNVs, focusing mainly on MSA, we performed additional FISH for *SNCA*, combined in selected experiments with markers for oligodendrocytes, neurons and alpha-synuclein, and low coverage single-cell whole genome sequencing (scWGS) for CNVs.

1) **FISH:** We analysed the nigra in 15 MSA cases (10 striatonigral- SND, 5 mixed). *SNCA* gains were present in 3.3% of DA neurons, and 3.1% of other cells, notably 9.4% of oligodendrocytes in 3 cases. CNVs and inclusions often occurred in the same cells in a disease-relevant pattern: DA neurons in Lewy body (LB) cases (1 PD, 4 incidental;  $p=0.004$ ), and other cells in SND MSA ( $n=7$ ;  $p=0.001$ ), mostly oligodendrocytes in 2 of 3 cases. Most cells with CNVs did not, however, have detectable inclusions, and vice versa.

In the cingulate cortex (12 PD, 6 MSA, 9 controls to date), there was a clear difference in the fraction of cells with CNVs between all groups / cell types ( $p=0.001$ ), with higher levels in neurons (MSA 4.3%, PD 2.8%, controls 1.4%).

We investigated possible correlations of the fractions of cells with CNVs. In the nigra, there was a possible negative correlation in non-DA cells with onset age in SND MSA ( $r=-0.6$ ,  $p=0.07$ ). In the cingulate, there was a clear correlation with age of death in non-neuronal cells in controls ( $r=0.9$ ,  $p=0.002$ ).

2) **scWGS:** We performed nuclear isolation and whole genome amplification by Picoplex Gold followed by scWGS. We obtained successful results so far from 9 neurons and 12 other cells from the pons of a case with MSA, with sequencing of an additional 250 cells from MSA brain underway. Among 5 neurons with nuclear a-syn inclusions, 3 had somatic gains, one of which was in the *SNCA* vicinity. Apparent neuronal multiple chromosomal losses were also seen, and gains in a non-neuron.

**Conclusion:** We demonstrate that somatic *SNCA* CNVs are commoner in MSA, and also occur in the cingulate. *SNCA* CNVs and inclusions often occur together, in a cell-type specific pattern, suggesting a functional role, supported by the possible negative correlation with age of onset in SND MSA. Ongoing scWGS will provide more details on CNVs detected.

## PgmNr 2120: *TLR* gene polymorphisms and susceptibility to Guillain-Barré syndrome: A study from India.

### Authors:

D. Dutta <sup>1</sup>; M. Debnath <sup>1</sup>; M. Nagappa <sup>2</sup>; P. Sundaravadivel <sup>1</sup>; S.K. Joshi <sup>1</sup>; S.K. Das <sup>3</sup>; R. Wahatule <sup>2</sup>; S. Sinha <sup>2</sup>; S. Periyavan <sup>4</sup>; G.S.U. Rao <sup>5</sup>; V. Ravi <sup>6</sup>; A.B. Taly <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, National Institute of Mental Health & Neuro Sciences (NIMHANS), Bengaluru, Karnataka, India; 2) Department of Neurology, National Institute of Mental Health & Neuro Sciences (NIMHANS), Bengaluru, Karnataka, India; 3) Department of Biostatistics, National Institute of Mental Health & Neuro Sciences (NIMHANS), Bengaluru, Karnataka, India; 4) Department of Transfusion Medicine, National Institute of Mental Health & Neuro Sciences (NIMHANS), Bengaluru, Karnataka, India; 5) Department of Neuroanaesthesia, National Institute of Mental Health & Neuro Sciences (NIMHANS), Bengaluru, Karnataka, India; 6) Department of Neurovirology, National Institute of Mental Health & Neuro Sciences (NIMHANS), Bengaluru, Karnataka, India

Guillain Barre Syndrome (GBS), the commonest cause of acute neuromuscular paralysis, is a post-infectious autoimmune disorder of peripheral nerves. The innate immune-molecule Toll-like receptors (TLRs) have emerged as an important mediator of inflammatory and autoimmune diseases including GBS. The impact of genetic variations within *TLR* genes on the risk and clinical phenotypes of GBS is inadequately known. Therefore, we investigated the role of *TLR* gene polymorphisms in GBS.

A total of 150 patients with GBS and 150 healthy subjects were assessed for Insertion/Deletion and single nucleotide polymorphisms (SNPs) within *TLR2* (rs3804099 and rs111200466), *TLR3* (rs3775290 and rs3775291) and *TLR4* (rs1927911 and rs11536891) genes by the polymerase chain reaction and restriction fragment-length polymorphism (PCR-RFLP). Haplotype analysis was performed by SHEsis, a web-based interface. Interactions among variants were determined by multifactor dimensionality reduction (MDR), a machine learning tool. Influence of genotypes on clinical features was analyzed using SPSS 20.

The most significant finding was 24-fold increased risk ( $p < 0.001$ ) to GBS in individuals with Insertion/Insertion (I/I) genotype of rs111200466 of *TLR2* gene. Specific genotypes of rs3804099, rs1927911 and rs11536891 were also associated with the risk of GBS. However, polymorphisms of the *TLR3* gene were not associated with GBS. The haplotype TC ( $p = 0.0001$ ) of *TLR4* SNPs confers 2.5-fold risk for GBS compared to controls. Specific haplotypes of *TLR2*, *TLR3* and *TLR4* variants have also imposed the risk of GBS. Two-way interaction between rs111200466 of *TLR2* and rs1927911 of *TLR4* selected as the best predictor of disease, having the highest average testing accuracy (0.67) and CVC value (4/5). The interaction dendrogram indicates a synergistic effect among rs111200166 and rs3804099 SNPs of *TLR2* and rs1927911 of *TLR4*. Patients with demyelinating subtypes with TT homozygosity had significantly higher MRC sum scores at admission ( $p = 0.0285$ ) and higher creatine kinase (CK) level ( $P = 0.0137$ ) compared with heterozygous CT genotype of rs3804099.

In conclusion, *TLR2* polymorphisms could influence genetic susceptibility to GBS and can emerge as a potential genetic marker of GBS. Leverage in CK level in demyelinating subtypes with TT homozygosity in rs3804099 is intriguing. Further studies in ethnically diverse populations may provide more insights on the role of *TLR* gene in pathogenesis and pathophenotypes of GBS.

## PgmNr 2121: Unbiased genome-wide screen identifies new ALS risk variants within gene-regulatory elements.

### Authors:

J. Cooper-Knock <sup>1</sup>; T. Moll <sup>1</sup>; S. Allen <sup>1</sup>; S. Zhang <sup>2</sup>; K. Kenna <sup>3</sup>; E. Hornstein <sup>4</sup>; M. Snyder <sup>2</sup>; P. Shaw <sup>1</sup>;  
Project MinE ALS Sequencing Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, United Kingdom; 2) Department of Genetics, Center for Genomics and Personalized Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA; 3) Brain Center Rudolf Magnus, Univ Med Ctr Utrecht, Utrecht, The Netherlands; 4) Department of Molecular Genetics, Weizmann Institute of Sciences, Rehovot, Israel

---

**Background:** Amyotrophic lateral sclerosis (ALS) is a universally fatal and relatively common neurodegenerative disease; lifetime risk is 1/350 in the UK. Currently there is no effective disease modifying therapy. 10% of ALS is autosomal dominant, but even in sporadic ALS observed broad-sense heritability is 61%. Currently, described genetic risk factors for ALS are found in <10% patients. Genome-wide association studies suggest the majority of missing ALS heritability is distributed throughout non-coding chromosomal regions. Enhancers are cis-acting DNA sequences that modulate expression of target genes primarily through binding of transcription factors.

**Methods:** Enhancers were associated with genes using manually curated data including (i) tissue-specific co-expression correlation between genes and enhancer RNAs, and between genes and transcription factors; (ii) expression quantitative trait loci (eQTL) for variants within enhancers; (iii) capture Hi-C; and (iv) gene-enhancer genomic distances. Genetic variants within enhancers were prioritised for further analysis if they were rare (MAF<0.01) and evolutionary conserved (LINSIGHT score >0.8).

Whole genome sequencing data was obtained from whole blood for 4,495 ALS patients and 1,925 controls. Reads were aligned to the hg19 human genome and SNP QC was performed using PLINK. Enhancers were grouped according to their target gene(s), and a rare variant burden test was conducted for ALS-association within grouped enhancers.

**Results:** A genome-wide screen of genetic variation within enhancers grouped around individual genes (n=15,603) revealed two new ALS genes with genome-wide significant disease-association: *METTL8* (SKAT-O, p=6.05E-07) and *CAV2* (p=1.28E-06). Association within *CAV1*-enhancers was near genome-wide significant (p=1.02E-05). Interestingly rare genetic variation within coding and promoter regions of *CAV1* is also significantly associated with ALS (Firth logistic regression, p<0.05). Genetic variation within all three regions could conceivably impair *CAV1* function; consistent with this immunoblotting for *CAV1* protein in patient-derived lymphoblastoid cells carrying two identified ALS-associated *CAV1*-enhancer variants revealed reduced protein expression.

**Conclusion:** An unbiased genome-wide screen for ALS-associated genetic variation within enhancer elements has identified new risk variants. Our work highlights new disease biology and new

methodology that could be applied in other disease areas.

## PgmNr 2122: Identification of rare variant burden in neurodegenerative disease.

### Authors:

A.A. Dillio<sup>1,2</sup>; J.S.. Dron<sup>1,2</sup>; S.M.K. Farhan<sup>3</sup>; D. Kwan<sup>4</sup>; B. Tan<sup>5</sup>; M. Ghani<sup>6</sup>; C. Sato<sup>6</sup>; M. Zhang<sup>6</sup>; A.D. McIntyre<sup>1</sup>; H. Cao<sup>1</sup>; J.F. Robinson<sup>1</sup>; M. Freedman<sup>5,7,8</sup>; S.E. Black<sup>9,10</sup>; M. Borrie<sup>11</sup>; L. Zinman<sup>9</sup>; J. Turnbull<sup>12</sup>; E. Finger<sup>13</sup>; M.C. Tartaglia<sup>6,7</sup>; A. Frank<sup>14</sup>; A. Lang<sup>7,15</sup>; D. Grimes<sup>16</sup>; R.H. Swartz<sup>9</sup>; J. Mandzia<sup>13</sup>; D.E. Bulman<sup>17</sup>; E. Rogaeva<sup>6</sup>; R.A. Hegele<sup>1,2</sup>; The ONDRI Investigators

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, ON, CAN; 2) Department of Biochemistry, Schulich School of Medicine and Dentistry, Western University, London, ON, CAN; 3) Analytic and Translational Genetics Unit, Center for Genomic Medicine, Harvard Medical School, Massachusetts General Hospital, Stanley Centre for Psychiatric Research, Broad Institute of MIT and Harvard, Boston, MA, USA; 4) Neuropsychology Operations, Ontario Neurodegenerative Disease Research Initiative, Toronto, ON; 5) Rotman Research Institute, Baycrest Health Sciences, Toronto, ON; 6) Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, ON, CAN; 7) Division of Neurology, Department of Medicine, University of Toronto, Toronto, ON, CAN; 8) Division of Neurology, Department of Medicine, Baycrest Health Sciences and Mt. Sinai Hospital, Toronto, ON; 9) Division of Neurology, Department of Medicine, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, CAN; 10) LC Campbell Cognitive Neurology Research Unit, Hurvitz Brain Sciences Research Program Sunnybrook Research Institute, Toronto, ON, CAN; 11) Division of Geriatric Medicine, Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, ON, CAN; 12) Andrew Bruce Douglas Chair in Neurology, Faculty of Health Sciences, McMaster University, Hamilton, ON, CAN; 13) Department of Clinical Neurological Sciences, Schulich School of Medicine and Dentistry, Western University, London, ON, CAN; 14) Bruyère Memory Program, Élisabeth Bruyère Hospital, Ottawa, ON, CAN; 15) Edmond J. Safra Program in Parkinson's Disease and the Morton and Gloria Shulman Movement Disorders Clinic, Toronto Western Hospital, Toronto, ON, CAN; 16) Neuroscience Program, Ottawa Hospital Research Institute, Ottawa, ON, CAN; 17) Faculty of Medicine, Department of Pediatrics, University of Ottawa, Ottawa, ON, CAN

---

Neurodegenerative diseases are characterized by cognitive decline and motor dysfunction as a result of neuronal loss and are estimated to affect over 50 million people worldwide (GBD 2016 Collaborators, 2017, *The Lancet*). Although substantial effort has been expended to genetically characterize these diseases, many studies have focused on identifying single nucleotide polymorphism associations, each with only a small effect size. We aim to elucidate the rare genetic landscape of neurodegeneration using targeted next-generation sequencing (NGS). In total, 519 participants were enrolled with one of five neurodegenerative diseases: 1) Alzheimer's disease and mild cognitive impairment (AD/MCI); 2) amyotrophic lateral sclerosis (ALS); 3) frontotemporal dementia (FTD); 4) Parkinson's disease (PD); and 5) vascular cognitive impairment (VCI). We sequenced participants and 189 cognitively normal controls using a custom NGS panel that covers 80 genes previously associated with neurodegeneration. We used rare variant association analysis burden testing to identify disease cohorts with an increased frequency of rare variants and

disease-associated genes with greater burdens of variants.

The ALS cohort had a significant burden of rare candidate variants in primary ALS genes compared to controls (OR 4.65 [1.46–14.8]  $P=0.0092$ ). Further, the AD/MCI and PD cohorts had significantly increased burdens of rare candidate variants in secondary AD/MCI and PD genes, respectively, compared to controls (OR 2.18 [1.10–4.32]  $P=0.0260$  and OR 2.81 [1.21–6.56]  $P=0.0166$ , respectively).

Using an Optimized Sequence Kernel Association Test custom script we only identified *PARK2* in the FTD cohort to have a significant burden of rare variants compared to controls ( $P=0.0127$ ). However, other genes approached significance, and further analysis confirmed a rare variant burden. The genes included *PARK2* in AD/MCI and VCI (OR 13.8 [1.65–114.6]  $P=0.0153$  and OR 13.1 [1.52–113.2]  $P=0.0194$ , respectively) and *PINK1* in AD/MCI and VCI (OR 7.65 [1.59–36.89]  $P=0.0113$  and OR 6.46 [1.26–33.1]  $P=0.0252$ , respectively).

Although the study has a relatively modest sample size, these pilot results indicate a genetic contribution by rare variants to neurodegeneration, which can now be further elucidated. Additionally, the burden of rare variants in *PARK2* and *PINK1* across multiple disease phenotypes suggests they could be contributing to overlapping features of diseases, which warrants further investigation.

## PgmNr 2123: Mid-life ethanol intake and cognitive decline: A gene x environment interaction study.

### Authors:

S. Love <sup>1</sup>; G. Heiss <sup>1</sup>; Y. Wang <sup>1</sup>; M. Liu <sup>2</sup>; M. Graff <sup>1</sup>; L. Loehr <sup>1</sup>; D. Zeng <sup>3</sup>; K. North <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.; 2) Institute for Behavioral Genetics, Psychology and Neuroscience, The University of Colorado, Boulder, Colorado.; 3) Department of Biostatistics, The University of North Carolina, Chapel Hill, North Carolina.

---

**Background:** Previous studies of the relationship between ethanol intake and cognitive decline have been inconsistent, yet possible interaction between ethanol intake and genetic susceptibility on the risk of cognitive decline has often not been considered. **Objective:** To investigate if unweighted genetic risk scores (GRSs) based on ethanol intake-associated single nucleotide polymorphisms (SNPs) modify the relationship between weekly ethanol intake in mid-life and 15-year rate of decline in general cognitive performance from mid-to-late life among African-American and European-American adults. **Methods:** A total of 9,183 participants (n=1,733 African-Americans and n=7,540 European-Americans) of the Atherosclerosis Risk in Communities (ARIC) study completed an interviewer-administered questionnaire and two neurocognitive examinations at 1996 and 2013. Twenty ethanol intake-associated SNPs for African-Americans and 11 for European-Americans were combined, separately, to create unweighted GRSs, uGRS<sub>20</sub> and uGRS<sub>11</sub>, respectively. Multivariable linear regression was used to assess modification of the ethanol intake-cognitive decline association by uGRS<sub>20</sub> and uGRS<sub>11</sub>. Multiple imputations by chained equations (MICE) were used to account for attrition. **Results:** Ethanol intake in mid-life was not associated with 15-year decline in general cognitive performance from mid-to-late life (African-American:  $\beta = -0.011$  (95% CI: -0.052, 0.031), European-Americans: -0.010 (-0.021, 0.002)). The uGRS<sub>20</sub> and uGRS<sub>11</sub> did not modify the association of ethanol intake in mid-life with 15-year change in general cognitive performance from mid-to-late life among African-Americans (P= 0.811) and European-Americans (P= 0.847), respectively. **Conclusions:** Ethanol intake in mid-life is not associated with cognitive decline from mid-to-late. There is no indication that an association between ethanol intake and cognitive depends on genetic susceptibility to ethanol intake among African-American and European-American adults.



## PgmNr 2124: Integrated exome sequencing and array-based comparative genomic hybridization in familial Parkinson disease.

### Authors:

L.A. Robak<sup>1</sup>; R. Du<sup>1</sup>; B. Yuan<sup>1</sup>; S. Gu<sup>1</sup>; I. Alfradique-Dunham<sup>2</sup>; V. Kondapalli<sup>1</sup>; E. Hinojosa<sup>1</sup>; A. Stillwell<sup>1</sup>; E. Young<sup>1</sup>; C. Zhang<sup>1</sup>; X. Song<sup>1</sup>; H. Du<sup>1</sup>; S.N. Jhangiani<sup>3</sup>; Z. Coban Akdemir<sup>1,4</sup>; D.M. Muzny<sup>3</sup>; A. Tejomurtula<sup>1</sup>; C. Shaw<sup>1</sup>; W. Bi<sup>1,4</sup>; J.E. Posey<sup>1</sup>; J. Jankovic<sup>2,5</sup>; J.R. Lupski<sup>1,3,6,7</sup>; J.M. Shulman<sup>1,2,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Neurology, Baylor College of Medicine, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 4) Baylor Genetics, Houston, TX; 5) Parkinson's Disease Center and Movement Disorders Clinic, Houston, TX; 6) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 7) Department of Pediatrics, Texas Children's Hospital, Houston, TX; 8) Department of Neuroscience, Baylor College of Medicine, Houston, TX; 9) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX

---

Parkinson disease (PD) is a genetically heterogeneous condition; both single nucleotide variants (SNVs) and copy number variants (CNVs) are important genetic risk factors. Here, we examined the utility of combining exome sequencing (ES) and genome-wide array-based comparative genomic hybridization (aCGH) for identification of PD genetic risk factors. We performed ES on 110 subjects with PD and a positive family history; 99 of these subjects were also evaluated using genome-wide aCGH. We interrogated ES and aCGH data for pathogenic SNVs and/or CNVs at Mendelian PD gene loci, respectively. All SNVs were confirmed via Sanger sequencing. All CNVs were confirmed using orthogonal experimental approaches including a custom-designed high-density array comparative genomic hybridization (aCGH), droplet digital PCR, and breakpoint sequencing. Based on ES, we discovered individuals with known pathogenic SNVs in *GBA* (p.E365K, p.N370S, p.T408M, p.L444P) and *LRRK2* (p.R1441G and p.G2019S). Two subjects were each double heterozygotes for variants in *GBA* and *LRRK2*. Using aCGH to identify CNVs, we additionally discovered 1 *SNCA* duplication and a heterozygous intragenic *GBA* deletion. Five additional cases harbored both an SNV (p.N52fs, p.T240M, p.P437L, p.W453\*) and a likely disrupting CNV at the *PARK2* locus, consistent with compound heterozygosity. In nearly all cases, breakpoint sequencing revealed microhomology; mutational signatures consistent with CNV formation due to DNA replication errors. Integrated ES and aCGH yielded a genetic diagnostic rate of 21.1% of our PD cohort. Our analyses also highlight potential mechanisms for *SNCA* and *PARK2* CNV formation, uncover multilocus pathogenic variation, and identify novel SNVs for further investigation as potential PD risk alleles.

## PgmNr 2125: Parkinson's disease genetics: Modifiers of risk and age at onset in GBA positive individuals.

### Authors:

C. Blauwendraat <sup>1</sup>; M.A. Nalls <sup>1</sup>; L. Krohn <sup>2</sup>; S. Bandres-Ciga <sup>1</sup>; M. Tan <sup>3</sup>; J.R. Gibbs <sup>1</sup>; D.G. Hernandez <sup>1</sup>; R. Kumaran <sup>1</sup>; H. Iwaki <sup>1</sup>; H. Leonard <sup>1</sup>; M.B. Makarious <sup>4</sup>; A. Brice <sup>5</sup>; A.J. Noyce <sup>3</sup>; N. Wood <sup>3</sup>; M. Cookson <sup>1</sup>; J. Hardy <sup>3</sup>; J. Jankovic <sup>6</sup>; J.M. Shulman <sup>6</sup>; P. Heutink <sup>7</sup>; T. Gasser <sup>7</sup>; S.W. Scholz <sup>4</sup>; H. Morris <sup>3</sup>; Z. Gan-Or <sup>2</sup>; A.B. Singleton <sup>1</sup>; International Parkinson's Disease Genomics Consortium (IPDGC)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 2) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 3) Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK; 4) Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA; 5) Inserm U1127, Sorbonne Universités, UPMC Univ Paris 06 UMR S1127, Institut du Cerveau et de la Moelle épinière, ICM, Paris, France; 6) Department of Neurology, Baylor College of Medicine, Houston, USA; 7) German Center for Neurodegenerative Diseases (DZNE), Tuebingen, Germany

---

Parkinson's disease (PD) is a genetically complex disorder. Multiple genes have been shown to cause monogenic PD, and currently 90 independent risk loci have been identified by genome-wide association studies. Thus far, a number of genes have been shown to be pleomorphic risk loci, containing variability across a spectrum of frequency and effect, from rare highly penetrant variants to common risk alleles; these include *SNCA*, *LRRK2*, *VPS13C*, and *GBA*.

*GBA*, encoding the enzyme glucocerebrosidase, contains variants associated with PD. These variants, which reduce or abolish enzymatic activity, confer a spectrum of disease risk, from 1.4- to ~5-fold. An outstanding question in the field is whether there are genetic factors that influence *GBA*-associated disease risk, and whether these overlap with known PD risk variants. Using multiple, large case-control datasets (including IPDGC and UK Biobank), totaling 217,165 individuals including: 22,757 PD cases, 13,431 PD proxy cases, and 180,355 controls, we performed a genome-wide association study and analyzed the most recent PD-associated genetic risk score (GRS) to detect genetic influences on *GBA* risk and age at onset including in total 1,772 *GBA*+ cases, 711 *GBA*+ proxy cases and 7,624 of *GBA*+ controls.

We identified that the PD GRS modifies risk for disease and lowers age at onset in carriers of *GBA* variants. Notably, this effect was consistent across all tested *GBA* risk variants. Dissecting this signal demonstrated that variants in close proximity of *SNCA* and *CTSB* are the most significant contributors. Additional analyses suggest that there is might be a potential genetic interaction between *CTSB* and *GBA* and that the likely functional mechanism of disease risk is a difference of expression of *CTSB*.

These data provide a genetic basis for modification of *GBA* parkinsonism risk and age at onset and support the hypothesis that variability at genes implicated in lysosomal dysfunction exert the largest effect on *GBA* associated risk for disease. The finding of a link between PD associated GRS with *GBA*

associated PD indicates that *GBA* linked PD may not simply be an etiologic subtype of disease but rather that *GBA* and lysosomal dysfunction are more general features in etiopathogenesis of certain forms of alpha synucleinopathies. Further, these results have important implications for selection of *GBA* carriers for therapeutic interventions. HiC

## PgmNr 2126: Using whole genome sequencing to dissect Lewy body dementia: An update.

### Authors:

M.B. Makarious <sup>1</sup>; C. Viollet <sup>2</sup>; R. Chia <sup>3</sup>; M.S. Sabir <sup>1</sup>; E. Martinez-McGrath <sup>2</sup>; S. Bandres-Ciga <sup>3</sup>; H. Leonard <sup>3</sup>; F. Faghri <sup>3</sup>; C. Alba <sup>2</sup>; M. Diez-Fairen <sup>3</sup>; J. Ding <sup>3</sup>; J. Hardy <sup>4</sup>; A. Chio <sup>5</sup>; M. Nalls <sup>3</sup>; J.R. Gibbs <sup>3</sup>; A.B. Singleton <sup>3</sup>; C.L. Dalgard <sup>2</sup>; O. Ross <sup>6</sup>; B.J. Traynor <sup>3</sup>; S.W. Scholz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA; 2) Uniformed Services University of the Health Sciences, Anatomy Physiology & Genetics, Bethesda, MD, USA; 3) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 4) Department of Molecular Neuroscience, University College London, London, UK; 5) Department of Neurology, University of Turin, Turin, Italy; 6) Department of Neuroscience, Mayo Clinic Jacksonville, Jacksonville, FL, USA

---

Genetic analyses of complex neurodegenerative syndromes have become more advanced with the development of whole genome sequencing (WGS) technologies. In this project, we assembled and profiled genomes from ~3,000 samples from Lewy body dementia (LBD) alongside ~2,000 neurologically-healthy controls. This project is the largest WGS effort for this common age-related dementia syndrome. PCR-free DNA libraries were sequenced on the Illumina HiSeq X10 system to study the genetic architecture underlying these two heterogeneous dementia syndromes. Here, we present analyses on both common and rare variant discovery testing. In addition, we applied modern data-driven approaches, such as the construction of polygenic risk scores, applying machine learning to determine which features are explaining the variance, gene and pathway analyses, and two-way Mendelian randomization to better understand the complex genetic composition of these dementia syndromes. All data from this initiative will be made available on the Accelerating Medicines Partnership - Parkinson's Disease (AMP-PD) platform as a public resource, so that other researchers can access, analyze, and combine these genomes with their own cohorts for downstream analyses. With this sequencing initiative, combined with the online platform available for collaboration, we aim to gain a better understanding of the genetic underpinnings of LBD.

## PgmNr 2127: Genetic analysis of risk alleles associated with frontotemporal dementia in a group of Colombian patients.

### Authors:

A.I. Sanchez <sup>1,2</sup>; A. López-Cáceres <sup>3,4</sup>; F. Cruz <sup>4</sup>; P. Mayorga <sup>4</sup>; P. Ayala <sup>3</sup>; D. Matallana <sup>3,4</sup>; I. Zarante <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pontificia Universidad Javeriana Cali; 2) Centro Médico Imbanaco; 3) Pontificia Universidad Javeriana; 4) Fundación Santa Fé de Bogotá

---

**Introduction:** Frontotemporal Dementia (FTD) is a common type of dementia in people younger 65 yo, with strong genetic basis. FTD encompasses distinct clinical symptoms including progressive changes in behavior, language, executive and motor function. The FTD clinical spectrum includes a behavioral variant (bvFTD), a non-fluent/agrammatic variant primary progressive aphasia (nfvPPA), and a semantic variant (svPPA). There are risk loci and causative mutations associated with FTD that have been identified by Genome-Wide Association studies. These genes include *C9orf72*, *GRN*, *MAPT*, and *TMEM106B* among others. **Methodology:** We studied 104 patients clinically diagnosed with FTD with a Pan-Neurodegenerative Disease-Oriented Risk Allele (PANDORA) panel that was performed on a mass ARRAY instrument. The panel was designed by University of Pennsylvania as a method for genotyping common single nucleotide polymorphism (SNP) risk alleles for common neurodegenerative disorders: Alzheimer Disease (AD), Parkinson Disease, FTD and Amyotrophic lateral sclerosis. We analyzed 56 SNPs, calculated minor allele frequencies and Hardy-Weinberg (HW) law for each SNP. All patients had neuropsychological evaluations (Stroop, ineco frontal screening, Rey-Osterrieth complex figure etc). We analyzed data to look for associations between these tests and the risk alleles in each SNP. **Results:** 28 patients were classified with nfvPPA, 60 with bvFTD, and 16 with svPPA. Median age of diagnosis was 61 yo. All SNPs related to FTD genes reached HW. The following SNPs related to AD did not reach HW: rs7412 in *APOE* gene ( $p=0.029$ ); rs6656401 in *CR1* gene ( $p=0.024$ ), and rs983392 in *MS4A6A* gene ( $p=0.009$ ). rs1411478 in *STX6* gene, related to progressive supranuclear palsy, didn't reach HW ( $p=0.014$ ). We found low performance in Total IFS test when the allele A for rs5848 (*GRN* gene) was present (OR: 13 [1.47,115.34],  $p=0.021$ ); low performance in SDMT test when allele A in rs1020004 (*TMEM106B* gene) was present (OR:3.4 [1.06,10.89],  $p=0.039$ ). Other alleles in SNPs of *TMEM106B* gene were also significant associated with low performance in Stroop test. **Discussion:** a pleiotropic effect is seen in FTD, meaning that multiple loci have small additive effects for increasing its risk. Having any of these SNPs, should not be considered as a necessary factor to develop FTD. More studies are needed to determine if having one or several SNPs confers additional risk for the development and severity of FTD.

## PgmNr 2128: Genetic differences between neuropathologically confirmed pure dementia with Lewy bodies and Parkinson's disease dementia.

### Authors:

D. Tsuang<sup>1,2</sup>; O. Lorenzo Betancor<sup>1,2</sup>; K. Nuytmans<sup>3</sup>; I. Mata<sup>4</sup>; M. Pericak-Vance<sup>3</sup>; D. Bennett<sup>5</sup>; D. Galasko<sup>6</sup>; D. Irwin<sup>7</sup>; G. Jicha<sup>8</sup>; J. Kofler<sup>9</sup>; J. Kaye<sup>10</sup>; J. Leverenz<sup>4</sup>; O. Lopez<sup>9</sup>; E. Masliah<sup>6</sup>; R. Mayeux<sup>11</sup>; T. Montine<sup>12</sup>; P. Nelson<sup>8</sup>; J. Quinn<sup>10</sup>; J. Schneider<sup>5</sup>; J. Trojanowski<sup>7</sup>; D. Weintraub<sup>7</sup>; R. Woltjer<sup>10</sup>; C. Zabetian<sup>1,2</sup>; Alzheimer's Disease Sequencing Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) VAPSHCS, Seattle, Washington.; 2) University of Washington; 3) University of Miami; 4) Cleveland Clinic; 5) Rush University; 6) University of California at San Diego; 7) University of Pennsylvania; 8) University of Kentucky; 9) University of Pittsburgh; 10) Oregon Health Science University; 11) Columbia University; 12) Stanford University

### Objectives

Parkinson's disease dementia (PDD) and dementia with Lewy bodies (DLB) are differentiated based on the temporal onset of parkinsonism versus cognitive impairment. The disorders share classical Lewy pathologies and may be neuropathologically indistinguishable. We sought to identify genetic differences between the disorders.

### Methods

PDD (n=54) Caucasian subjects presented with parkinsonism and pure DLB (pDLB; n=78) subjects presented with dementia. Both groups were defined by the neuropathologic criteria. Coding regions were screened for 27 causative AD, PD, or neurodegeneration-related genes: sample libraries were constructed and enriched for these genes using xGEN custom probes; >100X mean sequence coverage was generated per sample (Illumina); and sequences were aligned to hg19 (Burrows-Wheeler Aligner) and variants were called and annotated with custom Genome Analysis Tool Kit and ANNOVAR, respectively. To evaluate disease specificity, variants of interest identified in the PDD and DLB patients were screened in the 2237 Alzheimer disease (AD) cases and 2207 controls of the 1st data release of the AD Sequencing Project.

### Results

883 SNPs from 27 genes passed QC filters and were advanced to the analyses.

We compared frequencies of variants with MAF <5% between the pDLB and PDD groups using the gene burden tests (SKAT-O) for all 27 genes. We found significant differences in only 1 gene, *LRRK2*. We found that the burden of rare missense variants in *LRRK2* (not known to be pathogenic) were significantly higher in PDD compared to pDLB ( $p = 0.003$ ). Both groups also harbored unique *LRRK2* variants (7 for PDD; 8 for pDLB). Of the unique MS variants, only 2 were observed in the ADSP; making p.M712T and p.T776M very rare and specific to pDLB and PDD respectively). p.N2081D was previously reported in gnomAD NFE (1.7%) and subsequently identified in both AD cases and controls in the ADSP. Interestingly, p.R1941H is extremely rare in gnomAD NFE, but was identified in 8 AD and 6 controls of the ADSP; all of Hispanic descent.

**Conclusions:** Genetic differences in *LRRK2* exist between neuropathologically confirmed pDLB and

PDD. These findings suggest that genetic signatures may differentiate PDD from DLB. Though we identified some variants specific to PDD and pDLB, 1 of the rare variants also appears in Hispanic AD and controls, confirming that the inclusion of subjects from the same ethnicity is crucial to fully understand the underlying genetics to disease.

## PgmNr 2129: Neuropsychiatric features differ in individuals of African and European descent with Alzheimer disease (AD).

### Authors:

P.R. Mena <sup>1</sup>; J. Ramos <sup>1</sup>; K. Celis <sup>1</sup>; L.D. Adams <sup>1</sup>; A. Scott <sup>1</sup>; T. Stark <sup>2</sup>; F. Lacroix <sup>1</sup>; J.L. Haines <sup>3</sup>; C. Reitz <sup>4</sup>; G. Beecham <sup>1</sup>; G.S. Byrd <sup>2</sup>; M.A. Pericak <sup>1</sup>; J.M. Vance <sup>1</sup>; M.L. Cuccaro <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John Hussman Institute of Human Genomics, University of Miami, Miami, Florida.; 2) Department of Biology, North Carolina A&T State University, Greensboro NC; 3) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland OH; 4) Taub Institute, Columbia University, New York, NY

**Background:** There is evidence that genetic risk for AD differs between African Americans (AA) and individuals of European descent (EU). However, whether these differences in genetic risk for AD translate into phenotypic differences between the groups is not known.

We investigated neuropsychiatric phenotypes in our AA and EU groups. Prior studies indicate that these phenotypes may vary by race and ethnicity in AD. Understanding neuropsychiatric phenotypes in AD may provide insights into genetic mechanisms.

**Methods:** Participants were enrolled in genetic studies of AD and met criteria for mild cognitive impairment or AD. Self-reported race was confirmed by genetic analysis. Neuropsychiatric features were assessed using the 12-item Neuropsychiatric Inventory Questionnaire (NPI-Q). NPI-Q items (e.g., agitation and depression) are scored as present/absent and summed to create a total score (higher scores=more problems).

Our primary hypothesis was that our AA and EU groups would differ on NPI-Q outcomes. To test this, we compared mean NPI-Q total scores for our groups using an independent samples t-test and performed logistic regression to identify those NPI-Q items that differentiate AA and EU cases.

**Results:** We had 675 cases (EU=53%; AA=47%) with complete NPI-Q data. Our groups did not differ on age of onset of AD ( $p=0.13$ ) or distribution of CDR scores ( $p=0.79$ ). Mean NPIQ total scores were nominally significantly higher in our AA group ( $p=0.04$ ). Our logistic model was significant ( $p<0.001$ ) and the odds of being in the AA group was 3.1 times greater (95% CI 2.0, 4.8) when agitation was present while the odds of being in the EU group was 2.2 time greater (95% CI 1.5, 4.4) when anxiety was present and 1.7 times greater (95% CI 1.2, 2.5) when apathy was present.

**Conclusions:** We identified differences in our AA and EU groups on select neuropsychiatric features. Among EU, there was a greater likelihood of anxiety and apathy (i.e., apprehension or indifference) while our AA group showed a greater likelihood of agitation (i.e., resistant, uncooperative behavior). While our findings may be explained by non-genetic causes (e.g., SES), the presence of distinctive neuropsychiatric problems may be phenotypically and genetically relevant to AD in these populations.



## **PgmNr 2130: Longitudinal analysis of 4500 whole blood transcriptomes across over 1500 whole-genome sequenced individuals within the Parkinson's Progression Markers Initiative.**

### **Authors:**

D.W. Craig <sup>1</sup>; I. Violich <sup>1</sup>; E. Hutchins <sup>2</sup>; E. Alsop <sup>2</sup>; C. Blauwendraat <sup>4</sup>; S. Levy <sup>3</sup>; A. Singleton <sup>4</sup>; M.R. Cookson <sup>4</sup>; J.R. Gibbs <sup>4</sup>; K. Van Keuren-Jensen <sup>2</sup>; The Parkinson's Progression Markers Initiative

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Translational Genomics, USC, Keck School of Medicine, Los Angeles, California.; 2) Neurogenomics, TGen, Phoenix, AZ; 3) Genomic Services Laboratory, HudsonAlpha Institute for Biotechnology, Huntsville, AL; 4) Laboratory of Neurogenetics, NIA, Bethesda, MD

---

Transcriptome analyses of whole blood samples from normal healthy control subjects and patients with Parkinson's disease (PD) have the potential to uncover biological pathways disrupted by disease processes. Detectable changes in the transcriptome of a readily accessible biofluid, such as blood, have the potential to be valuable biomarkers. Moreover, integration of transcriptomic data with genomic whole-genome sequencing data has the potential to give insight to variants of unknown significance and previously identified genetic loci. In this study, we report transcriptome profiling of whole blood from 4,600 samples longitudinally collected from over 1,600 individuals who also have whole genome DNA sequence, as part of Michael J. Fox Foundation's Parkinson's Progression Markers Initiative (PPMI) cohort. Paxgene derived whole-blood samples were sequenced at baseline and at months 6, 12, 24, and 36 to a depth of over 200 million reads, with many individuals having over 1 billion reads over the time series. Longitudinal analysis was conducted using across a series of clinical variables, examining clinical phenotypes, carrier status, and therapeutic drug usage. We analyzed expression data with existing whole-genome sequencing data, reporting on the robustness of expression quantitative loci (eQTL) over the 3 year period, examining both novel and known eQTLs in PD associated genes. We detail the utility of longitudinal analysis and will present on differentially expressed transcripts between PD and normal controls, as well as transcript changes associated with time. Finally, we describe a series of resources to enable other researchers, including a series of tools for querying and searching through the clinical/genomics data via an analysis portal.

## PgmNr 2131: Differential gene expression analysis of CD8<sup>+</sup> T cells from multiple sclerosis patients and healthy controls.

### Authors:

S.D. Bos<sup>1,2</sup>; I.S. Brorson<sup>1,2</sup>; A. Eriksson<sup>1,2</sup>; I.S. Leikfoss<sup>1,2</sup>; V. Vitelli<sup>3</sup>; E.G. Celius<sup>1,2</sup>; T. Lüders<sup>4</sup>; T. Berge<sup>2,5</sup>; H.L. Nilsen<sup>4</sup>; H.F. Harbo<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 2) Department of Neurology, Oslo University Hospital Ullevål, Oslo, Norway; 3) Oslo Centre for Biostatistics and Epidemiology, Department of Biostatistics, University of Oslo, Oslo, Norway; 4) Department of Clinical Molecular Biology, University of Oslo and Akershus University Hospital, Oslo, Norway; 5) Department of Mechanical, Electronics and Chemical Engineering, Oslo Metropolitan University, Oslo, Norway

---

### Introduction

Through recent genome-wide approaches the adaptive immune system has been implicated as a major contributor to the risk for developing multiple sclerosis (MS). Furthermore, post-mortem analysis of MS lesions has pointed out that CD8<sup>+</sup> T cells may be involved in the disease processes. A detailed insight in the activity of genes involved in the MS processes will contribute to a better understanding of disease mechanisms and may provide valuable markers for disease and disease activity.

### Materials and methods

CD8<sup>+</sup> T cells were purified from whole blood from 20 MS patients and 20 controls. RNA was extracted and sequenced in 12-plex 75-basepair paired-end reads on an Illumina NextSeq sequencer in order to get detailed insights in gene expression. The raw reads were mapped using the program Kallisto. The per-sample-per-gene read count matrix was imported into DeSeq2 for differential expression analysis. We included surrogate variable analysis to account for hidden confounders.

### Results

After correction for multiple testing according to Benjamini and Hochberg we identified 73 genes with significant differential expression. These genes may provide valuable insights in which biological pathways are affected in CD8<sup>+</sup> T cells in MS patients.

### Discussion and conclusion

CD8<sup>+</sup> T cells from MS patients show differential gene expression compared to healthy controls. Further analysis of the differentially expressed genes in a pathway analysis may provide a much-needed insight in how these cells contribute to MS disease processes. Furthermore, by analysis of the RNA sequencing dataset in the perspective of the large-scale genome wide association studies by the international MS genetics consortium (IMSGC) may provide additional clues to which genes in CD8<sup>+</sup> T cells are driving or marking MS disease processes.

## **PgmNr 2132: Single-cell transcriptomics analyses of cells from a novel tri-culture model of HIV neuropathogenesis reveals cellular heterogeneity and inflammatory gene expression signatures.**

### **Authors:**

M. Gonzalez <sup>1</sup>; S. Ryan <sup>2</sup>; J. Smiler <sup>1</sup>; F. Mafra <sup>1</sup>; H. Hakonarson <sup>1,3</sup>; S. Anderson <sup>2,3</sup>; K. Jordan-Sciutto <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Psychiatry, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pathology, University of Pennsylvania Dental Medicine, Philadelphia, PA

---

HIV-Associated Neurocognitive Disorders (HAND) affect 55% of HIV-infected individuals worldwide. While antiretroviral treatments have reduced the severity of HAND, the prevalence has increased due to increased life expectancy. In addition, little progress has been made in developing therapeutics to reduce the prevalence of HAND. While the major pathological manifestation of HAND is synaptodendritic damage, the full, underlying mechanism is unknown partly since there is no in vitro model to study the direct interactions between HIV-infected macrophages/microglia and neurons. In order to address this problem, we developed a human-induced pluripotent stem cell (HiPSC) based model; whereby, we separately differentiate HiPSCs into glutamatergic-like neurons, astrocytes, and microglia and create a co-culture of the three cell types with or without HIV-infection. We used single-cell RNA-sequencing (scRNA-seq) to understand gene expression changes under four separate conditions including during HIV infection in the presence or absence of Efavirenz (EFZ), an antiretroviral treatment that prevents integration of the HIV virus into host cells. The cellranger pipeline (10x Genomics, v.3.0.1) was used for initial analysis and aggregation of samples. Secondary analysis was performed using the Seurat package in R. After QC, 31,177 cells (39,227 post-normalization mean reads per cell, 2,165 median genes per cell) were characterized from the separate conditions (6,564 cells from control condition, 7,431 from HIV infected, 7,111 from HIV infected+EFZ, and 10,071 from control+EFZ). Automated clustering generated 16 distinct cell clusters. Cells were identified into one of the 3 cell populations based on canonical gene markers in each cluster. Inflammatory pathways were altered in all HIV infected cell types compared with uninfected controls, suggesting that each cell type may play a role in HIV mediated neuropathogenesis. The upregulation of the inflammatory gene signature was most pronounced in HIV infected microglia, indicating that neuroinflammatory roles of each cell type may not be equal. Additionally, a markedly different gene expression profile between infected and infected with antiretroviral treatment conditions was observed, suggesting different mechanistic drivers of damage between pure HIV infection and a combination of HIV infection with antiretroviral treatment.

## PgmNr 2133: Human iPSC-derived microglia are genetically relevant to Alzheimer's disease.

### Authors:

R.R. Butler<sup>1,2</sup>; A. Kozlova<sup>1,2</sup>; H. Zhang<sup>1</sup>; S. Zhang<sup>1,2</sup>; M. Streit<sup>1</sup>; A.R. Sanders<sup>1,2</sup>; Z.P. Pang<sup>3</sup>; P.V. Gejman<sup>1,2</sup>; J. Duan<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Psychiatric Genetics, NorthShore University HealthSystem, Evanston, IL; 2) Department of Psychiatry of Behavioral Neuroscience, The University of Chicago, Chicago, IL; 3) Department of Neuroscience and Cell Biology and Child Health Institute of New Jersey, Rutgers University, New Brunswick, NJ

---

Microglia are the primary innate immune cell type in the brain that have been implicated in the pathogenesis of neurodegenerative disorders such as Alzheimer's disease (AD) and neuropsychiatric disorders such as schizophrenia. Microglia generated from human-induced pluripotent stem cells (hiPSCs) represent a promising *in vitro* cellular model for studying the neuroimmune interactions involved in AD. Among several methods of generating hiPSC-derived microglia (iMG)—varying in duration and resultant purity—a recent protocol by Brownjohn et al. is particularly simple and efficient. However, the replicability of this method, transcriptomic similarity of these iMG to primary adult microglia, and their genetic relevance to AD remain unclear. Using two hiPSC lines, we demonstrated that Brownjohn's protocol can rapidly generate iMG that morphologically and functionally resembled microglia. The iMG cells we generated were found to be transcriptionally similar to previously reported iMG lines, as well as fetal and adult microglia. Furthermore, by using cell type-specific gene expression to partition disease heritability, we showed that iMG cells are genetically relevant to AD. Across a range of neuronal and immune cell types we found only iMG, primary microglia, and microglia-like cell types exhibited a significant enrichment for AD heritability. Our results thus support the use of iMG as a human cellular model for understanding AD biology and underlying genetic factors, as well as for developing and efficiently screening new therapeutics.

## PgmNr 2134: Longitudinal behavioral characterization of Gaucher-associated Parkinson murine models.

### Authors:

J. Do; B. Berhe; N. Tayebi; E. Sidransky

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Mutations in the glucocerebrosidase gene, *GBA1*, impart an increased risk for Parkinson disease. Murine models created to elucidate the link between *GBA1* and Parkinson vary widely in mutation type and severity and have been evaluated at different ages, complicating comparative analyses between different mouse models. As *GBA1*-associated Parkinson is an age-related disorder, elucidating age-dependent alterations in behavior in mouse models can provide an appropriate timeline for future therapeutic studies. We crossed mice overexpressing the mutated human  $\alpha$ -synuclein transgene (*SNCA*<sup>A53T</sup>) with a heterozygous *GBA1* knockout mouse (*gba*+/-), and the behavior of the corresponding WT/WT, *gba*+/+//*SNCA*<sup>A53T</sup> and *gba*+/- mice were initially analyzed at 18 months and then monthly for four consecutive months. We then tested younger groups of mice at eight months and re-tested them every other month for six months. Three phenotypes were tested: memory by the novel object recognition test, olfaction by the buried pellet test, and motor coordination by the beam walk test. Gait, assessed by a tunnel-guided gait walk, also provided a qualitative comparative measure. The collected data were analyzed in GraphPad Prism 8.

The study revealed mild, age-associated differences between experimental groups on the behavioral tasks. Over-expression of *SNCA*<sup>A53T</sup>, regardless of genotype, resulted in impaired motor coordination starting from 10 months of age ( $p = 0.037$ ). Olfaction and memory were well-preserved throughout this time span in all mice, regardless of genotype, although *gba*+/+//*SNCA*<sup>A53T</sup> mice exhibited a trend suggestive of impaired memory after age 12 months. Physiological parameters assessed in *gba*+/-//*SNCA*<sup>A53T</sup> mice at 22 months showed decreased weight ( $p < 0.05$ ) and reduced glucocerebrosidase (GCase) activity in brain tissue ( $p < 0.05$ ), confirming the late-onset physiological effects of combining *gba* and *SNCA* mutations.

This study indicates that even prior to the development of symptoms, *gba*+/-//*SNCA*<sup>A53T</sup> mice model the age-dependent decline in motor behavior seen in Parkinson disease from as early as 10 months of age. These results suggest that motor coordination in *gba* haploinsufficient mice can be utilized as an efficacy endpoint for testing novel therapeutics at different timepoints.

## PgmNr 2135: Spinocerebellar ataxia STRs identified in Mexican Americans families show association with decreased cerebellum volume and decreased working memory.

### Authors:

N.B. Blackburn <sup>1</sup>; D.C. Glahn <sup>2,3</sup>; M.F. Bennett <sup>4,5</sup>; J.M. Peralta <sup>1,6</sup>; J.C. Charlesworth <sup>6</sup>; E.E.M. Knowles <sup>2</sup>; S.R. Mathias <sup>2</sup>; S. Kumar <sup>1</sup>; A.C. Leandro <sup>1</sup>; R.L. Olvera <sup>7</sup>; P.T. Fox <sup>8,9</sup>; J.E. Curran <sup>1</sup>; M. Bahlo <sup>4,5</sup>; J. Blangero <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, and South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX, USA; 2) Department of Psychiatry, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA; 3) Olin Neuropsychiatry Research Center, Institute of Living, Hartford Hospital, Hartford, CT, USA; 4) Population Health and Immunity Division, the Walter and Eliza Hall Institute of Medical Research, Parkville, VIC, Australia; 5) Department of Medical Biology, University of Melbourne, Parkville, VIC Australia; 6) Menzies Institute of Medical Research, University of Tasmania, Hobart, TAS, Australia; 7) Department of Psychiatry, University of Texas Health Science Center San Antonio, San Antonio, TX, USA; 8) Research Imaging Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA; 9) Department of Radiology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

---

Short-tandem repeats (STRs) are a typically uncharacterized but clinically relevant form of genetic variation (DNA repeat expansions) that have been shown to cause over 40, primarily neurological, disorders. We used two methods for STR detection (ExpansionHunter and exSTRA) to profile 21 clinically relevant STRs in short-read Illumina WGS from 1159 individuals from the Genetics of Brain Structure and Function Study (GOBS); a randomly ascertained, pedigree-based cohort of Mexican Americans from San Antonio, Texas that also includes comprehensive neuroimaging and neurocognitive assessments.

We identified 4 individuals with expansions at the SCA8 locus (*ATXN8*) and 16 carriers of a SCA10 (*ATXN10*) expansion. These STRs are responsible for the development of autosomal dominant spinocerebellar ataxia 8 and 10, respectively. GOBS does not specifically record clinical ataxias but this disease may be undiagnosed in this clinically underserved minority group. To date, SCA10 expansions have only been observed in ataxia patients with apparent Amerindian ancestry, supporting our findings. Our ExpansionHunter estimates suggest the SCA8 carriers have an expansion in the pathogenic range of >74 repeats, whereas the SCA10 carriers are expanded beyond the normal range with >30 repeats, but not within the pathogenic range (>500 repeats). We therefore hypothesized that, in the absence of overt ataxia diagnoses, we would detect association with ataxia related quantitative phenotypes in STR carriers.

To test for association with quantitative phenotypes we used variance components modelling in SOLAR. This incorporates the non-independence of related individuals in this cohort into the analysis

to identify mean unit differences in phenotypes in variant carriers versus non-carriers. For carriers of the SCA8 expansion we identified 1.19 and 0.89 standard deviation unit (SDU) decreases in the left and right cerebellum cortex volumes ( $P=0.002$  and  $0.02$  respectively) and a 1.05 SDU decrease in bilateral cerebellum cortex volume ( $P=0.006$ ). For carriers of the SCA10 expansion we did not detect cerebellum size changes however a 0.5 SDU decrease in response time to a spatial capacity test as a measure of working memory was identified ( $P=0.009$ ).

STRs have been posited as a potential source for some of the 'missing heritability' of complex disease. Here we show that two clinically relevant STRs are associated with spinocerebellar ataxia relevant quantitative phenotypes.

## PgmNr 2136: *De novo*/shared mutations identified for autism in the US Hispanic population.

### Authors:

P. Acevedo <sup>1</sup>; J. Peralta <sup>2</sup>; K. Wang <sup>3</sup>; R. Crutchfield <sup>1</sup>; C. Xu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Health and Biomedical Sciences, College of Health Professions, University of Texas Rio Grande Valley, Brownsville, TX; 2) Department of Human Genetics, South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX; 3) Department of Family and Community Health, School of Nursing, West Virginia University, Morgantown, WV

---

**Introduction:** Autism, a complex neurodevelopmental disorder with a prevalence rate of 1% worldwide, has been seen to have varying etiologies, trajectories and has been found to affect everyone differently. ASD has been found to deficits in communication, repetitive behavior, abnormal social functioning, etc. Autism has been found to be caused by the interactions between genetic factors and environmental factors such as exposure to toxic chemicals, advanced parental age, and preterm birth among other factors. Although there have been recent advancements in the molecular techniques used for the discovery of causative variants for multiple diseases, there has not been much progress in the discovery of causative variants for Autism. The limited studies conducted for the discovery of causative variants has been conducted on a predominantly Caucasian population. Very few studies have been performed on the US Hispanic/Latino population.

**Material and Method:** Whole exome sequencing (WES) was used for the discovery of causative mutations in nine affected children and one affected parent from five families.

**Results:** We identified two genes/mutations in the AR gene and the ACOXL gene, where the mutation found in the AR gene is a *de novo* mutation and the mutation found in the ACOXL gene is a mutation found to be shared between an affected child and affected parent from the same family. Gene information such as expression in the central nervous system, the function, and if they were previously reported was gathered and used to identify these mutations. These variants will then go through PCR and will be sent to a biotech company for Sanger Sequencing for further confirmation.

**Conclusion:** This WES led to the discovery of additional causative mutations in the US Hispanic population; however, future validation using a large sample are needed to confirm the current findings.



## **PgmNr 2137: Assessment of frequency and pathogenicity of loss-of-function and missense variants in candidate genes of autistic spectrum disorders.**

### **Authors:**

T.F. Almeida <sup>1</sup>; R.M. Minillo <sup>1</sup>; M.R.S Passos-Bueno <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Laboratório clínico do Hospital Israelita Albert Einstein; 2) Instituto de Biociências, Departamento de Genética e Biologia Evolutiva

---

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects around 1-2% of the world's population. The etiology of ASD is still not fully understood, however, many advances have been made to understand its genetic component. Next generation sequencing pointed that rare alterations of loss-of-function (LoF) and missense changes with high *in silico* prediction of damage in neurodevelopment genes are contributing to the mechanism of the disease. Whole exome sequencing (WES) of 48 individuals with TEA and 119 controls were evaluated, 11 lists of genes of relevance for TEA ranging from 23 to 7764 genes were evaluated. The number of rare variants (frequency less than 0.01) and singleton variants were counted for each individual. There was no difference (measured by Chi-Square test) in the number of LoF variants found in individuals with ASD and controls in any of the scenarios tested. The missense variants were ordered for their pathogenicity by the SPRING software and the distribution of the variants compared in the interest groups. There was no difference in the mean of non-synonymous variants per individual, nor in the degree of pathogenicity in individuals with ASD and controls in any of the tested scenarios. Due to the small sample size, the results of this study should be interpreted with limitations, and it is necessary to replicate this scenario in other databases. However, these findings suggest that the choice of candidate genes for ASD may increase the number of false positives in clinical reports and the determination of causative variant must be accompanied by further investigations other than WES.

## PgmNr 2138: Whole exome sequencing of a monozygotic triplet with autism spectrum disorder identifies two homozygous mutations in genes involved in neuronal development.

### Authors:

Y. AlSarraj <sup>1,2</sup>; E. AlDous <sup>2</sup>; R. Taha <sup>2</sup>; I. Thompson <sup>2</sup>; F. AlShaban <sup>2</sup>; H. ElShanti <sup>3</sup>; O. Albagha <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar; 2) Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Doha, Qatar; 3) University of Iowa, Carver College of Medicine, Division of Medical Genetics & Genomics, Iowa City

---

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects 1-1.5% of individuals and characterized by impairments in social interaction and communication, along with repetitive and restrictive behaviors. ASD is usually accompanied by heterogeneous phenotypic comorbidities, such as mild to severe intellectual disability and epilepsy. ASD is highly heritable with complex genetic architecture where both rare and common genetic variations contribute to its etiology. Rare genetic variations in many genes have been identified in ASD patients and genome-wide association studies have identified multiple common genetic risk variants for ASD. In this study, we performed whole exome sequencing of a consanguineous middle-eastern family in which a monozygotic triplet was affected with ASD. The Autism Diagnostic Interview-Revised scores for the proband were consistent with ASD diagnosis with developmental delay and epileptic seizures. Exome capture was performed using SureSelect Human all exon v6 followed by sequencing on Illumina platform. Identity by state analysis confirmed monozygosity of the triplet.

Results showed a homozygous missense mutation in the *VPS13B* gene. Both parents were heterozygotes for this mutation and all three ASD patients were homozygotes. The mutation was very rare in the gnomAD database (MAF=0.00002, homozygous count=0) and predicted to be pathogenic by functional prediction tools. Current evidence indicates an important role for *VPS13B* in normal growth and development of neurons. Mutations in *VPS13B* have been reported in patients with Cohen syndrome, a rare autosomal recessive disease characterized by intellectual disability, dysmorphism, and microcephaly.

Sequencing results also revealed a homozygous missense mutation in the *PIGN* gene in the three ASD patients. The mutation was rare in the gnomAD database (MAF=0.0007, homozygous count=1) and predicted to be pathogenic. The *PIGN* protein is involved in glycosylphosphatidylinositol (GPI)-anchor biosynthesis and homozygous mutations in *PIGN* have been reported in patients with multiple congenital anomalies including delayed psychomotor development, hypotonia, and seizures.

Our data reveal the complex genetic architecture of ASD and identify two homozygous mutations in genes involved in neuronal development. The mutation in *PIGN* could explain the observed seizures in the patients but further functional studies will be required to confirm this finding.

## PgmNr 2139: Bioinformatics analysis of GWAS for autism spectrum disorders and related traits.

### Authors:

R.M. Cantor-Chiu <sup>1,2</sup>; J.K. Lowe <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Human Genetics,; 2) Dept Psychiatry and Biobehavioral Sciences, David Geffen UCLA Sch Med, Los Angeles, California.

---

Autism Spectrum Disorders (ASDs) are neuropsychiatric conditions affecting communication skills and social interactions. They feature related traits such as restricted and repetitive behaviors and have an onset prior to 3 years of age. ASDs, with a 4 male:1 female sex ratio has current prevalence estimate of 1/37 in males. The consensus is that ASDs are polygenic, with as many as 1,000 predisposing genes. Robust detection of these common risk genes having small effect sizes is contingent upon the development of large study samples that permit well-powered GWAS. A recent GWAS meta-analysis of approximately 45,000 Danes that identified 5 loci for ASDs at  $p < 5e-8$  provides a marked improvement over previous studies, however, this sample is still smaller than those for other complex disorders with robust findings. To address robustness using a different approach, we conducted a bioinformatics analysis aimed at detecting associations for ASDs and related traits from published GWAS using 2 publicly available databases with different inclusion criteria. We hypothesized that associations meeting stringent criteria in both GWAS databases are more likely to be robust, and searched for concordance between the Simons Foundation Autism Research Initiative SFARI Gene-Human Gene Module database and the NIH/NHGRI GWAS Catalog (GC) for ASDs and related traits. These databases apply markedly different approaches to curation. SFARI employs a panel of scientists who score the published evidence for GWAS results on an ordinal scale (score) ranging from 1 to 6, with 1 representing the most stringent criterion. GC reports all genes published in the GWAS literature along with their p-values and other relevant information. When we applied a score threshold of 3 for SFARI and a p-value of  $5e-8$  for GC the analysis identified 822 GC and 18 SFARI genes. Their overlap consisted of 3 genes: (*GRIN2A* ( $p=2e-9$ ,  $score=3$ ), *MACROD2* ( $p=4e-8$ ,  $score=2$ ) and *PHB* ( $p=7e-9$ ,  $score=3$ )). There was a .4% concordance in the results. Similar patterns were observed with a GC criterion of  $p < 5e-8$  and SFARI thresholds of 2 and 4. These analyses 1) illustrate the feasibility of identifying robust GWAS results using publicly available databases, 2) illustrate that different methods of curation allow the application of diverse criteria, and 3) provide additional support for *GRIN2A*, *MACROD2*, and *PHB* as risk genes for ASDs and their related traits.

## PgmNr 2140: Investigating the genetic architecture of different autism spectrum disorder subtypes through deep phenotyping and whole-genome sequencing.

### Authors:

A. Chan <sup>1,2</sup>; W. Engchuan <sup>1</sup>; M.S. Reuter <sup>1</sup>; Z. Wang <sup>1</sup>; B. Thiruvahindrapuram <sup>1</sup>; G. Pellecchia <sup>1</sup>; T. Nalpathamkalam <sup>1</sup>; B. Trost <sup>1</sup>; S. Lamoureux <sup>1</sup>; K. Tammimies <sup>3</sup>; S. Walker <sup>1</sup>; M. Zarrei <sup>1</sup>; D.J. Stavropoulos <sup>4</sup>; D. Merico <sup>1,5</sup>; R.K.C. Yuen <sup>1,2</sup>; B.A. Fernandez <sup>6,7</sup>; S.W. Scherer <sup>1,2,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 3) Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Karolinska Institutet, Solna, Sweden; 4) Department of Pediatric Laboratory Medicine, Genomic Diagnostics, The Hospital for Sick Children, Toronto, Ontario, Canada; 5) Deep Genomics Inc., Toronto, Ontario, Canada; 6) Provincial Medical Genetics Program, Eastern Health, St John's Newfoundland and Labrador, Canada; 7) Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St John's Newfoundland and Labrador, Canada; 8) McLaughlin Center, University of Toronto, Toronto, Ontario, Canada

---

Although autism spectrum disorder is phenotypically and genetically heterogeneous, clinical morphology classification can help stratify ASD heterogeneity to resolve genotype-phenotype relationships in ASD. Using microarray and whole-exome sequencing, we previously determined a molecular diagnostic yield (MDY) of 15.8% in 95 ASD cases, with the yield higher in the complex (more dysmorphic) ASD subtype (Tammimies *et al*, JAMA, 2015). Here, we aimed to examine the MDY in a larger ASD cohort (n=327), and to dissect the genes and pathways involved in different ASD subtypes using whole genome sequencing (WGS), with the advantage of detecting a full size spectrum of variants in a single experiment. A total of 327 children with ASD from Newfoundland, Canada were enrolled using a population-based recruitment strategy. Standardized dysmorphology assessments were used to stratify the cohort into 193 essential (non-dysmorphic), 57 equivocal, and 77 complex ASD cases. We identified clinically significant single nucleotide variants, short insertions/deletions and copy number variants in 15.0% of ASD cases. The yield was higher in the complex subgroup (39.0%; 30/77) compared to the essential subgroup (6.7%; 13/179) ( $p=7.61 \times 10^{-10}$ ). Clinically significant variants included 16p11.2 microdeletion and microduplication, and variants impacting *SHANK3* and *CHD8*. Two ASD cases (~0.6%) had two clinically significant and/or ASD-risk variants. By performing a global burden analysis, we found a significant enrichment of rare deletions impacting genes in ASD cases with more dysmorphologies ( $p=7.03 \times 10^{-4}$ ). Via gene-set enrichment analysis, cases with more dysmorphologies had a difference in burdens of rare deletions and duplications in genes responsible for neuronal function (e.g., neurotransmission,  $p=4.25 \times 10^{-4}$ , FDR=0.02) and expressed specifically in brain ( $p=9.30 \times 10^{-3}$ , FDR=0.1), respectively. Moreover, the number of dysmorphologies was positively correlated with the number of missense variants in genes responsible for neuronal (e.g., neurogenesis,  $p=4.32 \times 10^{-4}$ , FDR=0.03) and mental function ( $p=4.48 \times 10^{-4}$ , FDR=0.02). Our findings show that WGS captures clinically significant variants with different size spectra for molecular diagnosis of ASD. By using a combination of rich genome-wide

data and extensive clinical data, we also dissect the specific brain and neuronal functions that are impacted by rare variants in ASD cases with more dysmorphologies.

## PgmNr 2141: Microbiome-wide analysis of children with autism spectrum disorder and their neurotypical siblings identifies numerous differentially enriched taxa.

### Authors:

B.S. Chrisman <sup>1</sup>; J.Y. Jung <sup>2,3</sup>; N.T. Stockham <sup>4</sup>; M. Varma <sup>5</sup>; P.Y. Washington <sup>1</sup>; K.M. Paskov <sup>2</sup>; M.W. Sun <sup>2,3</sup>; D.P. Wall <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Bioengineering, Stanford University, Stanford, CA; 2) Department of Biomedical Data Science, Stanford University, Stanford, CA; 3) Department of Pediatrics, Stanford University, Stanford, CA; 4) Department of Neuroscience, Stanford University, Stanford, CA; 5) Department of Computer Science, Stanford University, Stanford, CA

---

Despite the surge in genome-wide association studies (GWAS) in the past decade, no genotype-to-phenotype models currently exist for most complex diseases. One potential explanation is that GWASes ignore an important component of genetic diversity: - the trillions of microbes living in the human gut, which cumulatively contain more than 3 million genes. Autism Spectrum Disorder (ASD) is one such disease for which GWASes have struggled to find the causal genes. Microbial therapy has decreased ASD symptoms in mouse models, hinting that the microbial genes in the gut microbiome may play a significant role. Moreover, several human studies have found differences in the gut microbiomes of ASD and neurotypical children, despite being underpowered and struggling with the many confounding variables that influence gut microbiota. We present one of the largest microbiome-association studies to date across children with ASD and their neurotypical siblings, and identify 50 operational taxonomic units (OTUs) as well as 16 genera that are differentially enriched between children with ASD and their neurotypical siblings.

We performed 16s rRNA sequencing of 115 ASD-neurotypical sibling pairs over 3 timepoints, achieving an average read depth of on over 100,000 reads/sample after quality filtering. We sequenced across multiple timepoints to control for quality and stability, and took advantage of similarities of lifestyles between siblings as well as measured over 100 lifestyle related variables, such as diet, birth mode, and medication usage to control for confounders. We highlight three key findings: (1) Lifestyle factors such as diet, birth mode, and medication use can explain much of individual variation in higher level taxonomic features such as diversity and abundances of certain phyla, but (2) children with ASD and their neurotypical siblings differ with respect to which specific lower level taxa are present. We identified 50 OTUs and 16 genera differentially present in ASD vs neurotypical children, primarily composed of members of the of Lachnospiraceae, Ruminococcaceae, and Veillonellaceae families. (3) We show that pairwise statistical testing yields more powerful results than non-paired tests, and the implications for metagenome-wise association studies (MWAS), outlining a framework and bootstrapping methods for such future studies.

## PgmNr 2142: Genomic assay in autism spectrum disorders: A study of 200 cases.

### Authors:

I.O. Focsa<sup>1</sup>; S.M. Papuc<sup>2</sup>; A. Tutulan-Cunita<sup>2</sup>; F. Rad<sup>3</sup>; L. Mateescu<sup>3</sup>; R. Grozavescu<sup>3</sup>; B. Budisteanu<sup>3</sup>; C. Iliescu<sup>3</sup>; D. Birca<sup>3</sup>; C. Angheliescu<sup>3</sup>; D. Ioana<sup>3</sup>; D. Riga<sup>3</sup>; S. Riga<sup>3</sup>; I. Dobrescu<sup>3</sup>; L.C. Bohiltea<sup>1</sup>; A. Arghir<sup>2</sup>; M. Budisteanu<sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, C Davila University of Medicine, Bucharest, Romania; 2) V. Babes, National Institute of Pathology, Bucharest, Romania; 3) Prof. A. Obregia Psychiatry Hospital, Bucharest, Romania; 4) Titu Maiorescu University, Faculty of Medicine, Bucharest, Romania

---

Autism spectrum disorders (ASD) are complex neuropsychiatric conditions with an important genetic core. Over 1000 genes are estimated to be involved in outlining the pathological phenotype. Its genetics basis is extremely variable from point mutation to chromosomal abnormalities. The contribution of comparative genomic hybridization based on microarray (aCGH) in identification of small rearrangements and of new candidate genes, was valuable for defining the genetic architecture of ASD.

Our study includes 200 cases with ASD investigated by aCGH and CGH+SNP array in combination with classical cytogenetic methods (karyotype, FISH). The phenotype of our ASD children encompassed also intellectual disability, speech delay, epilepsy, dysmorphic features.

The genomic profile obtained in our patients ranged from benign to pathogenic variants. Some pathogenic variants are already associated with ASD, such as: Xq28 duplication (including *MECP2* gene), 22q11.2 deletion, 22q13.3 deletion (including *SHANK3* gene), 9q34.1 (contained *TSC1* gene) and 1q21.3q22 (including *POGZ* gene). Several rare variants were identified as well: deletion of the region 8p11.2p21.2 (only two autistic patients described so far), 9p13, 3p26.3 (including *CHL1* gene - 15 patients reported), and duplications of 2q13 (including *NPHP1* and *MALL* genes - ten patients described earlier), 4p16.1p16.3 (including *MRT54* and *PPP2C* genes- 4 patients reported), and 13q14.13q34 (3 patients have been described).

The genomic investigation in our group of patients led to delineation of disease etiology with a strong impact in management of the patients and in counselling of their families.

## PgmNr 2143: Naturally-occurring mutations in *CHD8* and *KDM6B* in rhesus macaques are associated with differences in social behavior.

### Authors:

C. Gunter<sup>1</sup>; R.A. Harris<sup>2</sup>; Z. Kovacs-Balint<sup>3</sup>; M. Raveendran<sup>2</sup>; V. Michopoulos<sup>3</sup>; J. Bachevalier<sup>3</sup>; J. Raper<sup>3,4</sup>; M. Sanchez<sup>3,5</sup>; J. Rogers<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Marcus Autism Center, Emory University School of Medicine, Atlanta, Georgia.; 2) Human Genome Sequencing Center and Dept. of Molecular and Human Genetics, Baylor College of Medicine; 3) Yerkes National Primate Research Center, Emory University; 4) Dept. of Psychology, Emory University; 5) Dept. of Psychiatry & Behavioral Sciences, Emory University School of Medicine

---

Nonhuman primates (NHPs) provide models of social behavior and social interaction that offer enormous potential for the study of autism and other behavioral disorders in humans. Rodent models are useful for investigating some mechanistic aspects of genotype-behavior relationships, but NHPs provide models of social behavior and social interaction with greater translational value. Although NHPs have long been critical to studies of developmental psychobiology, investigation of the genetic basis of intra-species variation in behavioral traits among NHPs is in its early stages. We conducted two analyses of individual variation in social behavior among juvenile rhesus macaques (*Macaca mulatta*). We used a well-established and broad behavioral ethogram as well as a novel downward extension of the macaque Social Responsiveness Scale (mSRS) to characterize social behavior among 209 juvenile macaques living in large social groups at the Yerkes National Primate Center. We performed whole exome sequencing on 91 of these animals to identify genetic variation among them. Using FaST-LMM, we then performed quantitative trait association analyses restricted to 87 genes, selected based either on published studies or high scores in SFARI Gene. A missense mutation in *CHD8* that has a CADD score of 26 (predicting a major effect on protein function) is significantly associated ( $p < 3 \times 10^{-9}$ ) with a behavioral phenotype (exploratory factor score) in which “social play” and similar variables have high loadings. This gene has been associated with risk for autism in children and is a strong candidate for direct effects on measures of sociality. A missense variant in *KDM6B* is associated with the same macaque factor score ( $p < 2 \times 10^{-8}$ ). While this gene has also been associated with autism spectrum disorder, the specific variant showing association has a low CADD score. For the *CHD8* variant, three other macaque behavioral phenotypes show p-values  $< 4 \times 10^{-6}$ . For *KDM6B* four other phenotypes produce p-values  $< 5 \times 10^{-6}$ . Two study animals each carrying both genetic variants displayed high levels of atypical behavior. These studies implicate these alleles as determinants of social behavior in macaques although do not yet provide definitive evidence for genotype-phenotype effects. Whole exome sequencing is in progress for the remainder of the 209 study animals and will provide additional information about this potential primate model of specific gene effects on complex social behavior.



## PgmNr 2144: Regional variation in transcriptional dysregulation and patterning in postmortem cerebral cortex in ASD.

### Authors:

J.R. Haney<sup>1,2,3</sup>; S. Parhami<sup>1,2,3</sup>; G. Ramaswami<sup>2,3,4</sup>; D. Alba<sup>1,2,3</sup>; C.L. Hartl<sup>2,3,4</sup>; J. Ou<sup>2,3</sup>; N.N. Parikshak<sup>2,3,4,5</sup>; V. Swarup<sup>2,3,6</sup>; M.J. Gandal<sup>1,2,3,4</sup>; D.H. Geschwind<sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, 695 Charles E. Young Drive South, Los Angeles, CA 90095; 2) Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095; 3) Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, 695 Charles E. Young Drive South, Los Angeles, CA 90095; 4) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095; 5) Department of Neurology, School of Medicine, University of California, San Francisco, 1600 Divisadero St, San Francisco, CA 94115; 6) Institute for Memory Impairments and Neurological Disorders, University of California, Irvine, 2642 Biological Sciences III, Irvine, CA 92697

---

Autism Spectrum Disorder (ASD) is a highly heritable and common neurodevelopmental disorder characterized by deficits in social communication and repetitive behaviors. Previous work has identified transcriptomic alterations in frontal and temporal cortex in ASD, characterized by the upregulation of genes expressed in astrocytes and microglia, downregulation of genes associated with synaptic activity, and attenuation of cortical patterning. Here, we expand upon this work using RNA-sequencing to identify gene expression changes in a total of 808 samples across eleven cortical brain areas from 58 subjects with ASD and 54 neurotypical controls (on average, seven regions per individual). While we replicate previous patterns of glial up-regulation and synaptic signaling down-regulation, we also observe regional gradients in the severity of the pattern, with the greatest changes occurring in the occipital cortex, followed by parietal regions and the auditory cortex. We confirm previous observations of less regional differences in gene expression in ASD, identifying pervasive deficits in patterning throughout the cortex, driven by a strong attenuation of typical anterior-posterior transcriptomic gradients in ASD. Co-expression network analysis (WGCNA) connects both widespread and region-specific ASD dysregulation to specialized biological pathways within the previously observed 'synaptic' and 'glial-immune' functional categories. These findings suggest that although the dysfunction of convergent molecular pathways is widespread in the ASD cerebral cortex, it is regionally variable, possibly reflecting an increasing gradient of severity from anterior to posterior. Whether this is due to biological or technical factors at this point cannot be determined definitively. However, we note that there is a positive relationship between the strength of the changes in a given region and neuronal density determined in previously published studies in primates.

## PgmNr 2145: Rare recurrent copy number variations in children with neurodevelopmental disorders.

### Authors:

M.E. Khan <sup>1,2</sup>; J.T. Glessner <sup>1,2</sup>; X. Chang <sup>1,2</sup>; Y. Liu <sup>1,2</sup>; N. Williams <sup>1,2</sup>; F.G. Otieno <sup>1,2</sup>; M. Garriss <sup>1,2</sup>; J. Li <sup>1,2</sup>; C. Kao <sup>1,2</sup>; P.M.A. Sleiman <sup>1,2</sup>; H. Hakonarson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia; 2) Perelman School of Medicine, University of Pennsylvania

---

Neurodevelopmental disorders (NDDs) such as attention deficit hyperactivity disorder (ADHD) and autism spectrum disorder (ASD) are illustrative examples of clinical cases exhibiting phenotypes that often undergo exhaustive diagnostic odyssey due to lack of definitive corroborating genetic information. ASD and ADHD have complex genetic associations implicated by rare recurrent copy number variations (CNVs). Diseases associated with similar organ impairments are hypothesized to share similar biological etiologies. Platforms aimed at investigating genetic-based associations, such as high-density microarray technologies, have allowed for the application of groundbreaking techniques in fields aimed at elucidating the underlying disease biology associated with NDDs and related psychiatric diseases. Previous studies investigating NDDs have uncovered CNVs associated with genes within shared candidate genomic networks. In this study we investigated 1,400 ADHD and 1,059 ASD cases using 7,526 controls with the study design of identifying CNVs that increase susceptibility in both diseases in order to better elucidate shared disease biology across NDDs. Cases and controls were matched for Illumina chip subversions and observed vs. expected frequency in association was calculated. Quality control through visual inspection of genotype and intensity was undertaken to ensure CNV confidence prior to association studies. In order to assess the implication of metabotropic glutamate receptor (mGluR) network functions underlying the disease biology of NDDs, we exhaustively queried patients with ASD and ADHD for CNVs associated with 273 genomic regions of interest within the mGluR network. We saw CNTN4 deletions in 190 cases and 50 controls ( $P = 8.15E-73$ ,  $OR = 11.91$ ). Additionally we saw NPY2R deletions in 10 cases and 0 controls ( $P = 8.15E-7$ ,  $OR = \text{infinity}$ ). Together, these results suggest that disruption in neuronal cell-adhesion pathways confers significant risk to multiple NDDs and showcase that rare recurrent CNVs in CNTN4 and NPY2R are overrepresented in both ADHD and ASD cases.

## PgmNr 2146: Genome wide discovery of transcriptional targets by *FOXN1* unveils regulation of previously associated gene candidates for ASD.

### Authors:

C. Lattig<sup>1,2</sup>; O.A. Moreno-Ramos<sup>1</sup>; D. Mahecha<sup>7</sup>; A. Lopez<sup>2</sup>; Q. Gan<sup>4</sup>; A.M. Montaña<sup>4,5</sup>; D.L. Nuñez<sup>1</sup>; R. Chaskel<sup>3,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Facultad de Ciencias -Universidad de los Andes, Bogotá, Colombia; 2) SIGEN alianza Universidad de los Andes - Fundación Santa Fe de Bogotá, Bogotá, Colombia; 3) Instituto del Sistema Nervioso Central – Clínica Montserrat, Bogotá, Colombia;; 4) Department of Pediatrics, School of Medicine, Saint Louis University, St. Louis, MO, USA; 5) Department of Biochemistry and Molecular Biology, School of Medicine, Saint Louis University, St. Louis, MO, USA.1; 6) Facultad de Medicina – Universidad de los Andes, Bogota, Colombia; 7) Fundación Santa Fe de Bogotá, Bogotá, Colombia;

---

Autism spectrum disorders (ASD) are neurodevelopmental disorders that are characterized by difficulties in communication, social interactions and stereotyped behaviors. ASD has a heritability of 64 – 91% indicating a high genetic component. We previously reported a rare *de novo* mutation in *FOXN1* gene in an individual with autism through exome sequencing. *FOXN1* gene encodes a transcription factor essential for the development of the thymus which knockout mouse is known as Nude/SCID. Not much information is known regarding its expression in human brain except for two human SCID fetuses reported with severe neural tube defects such as anencephaly and spina bifida. We used ChIP-Seq technique to identify possible genome-wide targets of the transcriptional factor *FOXN1* in CNS tissue from mouse embryonic brain (E13) and adult brain piriform cortex (P40). Control samples, not treated with anti-*FOXN1* antibody, were also sequenced for both tissues. Data was processed according to the guidelines for analysis of ChIP-Seq data by Bailey, *et al.* Sequences in DNA from embryonic mouse tissue include target genes that have been associated with ASD such as: contactin associated protein, thioredoxin-related transmembrane protein and vomeronasal receptor. Transcription factor DNA motifs for *FOXN1* include high homology for the *FOXP1* (p=0,0001) and *FOXK2* (p=0,001) genes as well as other members of the Forkhead family of transcription factors. The findings in this study confirm that *FOXN1* plays an important role in the development of the CNS during its early stages and will enable a better understanding of early stage neural pathways involved in ASD.

## PgmNr 2147: Copy number variations in autism spectrum disorder and specific language impairment.

### Authors:

V. Mahaganapathy <sup>1</sup>; J.F. Flax <sup>1</sup>; J. Xing <sup>1</sup>; C. Gwin <sup>2</sup>; S. Wilson <sup>1</sup>; C.W. Bartlett <sup>3</sup>; R. Thomas <sup>1</sup>; A. Zhou <sup>1</sup>; L.M. Brzustowicz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Department of Biomaterials, Rutgers University, Piscataway, NJ; 3) The Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital and Department of Pediatrics, The Ohio State University, Columbus, OH

---

Autism spectrum disorder (ASD) is a collection of neurodevelopmental disorders characterized by restrictive interests and/or repetitive behaviors and deficits in social interaction and communication. Studies have shown that individuals with ASD have a higher global burden of rare copy number variations. We have previously demonstrated a shared genetic etiology for autism and a specific language impairment (SLI) phenotype characterized by individuals who do not show other neurological or psychological impairments. We conducted an analysis of copy number variations (CNV) in our familial dataset collected as part of the New Jersey Language and Autism Study (NJLAGS). We have SNP genotyping microarray data for 570 individuals genotyped with Affymetrix and Illumina arraycards. The individuals are spread across 117 families with a median n of 4 members per family. The data consists of individuals affected by ASD, individuals affected by SLI and unaffected individuals. Comparative studies on algorithms to detect copy number variations have shown that there is greater variability but lower concordance and reproducibility among algorithms. This necessitates careful assessment of experimental design and stringent data curation in discovery studies. We therefore employ multiple algorithms (PennCNV, QuantiSNP, iPattern, GenomeStudio, Axiom CNV Summary Tools) to analyze intersection of results from at least 3 CNV callers for each sample. Extensive quality control with strict cutoffs for parameters like CNV length (10kb - 7.5 Mb) and probe count (>4) is carried out to investigate CNVs detectable by SNP genotyping arrays. Using families allows us to identify de-novo CNV calls that are observed only in ASD affected sibling and not in the unaffected siblings. Our preliminary results indicate greater number of CNVs in our ASD affected individuals (mean n=16.6) when compared to unaffected individuals (mean n=15.1). We also observe a greater number of CNVs in ASD affected individuals compared to SLI affected individuals (mean n=14.1). These results are statistically significant with p values 0.046 and 0.007 respectively. We did not see a statistically significant difference in number of CNVs in SLI affected and unaffected individuals. We have also identified previously reported CNVs associated with ASD suggesting that our samples share significant genetic etiology with previously reported samples.

## **PgmNr 2148: Non-additive genetic effects in autism spectrum disorder and specific language impairment on reading achievement and symptoms of the broader autism phenotype.**

### **Authors:**

L.L. Mason <sup>1</sup>; S.I. Gross <sup>1</sup>; C.W. Bartlett <sup>2</sup>; J.F. Flax <sup>3</sup>; S.A. Petrill <sup>1</sup>; L.M. Brzustowicz <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Psychology, The Ohio State University, Columbus, OH; 2) The Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 3) Department of Genetics, Rutgers University, Piscataway, NJ

We examined traits associated with Autism Spectrum Disorder (ASD) and Specific Language Impairment (SLI) using quantitative genetic and linkage analysis to determine the effects of these disorders on reading achievement and social aptitude.

Nuclear and extended families (N=179) in which at least 1 member had a diagnosis of ASD and at least 1 member had a diagnosis of SLI were recruited as part of the New Jersey Language and Autism Genetics Study and given behavioral rating scales on a variety of topics, including the Social Responsiveness Scale, Second Edition (SRS-2; Constantino, 2012). Social subscales of interest were those of awareness, cognition, communication, motivation, and restricted interests and repetitive behaviors. Reading data were also collected using the Gray Oral Reading Test-Fourth Edition (GORT-4; Wiederholt & Bryant, 2001), specifically within subdomains of reading rate, accuracy, fluency, and comprehension.

Heritability ( $h^2$ ) estimates present the proportion of genetic variation that influences phenotypic variation within a specific group in a specific environment. Univariate heritabilities and bivariate genetic correlations were modeled with the Sequential Oligogenic Linkage Analysis Routines package v4.3.1 (Almasy & Blangero, 1998). We compared the heritability estimates of ASD symptoms and reading in the full sample (N=687), containing families with both ASD and SLI, to heritability estimates when participants with ASD were statistically removed. In a comparison analysis, individuals with SLI were statistically removed.

Heritability of the SRS-2 subscales significantly changed when participants with ASD were removed, while few changes were apparent when participants with SLI were removed. This may be evidence for non-additive genetic effects on social aspects of ASD, specifically for awareness ( $\Delta h^2_r=31.6\%$  to  $37.0\%$ ,  $p<.01$ ), communication ( $\Delta h^2_r=16.7\%$  to  $24.8\%$ ,  $p<.05$ ), and restricted interests and repetitive behaviors ( $\Delta h^2_r=30.9\%$  to  $38.7\%$ ,  $p<.01$ ). Few substantial changes in heritability were found on the GORT-4 subscales when either participants with ASD were removed or participants with SLI were removed. Notably, however, a non-linear effect was observed in oral reading rate when SLI participants were removed from the sample, supporting the influence of language impairment on this trait ( $\Delta h^2_r=43.7\%$  to  $52.8\%$ ,  $p<.01$ ). Ongoing genetic linkage analyses investigate the relation of ASD

and SLI to reading and social ability throughout families.

## PgmNr 2149: Identification of common variants associated with risk for autism spectrum disorder in the SPARK dataset.

### Authors:

N. Matoba <sup>1,2</sup>; H. Qian <sup>3</sup>; L.M. Raffield <sup>1</sup>; Y. Li <sup>1,4,5</sup>; H. Won <sup>1,2</sup>; J.L. Stein <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of North Carolina at Chapel Hill; 2) UNC Neuroscience Center, University of North Carolina at Chapel Hill; 3) Department of Statistics and Operations Research, University of North Carolina at Chapel Hill; 4) Department of Biostatistics, University of North Carolina at Chapel Hill; 5) Department of Computer Science, University of North Carolina at Chapel Hill

---

Autism spectrum disorder (**ASD**) is a highly heritable neurodevelopmental disorder characterized by deficits in social interaction and communication as well as repetitive and restrictive behaviors. Although current research has identified hundreds of *de novo*, likely gene-disrupting ASD rare variants, these variants do not explain a large component of inherited risk. Recently, a genome-wide association study (GWAS) including 18,000 individuals with ASD identified five ASD susceptibility loci (Grove *et al*, *Nat. Genet.* 2019). Given SNP-based heritability estimates of around 50% (Gaugler *et al.*, *Nat. Genet.* 2014), many more common variants are expected to impact risk for ASD. Identification of additional common genetic risk factors may aid understanding of biological mechanisms underlying risk for this behaviorally defined disorder.

In order to detect additional common variants associated with ASD risk, and to replicate previously identified associations in an independent dataset, here we perform a GWAS study on 6,376 trios from **SPARK** (Simons Foundation Powering Autism Research for Knowledge) genotyped on a Illumina Infinium Global Screening Array. Since most participants in SPARK are families (35% and 29% are complete quads and trios, respectively), we applied a case-pseudocontrol strategy which compares allele frequencies between transmitted versus untransmitted alleles to a proband from his/her parents. Genotype imputation was performed on Michigan Imputation Server using TOPMed reference panel.

The primary SPARK GWAS identified one novel genome-wide significant (GWS) locus (rs552214748,  $P < 5e-8$ ) and one suggestive locus (rs16887378,  $P < 1e-6$ ). In particular, rs16887378 (near the gene **FGFR1**) overlaps with a GWS locus identified from a cross-disorder meta-analysis of ASD and schizophrenia (The Autism Spectrum Disorders Working Group, *Mol. Autism.* 2017). This signal was replicated in Grove *et al* ( $P = 0.007$ ; OR=0.96), although the combined p-value after meta-analysis did not reach genome-wide significance ( $P = 1.9e-6$ ). Meta-analysis with Grove *et al.* identified four additional loci including a **KANSL1** locus that was previously found by a more lenient gene-based analysis (Grove *et al.*, *Nat. Genet.* 2019).

Collectively, SPARK GWAS recapitulates previously reported ASD GWAS signals. We now work on the functional interpretation of these findings using chromatin interaction, chromatin accessibility, expression, and brain structure QTL data.



## PgmNr 2150: Scoring neurotypicals for autism polygenic risk reveals the underlying genetic architecture of the autism spectrum.

### Authors:

V. Pillalamarri<sup>1,2,3</sup>; S.Y. Yang<sup>1,2</sup>; D.E. Arking<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine; 2) Predoctoral Program in Human Genetics, Johns Hopkins University School of Medicine; 3) Maryland Genetics Epidemiology and Medicine Training Program

Common inherited variation explains the majority of liability for autism spectrum disorder (ASD), a phenotypically heterogeneous neurodevelopmental disorder of increasing public health urgency. As such, many ASD risk alleles are also found within the general population of unaffected individuals ("neurotypicals") and contribute to a polygenic architecture composed of a multitude of common variants each having a small effect towards modulating overall ASD liability under a threshold model that includes rare and *de novo* mutations.

We reasoned that studying the effects of common, polygenic risk for ASD in neurotypicals would uncover shared mechanisms of disease amongst the overlapping ASD endophenotypes and comorbid conditions. Specifically, we hypothesized that downstream molecular targets of common, polygenic risk for ASD will have altered gene expression in neurotypicals in a tissue-specific manner.

Here, we estimated polygenic risk score (PRS) effects on gene expression to identify ASD PRS expression quantitative trait loci (PRS-eQTL) within the Genotype-Tissue Expression (GTEx) cohort. European-specific effect size estimates from a recent ASD GWAS (Grove 2019) were used to score neurotypicals of European-American descent (n=527). Normalized RNA-Seq expression estimates from post-mortem tissue samples were then modeled as a function of the ASD PRS along with confounding covariates, including latent factors captured through probabilistic estimation of expression residuals (PEER).

A tissue-by-tissue analysis of 48 tissues revealed nine tissues that had  $\geq 1$  ASD-specific PRS-eQTL at a tissue-specific FDR  $\leq 10\%$ , including three tissues sampled from the brain (cortex, caudate basal ganglia, and substantia nigra). Five genes were associated with an ASD PRS at tissue-specific Bonferroni significance, including an antisense RNA *RP11-342K6.3* ( $p = 2.07 \times 10^{-6}$ ;  $q\text{-value} = 0.045$ ; effect size = -0.60 normalized units of expression) in brain substantia nigra (SN) at the *GCFC2* locus, which has been previously associated with dyslexia and language impairment and replicated in ASD. A random effects meta-analysis of an ASD PRS effect on *GCFC2* expression across 13 brain regions revealed that the concomitant effects on *GCFC2* expression were in the opposite direction for all regions and the strongest in the SN. Combined, these results suggest the presence of a shared ASD genetic etiology and molecular mechanisms of disease driven by common inherited variation.



## PgmNr 2151: Autism spectrum disorder in the Amish: Exome sequencing unveils a novel missense variant in EvC ciliary complex subunit 1 (*EVC*), a known regulator of the sonic hedgehog signaling pathway.

### Authors:

C.G. Tise<sup>1,3</sup>; J.W. Kleinberger<sup>1</sup>; M.A. Pavlovich<sup>1</sup>; M.L. Daue<sup>1</sup>; D.P. Loesch<sup>1</sup>; J.R. O'Connell<sup>1</sup>; J.A. Perry<sup>1</sup>; L.M. Yerges-Armstrong<sup>1</sup>; A.R. Shuldiner<sup>1,2</sup>; N.A. Zaghoul<sup>1</sup>; Regeneron Genetics Center

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Maryland SOM, Baltimore, MD; 2) Regeneron Genetics Center LLC, Tarrytown, NY; 3) Stanford University SOM, Stanford, CA

**Purpose:** To date, sequencing efforts to identify genetic determinants of autism spectrum disorder (ASD) have focused largely on *de novo* copy number variants and loss-of-function (LOF) variants in severely affected children. In this study, we evaluated exonic variants for association with quantitative measures of social impairment within the autism spectrum among adult subjects.

**Methods:** The Social Responsiveness Scale™, Second Edition (SRS-2), Adult Self-Report Form was administered to 357 generally healthy Amish adult research participants of the University of Maryland Amish Research Clinic. The SRS-2 is an objective measure of symptoms associated with ASD, with a higher Total T-score correlating with increased severity. The SRS-2 also generates scores for five treatment and two DSM-5-compatible subscales. Using exome sequencing data, an exome-wide association analysis (ExWAS) of each SRS-2 score was performed using the MMAP program, adjusting for age, sex, and relatedness. Results were limited to exonic variants in HWE ( $P \geq 0.05$ ) with a MAC  $\geq 5$ .

**Results:** SRS-2 Total T-Score ranged from 36-75 with a mean of  $48.9 \pm 6.4$  (SD). A missense variant in EvC ciliary complex subunit 1 (*EVC* c.550G>A, p.D184N, rs41269549; MAF=0.7%) showed the strongest association with Total T-Score, reaching near exome-wide significance ( $\beta=14.5$ ,  $P=2.1 \times 10^{-6}$ ). In addition, *EVC* p.D184N demonstrated exome-wide significant associations with scores for Repetitive and Restrictive Behaviors (RRB,  $\beta=15.8$ ,  $P=3.3 \times 10^{-8}$ ) and Social Motivation (Mot,  $\beta=16.3$ ,  $P=7.7 \times 10^{-7}$ ). LOF variants in *EVC* are associated with Weyers acrofacial dysostosis and Ellis-van Creveld Syndrome, for which ASD is not a characteristic feature; however, the clinical significance of *EVC* p.D184N remains uncertain in ClinVar. *EVC* regulates the Sonic hedgehog signaling pathway which is involved in body and CNS patterning, providing mechanistic plausibility for this novel association.

**Conclusions:** We evaluated adults across the full spectrum of ASD-related behavioral traits, and through ExWAS, identified a germ-line variant in *EVC* associated with SRS-2 Total T-Score, as well as RRB and Mot scores. Possible explanations for this association include pleiotropy, incomplete dominance, and/or *EVC* p.D184N representing a gain-of-function variant. This work has potential to identify disease pathways and therapeutic targets for ASD, while highlighting the power of founder populations and the importance of deep phenotyping for gene discovery.

## PgmNr 2152: Whole genome sequencing and homozygosity mapping in a consanguineous cohort identifies new candidate genes for autism spectrum disorder.

### Authors:

I.O. Tuncay<sup>1</sup>; N. Parmalee<sup>2</sup>; B. Trost<sup>3</sup>; M. Zarrei<sup>3</sup>; J. Howe<sup>3</sup>; S. Scherer<sup>3</sup>; M.H. Chahrour<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neuroscience, The University of Texas Southwestern Medical Center, Dallas, TX.; 2) Eugene McDermott Center for Human Growth and Development, The University of Texas Southwestern Medical Center, Dallas, TX.; 3) The Centre for Applied Genomics, Hospital for Sick Children and McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 4) Department of Psychiatry, The University of Texas Southwestern Medical Center, Dallas, TX.

---

Autism spectrum disorder (ASD) is a neurodevelopmental disorder characterized by deficits in social communication and restricted, repetitive patterns of behavior or interests. The current estimate of prevalence for ASD is 1 in 59. ASD is one of the most highly heritable neuropsychiatric conditions, but is extremely genetically heterogeneous, reducing the power to identify causative genes. Here we utilized the enriched homozygosity in a consanguineous cohort with ASD to identify disease variants. In collaboration with the MSSNG Consortium we performed whole genome sequencing (WGS) in 68 individuals from 22 families. On average, each individual had 3.6-4.4 million variants after filtering for genotype quality  $\geq 99$  and read depth  $\geq 10$ , 160,000-220,000 of which were rare ( $MAF \leq 1\%$  in 1000 Genomes, ExAC, and gnomAD databases). Out of these rare variants, each proband had an average of 1,305 homozygous, 17 compound heterozygous, 164 X-linked, and 25,838 *de novo* variants. Focusing on coding nonsynonymous variants left 0-17 homozygous, 6-30 compound heterozygous, 0-6 X-linked, and 7-53 *de novo* variants per proband. Next, we identified runs of homozygosity (ROHs) in probands using a sliding window analysis of common SNPs ( $MAF \geq 5\%$ ) in PLINK, allowing for no more than 3 heterozygous SNPs in every 10 consecutive SNPs. We restricted our analysis to ROHs with  $\geq 75\%$  homozygosity. The largest ROH detected was over 17 Mb, consistent with expectations for multiple generations of consanguineous unions. For Sanger validation we selected rare coding nonsynonymous variants that were located either within an ROH or present in a SFARI-annotated ASD gene. Out of the 59 candidate variants (0-7 variants for each proband), 40 were in 35 SFARI-annotated ASD genes including *ANK2*, *ARID1B*, *SHANK1*, *SHANK3*, and *VPS13B*. The remaining 19 novel candidates included genes associated with neurological disorders such as epilepsy (*XIRP2*), intellectual disability (*ZNF721*), and Joubert syndrome (*CC2D2B*, *PIBF1*), as well as genes involved in cytoskeletal (*KIF26A*, *NEK3*) and transmembrane (*ABCC12*, *GPR142*) signaling, mitosis (*ANAPC1*, *CDK10*), and protein synthesis (*AARS*, *SMG7*). We combined homozygosity mapping with WGS to identify candidate ASD variants. Our data illuminate important biological pathways underlying ASD and emphasize the importance of leveraging shared ancestry to map disease variants in complex neurodevelopmental disorders.

## PgmNr 2153: The influence of common and rare variation on risk for autism spectrum disorders.

### Authors:

C. van der Merwe<sup>1,2</sup>; C. Carey<sup>1,2</sup>; E. Wigdor<sup>1</sup>; E. Robinson<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanley Center for Psychiatric Research, Broad Institute, Boston, MA.; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, MA, United States; 3) Department of Epidemiology, Harvard T.H. Chan School of Public Health, MA, United States

---

Heritability estimates for autism spectrum disorder (ASD) range between 64-91%, a substantial fraction of which is explained by common polygenic variation. Parent to child transmission of polygenic risk for ASD is highly complex, and may depend on factors such as cognition and sex of the affected individual. Here, we aim to understand common polygenic transmission in subgroups of individuals with ASD, and its affect on ASD heterogeneity.

Polygenic risk scores (PRS) for four traits (ASD, schizophrenia (SCZ), educational attainment (EA), and ADHD) were estimated for ASD probands and their unaffected parents in three independent cohorts – the Simons Simplex Collection (SSC, N=2580), Psychiatric Genomics Consortium (PGC)-AGRE (N=1478), and PGC-TASC (N=807). PRS were also calculated for unaffected siblings from the SSC cohort. We used the polygenic transmission disequilibrium test (pTDT) to examine the influence of each PRS in ASD cases. The difference between proband and average parent PRS scores was calculated for each affected child across the four traits, and split by sex and IQ. Overtransmission of a PRS is observed when the mean proband PRS exceeds the mean PRS of the probands' parents, and suggests that the common polygenic variation indexed by a PRS has contributed to risk of case status.

We found significant ( $p < 1 \times 10^{-7}$ ) overtransmission of common polygenic risk for ASD in all three cohorts, as previously reported. Consistent and significant overtransmission was also found for common polygenic risk for SCZ. For the first time, we found significant points of heterogeneity in common variant influence. In all cohorts, overtransmission of the ASD PRS to female probands was between 1.7- 2.0 times that of male probands, consistent with a female protective effect against ASD. While overtransmission of the EA PRS was highly significant in SSC ( $p = 1.84 \times 10^{-10}$ ) and PGC-TASC ( $p = 3.47 \times 10^{-7}$ ), there was no evidence of this in PGC-AGRE. Further analyses will be necessary to determine whether these between cohort effects could relate to AGRE being a multiplex cohort and SSC and TASC being simplex cohorts. Further, average polygenic influence differed by probands' mean level of cognition.

Addressing the heterogeneity of common polygenic risk for ASD will allow it to be used more profitably in downstream analyses. It will also highlight groups of ASD cases more etiologically similar, nominating potentially more homogeneous groups for treatment trials.

## PgmNr 2154: Affected sib-pair analyses identify signaling networks associated with social behavioral deficits in autism.

### Authors:

T. Wang <sup>1</sup>; T. Niranjan <sup>1</sup>; Y. Chen <sup>2</sup>; I. Tunc <sup>2</sup>; F. Goes <sup>3</sup>; D. Avramopoulos <sup>1</sup>; J. Potash <sup>3</sup>; R. Huganir <sup>4</sup>; P. Zandi <sup>3</sup>; M. Pirooznia <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine; 2) Bioinformatics and Computational Biology Core Facility, National Heart Lung and Blood Institute, NIH; 3) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine; 4) Department of Neuroscience, Johns Hopkins University School of Medicine

---

Autism spectrum disorders (ASDs) are characterized by deficits in three core behavioral domains: reciprocal social interactions, communication, and restricted interests and/or repetitive behaviors. Several hundreds of risk genes for autism have been identified, however, it remains a challenge to associate these genes with specific core behavioral deficits. In multiplex autism families, affected sibs often show significant differences in severity of individual core phenotypes. We hypothesize that a higher mutation burden contributes to a larger difference in the severity of specific core phenotypes between affected sibs. We tested this hypothesis on social behavioral deficits in autism. We sequenced synaptome genes (n=1,886) in affected male sib-pairs (n=274) in families from the Autism Genetics Research Exchange (AGRE) and identified rare (MAF $\leq$ 1%) and predicted functional variants. We selected affected sib-pairs with a large ( $\geq$ 10; n=92 pairs) or a small ( $\leq$ 4; n=108 pairs) difference in total cumulative Autism Diagnostic Interview-Revised (ADI-R) social scores (SOCT\_CS). We compared burdens of unshared variants present only in sibs with severe social deficits and found a higher burden in SOCT\_CS $\geq$ 10 compared to SOCT\_CS $\leq$ 4 (SOCT\_CS $\geq$ 10: 705.1 $\pm$ 16.2; SOCT\_CS $\leq$ 4, 668.3 $\pm$ 9.0;  $p=0.025$ ). Unshared SOCT\_CS $\geq$ 10 genes presented only in sibs with severe social deficits are significantly enriched in the SFARI gene set. Network analyses of these genes using InWeb\_IM, MSigDB, and GeNetMeta identified a significant enrichment for PI3K-AKT-mTOR (eScore P-value=3.36E-07; n=8 genes) and NGF (eScore P-value =8.94E-07; n=9 genes) networks. These studies support a key role for these signaling networks in social behavioral deficits and present a novel approach to associate risk genes and signaling networks with core behavioral domains in autism.

## PgmNr 2155: Phenotype-to-genotype approach reveals head-circumference-associated genes in a Chinese autism cohort.

### Authors:

H. Wu <sup>1</sup>; H. Li <sup>2</sup>; T. Bai <sup>1</sup>; K. Xia <sup>1</sup>; H. Guo <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Medical Genetics & Hunan Provincial Key, Central South University, Changsha, Hunan, China; 2) Key Laboratory of Developmental Disorders in Children, Liuzhou Maternity and Child Healthcare Hospital, Liuzhou, Guangxi, China

---

**Background** The genotype-first approach has been successfully applied and has elucidated several subtypes of autism spectrum disorder (ASD). However, it requires very large cohorts because of the extensive genetic heterogeneity. We investigate the alternate possibility of whether phenotype-specific genes can be identified from a small group of patients with specific phenotype(s).

**Methods** To identify novel genes associated with ASD and abnormal head circumference using a phenotype-to-genotype approach, we performed whole-exome sequencing on 93 families with ASD and abnormal head circumference.

**Results** Clinically relevant pathogenic or likely pathogenic variants account for 25.4% of patients with microcephaly or macrocephaly, and 76.5% of those variants or genes are head-size associated. Significantly, recurrent pathogenic mutations were identified in two macrocephaly genes (*PTEN*, *CHD8*) in this small cohort. *GIGYF1* was implicated as a novel gene associated with the increased head circumference. *De novo* mutations in several candidate genes (*UBN2*, *BIRC6*, *SYNE1*, and *KCNMA1*) were detected, as well as two new candidate genes (*TNPO3*, *PPP2R2B*) implicated in ASD and related neurodevelopmental disorders.

**Conclusions** We identify genotype-phenotype correlations for head-size-associated ASD genes and novel candidate genes for further investigation. Our results also suggest a phenotype-to-genotype strategy would accelerate the elucidation of genotype-phenotype relationships for ASD by using phenotype-restricted cohorts.

## PgmNr 2156: A large data resource of genomic copy number variation across neurodevelopmental disorders.

### Authors:

M. Zarrei <sup>1,2</sup>; C. Burton <sup>3</sup>; W. Engchuan <sup>1,2</sup>; E. Young <sup>4</sup>; E. Higginbotham <sup>1,2</sup>; B. Trost <sup>1,2</sup>; A. Chan <sup>1,2</sup>; G. Costain <sup>5,6</sup>; D. Merico <sup>1,7</sup>; M. Woodbury-Smith <sup>1,8</sup>; J. Howe <sup>1</sup>; R. Yuen <sup>1,2</sup>; J. Buchanan <sup>1</sup>; C. Marshall <sup>1,4,9</sup>; L. Zwaigenbaum <sup>10</sup>; M. Elsabbagh <sup>11</sup>; B. Fernandez <sup>12</sup>; M. Carter <sup>13</sup>; P. Szatmari <sup>14,15,16</sup>; W. Roberts <sup>17</sup>; X. Liu <sup>18</sup>; R. Nicolson <sup>19,20</sup>; R. Weksberg <sup>2,5,21</sup>; P. Arnold <sup>2,22,23</sup>; A. Bassett <sup>14,24,25</sup>; J. Crosbie <sup>3,14</sup>; R. Schachar <sup>3,14,26</sup>; D. Stavropoulos <sup>4</sup>; E. Anagnostou <sup>27</sup>; S. Scherer <sup>1,2,28</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 3) Neurosciences and Mental Health Program, The Hospital for Sick Children, Toronto, ON, Canada; 4) Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 5) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 6) Medical Genetics Residency Training Program, University of Toronto, Toronto, ON, Canada; 7) Deep Genomics Inc., Toronto, ON, Canada; 8) Institute of Neuroscience, Newcastle University, Newcastle upon Tyne, UK; 9) Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON, Canada; 10) Department of Pediatrics, University of Alberta, Edmonton, AB, Canada; 11) Department of Psychiatry, McGill University, Montreal, QC, Canada; 12) Discipline of Genetics, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada; 13) Regional Genetics Program, The Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 14) Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 15) Centre for Addiction and Mental Health, Toronto, ON, Canada; 16) Department of Psychiatry, The Hospital for Sick Children, Toronto, ON, Canada; 17) Autism Research Unit, The Hospital for Sick Children, Toronto, ON, Canada; 18) Department of Psychiatry, Queen's University, Kingston, ON, Canada; 19) Children's Health Research Institute, London, ON, Canada; 20) Western University, London, ON, Canada; 21) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 22) Mathison Centre for Mental Health Research and Education, University of Calgary, Calgary, AB, Canada; 23) Departments of Psychiatry and Medical Genetics, Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada; 24) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, ON, Canada; 25) The Dalglish Family 22q Clinic, Toronto General Hospital, Toronto, ON, Canada; 26) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 27) Holland Bloorview Kids Rehabilitation Hospital, University of Toronto, Toronto, ON, Canada; 28) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, ON, Canada

---

Copy number variations (CNVs) are implicated across many neurodevelopmental disorders (NDDs) and contribute to their shared genetic etiology. Multiple studies have attempted to identify shared etiology among NDDs, but this is the first genome-wide CNV analysis across autism spectrum disorder (ASD), attention deficit hyperactivity disorder (ADHD), schizophrenia (SCZ), and obsessive-compulsive disorder (OCD) at once. Using microarray (Affymetrix CytoScan HD), we genotyped 2,691 subjects diagnosed with an NDD (204 SCZ, 1,838 ASD, 427 ADHD and 222 OCD) and 1,769 family members, mainly parents, from the province of Ontario, Canada. We identified rare CNVs, defined as those found in less than 0.1% of 10,851 population control samples. We found clinically relevant CNVs

(broadly defined) in 284 (10.5%) of total subjects including 22 (10.8%) among subjects with SCZ, 209 (11.4%) with ASD, 40 (9.4%) with ADHD, and 13 (5.6%) with OCD. Among all NDD subjects, we identified 17 (0.63%) with aneuploidies and 115 (4.3%) with known genomic disorder variants. We searched further for genes impacted by different CNVs in multiple disorders. Examples of NDD-associated genes linked across more than one disorder (listed in order of occurrence) are *NRXN1*, *SEH1L*, *LDLRAD4*, *GNAL*, *GNG13*, *MKRN1*, *DCTN2*, *KNDC1*, *PCMTD2*, *KIF5A*, *SYNM*, and long non-coding RNAs: *AK127244*, *NRON*, and *PTCHD1-AS*. We demonstrated that identical CNVs or genes could potentially contribute to the etiology of multiple NDDs. The CNVs identified will serve as a useful resource for both research and diagnostic laboratories for prioritization of variants and clinical interpretations.



## PgmNr 2157: Rare variants involved in risk of suicide death identified from targeted genome-wide array analysis.

### Authors:

E. DiBlasi<sup>1</sup>; Q.S. Li<sup>2</sup>; A.A. Shabalin<sup>1</sup>; J.S. Anderson<sup>1</sup>; L. Jerominski<sup>1</sup>; W.B. Callor<sup>3</sup>; E. Christensen<sup>3</sup>; B. Keeshin<sup>1</sup>; E. Ferris<sup>1</sup>; A.V. Bakian<sup>1</sup>; D. Chen<sup>1</sup>; E.T. Monson<sup>1</sup>; D. Gray<sup>1</sup>; N.J. Camp<sup>1</sup>; A.R. Docherty<sup>1</sup>; H. Coon<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Utah School of Medicine, Salt Lake City, UT; 2) Research Information Technology, Janssen Research & Development LLC, Pennington, NJ; 3) Utah State Office of the Medical Examiner, Utah Department of Health, Salt Lake City, UT

---

Suicide kills >47,000 Americans annually, and rates have risen over 33% in the last decade. With estimated heritability at 50%, only a small fraction of the genetic variance contributing to risk of suicide death has been accounted for. Here we focus on the role of rare, functionally relevant, SNPs in suicide death using the large genetic resources available in the Utah Suicide Genetic Risk Study (USGRS). The USGRS has collected >6,000 de-identified DNA samples through a two-decade collaboration with the Utah Department of Health's centralized Office of the Medical Examiner. Genome-wide data from the Infinium PsychArray were available for >4,300 population-ascertained suicide deaths. Our study involved three stages. Stage 1: we performed a targeted discovery analysis of 30,377 rare PsychArray SNPs with potential functional gene consequences, in 2,672 USGRS suicides of non-Finnish European (NFE) ancestry and >50,000 NFE controls from gnomAD. Variants were selected for analysis if they passed rigorous QC filters and if variant annotation tools (SIFT, Polyphen, CADD) predicted that the variant has a high-impact effect on gene function. Allele frequencies were compared between cases and controls using Fisher's exact test (Bonferroni adjusted significance  $P < 1.69 \times 10^{-6}$ ). Stage 2: an analysis with stage 1 cases and independent controls was used to prioritize variants further. Stage 3: an analysis with 1,522 independent Utah suicide cases of mixed ancestry and 114,707 gnomAD mixed ancestry controls was performed. We identified nine novel, high-impact, rare SNPs with significant associations with suicide death in NFE individuals in both stage 1 and stage 2. These risk variants included loss-of-function variants in *LILRB1* and *PLEKHA4* and missense variants in *SNAPC1*, *TNKS1BP1*, *ADGRF5*, *PER1*, *ESS2*, *SLC25A41* and *SPRED1*. Both *PER1* and *SNAPC1* have other supporting gene level evidence of suicide risk. The SNP in *LILRB1* also replicated in our stage 3 mixed ancestry analysis. Rare variants identified from the PsychArray were validated in sequence data. Our study suggests an important role for rare variants in suicide risk. Furthermore, we highlight the utility of genotyping arrays in rare variant discovery.



## PgmNr 2158: Genetic variants within *HCRT1* are associated with nighttime eating behavior in Southwestern American Indians.

### Authors:

C. Koroglu Altok<sup>1</sup>; M.E. Gluck<sup>2</sup>; S. Votruba<sup>2</sup>; J. Krakoff<sup>2</sup>; M. Taurig<sup>1</sup>; E.J. Stinson<sup>2</sup>; P. Chen<sup>1</sup>; C. Bogardus<sup>1,2</sup>; P. Piaggi<sup>2</sup>; L.J. Baier<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Diabetes Molecular Genetics Section, Phoenix Epidemiology and Clinical Research Branch, NIDDK, NIH, Phoenix, Arizona.; 2) Obesity and Diabetes Clinical Research Section, Phoenix Epidemiology and Clinical Research Branch, NIDDK, NIH, Phoenix, Arizona.

---

Night eating syndrome (NES) is an eating disorder whose prominent feature is excess food intake during the evening and night. Prior studies have suggested a heritable component for this disorder. Affected individuals are prone to weight gain; therefore, our study focused on identifying genetic determinants of nighttime eating (NE) behavior in American Indians who are at elevated risk for obesity. We performed whole-genome sequencing on 23 individuals with NE behavior and 27 controls, who were all from the same Southwestern American Indian community. These individuals had been clinically characterized over a 3-day period as inpatients in our Clinical Research Center, and NE status was determined based on the occurrence of food intake on any of the 3 nights between 11 pm and 5 am recorded by a computerized vending machine. Among sequence data from these individuals, 4.3 million high quality SNPs with minor allele frequencies  $\geq 0.05$  were detected. As expected with this small sample size, no variant achieved genome-wide significance for NE in a case/control analysis. Therefore, 24,125 SNPs with a nominal association with NE ( $P < 0.01$ ) were queried in the GIANT dataset ( $N \approx 235,000$  Europeans) for evidence of association with BMI. Eleven of these SNPs, which represented 2 linkage disequilibrium groups, were directionally associated with BMI in the GIANT dataset ( $P < 1 \times 10^{-4}$ , where the risk allele for NE associated with higher BMI), and all mapped to a 72 kb region on chromosome 1. Among them, 2 tag SNPs ( $r^2 = 0.46$ ) were further genotyped in an additional 32 American Indians informative for NE (13 cases and 19 controls) and 2,679 American Indians characterized for BMI. Both associated with NE in an expanded sample of 82 subjects (rs3753612:  $P = 1.5 \times 10^{-3}$ ; rs4949459:  $P = 2.7 \times 10^{-4}$ ) while only one, rs3753612 (C/G) also associated with BMI in the 2,679 American Indians ( $\beta = 0.017$ ;  $P = 7.8 \times 10^{-3}$ ). The G allele carriers were found to consume a significantly higher fraction of their total calories during nighttime ( $P = 0.013$ ). The rs3753612 variant is an eQTL for *HCRT1* (GTEx Portal), where the risk (G) allele for NE and higher BMI associated with increased *HCRT1* expression in the pituitary. *HCRT1* is involved in the regulation of sleep and food intake and is an excellent candidate gene for NE behaviour and possibly NES. We hypothesize that genetic variants in the *HCRT1* locus may lead to increased wakefulness and nocturnal eating in Southwestern American Indians.

## PgmNr 2159: Alcohol consumption and health outcomes: New evidence from Mendelian randomization of China.

### Authors:

Y. Li <sup>3</sup>; J. Wang <sup>1</sup>; T. Beatty <sup>2</sup>; G. Chen <sup>3</sup>; C. Zhu <sup>4</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Renmin University of China, Beijing, China.; 2) University of California, Davis, CA, USA.; 3) WeGene, Shenzhen, Guangdong, China.; 4) China Agricultural University, Beijing, China.

---

Alcohol consumption has been shown to be associated with more than 60 diseases, and recent debates have challenged the conventional view that low-to-moderate alcohol drinking has a protective effect on health conditions. In this study, we investigate the causal influence of drinking behavior on various health outcomes by utilizing a Mendelian randomization (MR) on the genetic variation of aldehyde dehydrogenase 2 gene (*ALDH2*). The genetic basis of the MR approach relies upon the random allocation of genes at meiosis in humans, which likely resembles the random assignment into treatment groups in randomized controlled trials (RCT).

Through collaboration with WeGene, a leading personal genetics company in China, we conducted an online survey of alcohol drinking behavior to the participants with online consent since December 2018. The latest survey contained 1,534 respondents from mainland China with drinking behavior linked to genotyping data. A special variant of the *ALDH2* genotype that is common in Chinese (affecting about 35-45% of East Asians, but less than 1% of Caucasians) is known to cause the alcohol flushing response and adverse alcohol reaction (e.g. nausea, headache after drinking) due to decreased acetaldehyde metabolism, thereby plays a strong protective role against heavy drinking behavior and can be used as an instrumental variable (IV) to disentangle environmental confounders, which are otherwise very difficult to fully account for in observational studies.

Our results confirmed an adverse relationship between the presence of *ALDH2*\*2 allele (rs671) and general drinking frequency (-0.312,  $P < 0.01$ ), drinking times during the past 30 days (-1.903,  $P < 0.01$ ), binge drinking frequency (-0.968,  $P < 0.01$ ), as well as the maximum number of drinks consumed on a single occasion (-2.376,  $P < 0.01$ ). By using *ALDH2*\*2 allele as a genetic IV for drinking behavior, we found that higher alcohol consumption is causally associated with increased risks of chronic liver diseases ( $P < 0.1$ ). On the other hand, the IV point estimates of drinking behaviors on self-evaluated healthfulness, asthma, cancer, cardio cerebral diseases, diabetes, fatty liver, gout, and hypertension revealed no substantial associations. These findings were also robust with the pleiotropy-robust Mendelian randomization (PRMR) method. Our results highlighted the practicability of using genetic information to gain new insights into the long-term health outcomes of alcohol drinking behavior.

## PgmNr 2160: Mechanisms of parent-child transmission of tobacco and alcohol use with polygenic risk scores: Evidence for a genetic nurture effect.

### Authors:

G. Saunders; M. Liu; S. Vrieze; M. McGue; GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN)

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Psychology, University of Minnesota, Minneapolis, MN.

---

Parent-child substance use similarity is a function both of genetic (transmitted alleles) and environmental transmission (e.g., modeling of parent behavior). An additional possibility is that alleles not transmitted to offspring may drive parental behavior, thereby affecting the rearing environment of the child. Measuring genetic risk for substance use directly, through polygenic risk scores (PRSs), provides a way to test for an effect of the non-transmitted parental genotype on offspring outcome, termed a “genetic nurture” effect. In the current study we 1) evaluate the accuracy of tobacco and alcohol polygenic scores across developmental age and parent vs. offspring generation and 2) use these PRSs to test for a genetic nurture effect in a longitudinal, parent-offspring sample. Polygenic risk scores, with weights derived from the GWAS & Sequencing Consortium of Alcohol and Nicotine use, for smoking initiation ( $N = 1,232,091$ ), cigarettes per day ( $N = 337,334$ ), and drinks per week ( $N = 941,280$ ), were used to predict alcohol and tobacco consumption in a sample of 3,012 twins, assessed prospectively from age 17 to age 29, from the Minnesota Center for Twin and Family Research. Mixed effects models were used to test for a genetic nurture effect whereby parental PRSs predict offspring substance use after controlling for offspring’s own PRS. We find within-substance predictive accuracy of all polygenic scores (pseudo- $R^2$  ranging from 2.3%-4.1%) and cross-substance prediction for SI scores (pseudo- $R^2 = 2.2\%-3.6\%$ ). The predictive accuracy of scores remains relatively stable across adolescence and young adulthood, as well as over generation (parent vs. offspring). Parental smoking initiation PRS predicts offspring cigarettes per day at ages 17 and 24, and alcohol use at age 17, independent of shared genetics, consistent with a genetic nurture effect of parental smoking on adolescent tobacco and alcohol consumption. The parental effect attenuates to near zero after additionally controlling for parental education and occupation. Using genetic analysis tools, we find evidence of a genetic nurture effect on offspring tobacco and alcohol use in late adolescence and early adulthood. The parental genotype effect, through non-transmitted PRSs, is environmentally mediated by parental educational attainment and occupational status, suggesting a potentially causal environmental effect of rearing socioeconomic status on offspring substance use in late adolescence.

## PgmNr 2161: Investigation of the genetic modifiers of smoking-related environmental risks.

### Authors:

J. van Kampen <sup>1,2</sup>; J. Muscat <sup>1,3</sup>; J. Broach <sup>4</sup>; D. Liu <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Public Health Sciences, Pennsylvania State University College of Medicine, Hummelstown, Pennsylvania.; 2) Neural and Behavioral Sciences, Pennsylvania State University College of Medicine, Hummelstown, Pennsylvania; 3) Penn State Hershey Cancer Institute, Hummelstown, Pennsylvania; 4) Biochemistry and Molecular Biology, Pennsylvania State College of Medicine, Hummelstown, Pennsylvania

---

Smoking is considered to be the leading cause of preventable death in the United States. It has been strongly linked to the development of a variety of diseases, including heart disease, stroke, and cancer. However, there are populations of smokers, however, who do not develop such diseases. We hypothesize that individuals may contain genetic variants which modify the risk associated with smoking for the development of these diseases.

To investigate this hypothesis, we utilized the UK Biobank, a large biorepository containing phenotypic and genotypic data from ~ 500,000 individuals. We identified individuals with a history of cancers associated with smoking (lung, pharyngeal, stomach, bladder, etc.), heart disease, and stroke. The number of cigarettes per day (CPD) they smoked was used as a measure of smoking intensity. We seek to study the potential interactions between genetics and CPD, as a way to identify genetic modifiers for individuals with exposures to environmental risk factors. We performed logistic regression with the disease status as the response variable, CPD, genetic variants, the interaction term between CPD and genetic variants, age, sex, and ancestry measures as the predictors. To improve the robustness of the results in the presence of measurement errors for the environmental exposure, we utilized a sandwich-type estimator. The improved estimator yielded well-calibrated type I errors, while the standard score statistics from logistic regression produced many seemingly spurious associations.

Our results have identified genes which may modify smoking-related disease risk, such as: XKR4, a caspase-activated scramblase important for autophagy; DAPP1, a phospholipid binding protein; and PCDH20, an extracellular matrix protein in the brain. These genes modify the risk of cancer, heart disease, and stroke, respectively. Identified genes will be further analyzed to determine the most likely variants to cause such interaction effects. Our findings support the trend within the medical community toward personalized medicine, improving understanding of the variance of diseases within smoking populations. Such findings may provide insights into the etiology of these diseases, thereby identifying new therapeutic avenues for prevention.

## PgmNr 2162: Examining pleiotropic factors that jointly contribute to anxiety and depressive phenotypes in the UK biobank data.

### Authors:

B. Verhulst <sup>1</sup>; J. Hettema <sup>3</sup>; J. Pritikin <sup>2</sup>; M. Neale <sup>2</sup>; S. Clark <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Psychology, Michigan State University, East Lansing, Mi.; 2) Psychiatry, Virginia Commonwealth University, Richmond, Va; 3) Psychiatry, Texas A&M University, College Station Texas

---

Comorbidity between anxiety and depressive traits is substantial (Kessler et al. 2005; CDG-PGC, 2013), implying pleiotropic genetic factors may be shared between the phenotypes. To understand the both the comorbidity and the pleiotropy, we must explicitly model the correlation between the two phenotypes when testing association between a SNP and either comorbid phenotype. This type of analysis cannot be directly specified in current GWA methods as it requires a GWAS method that directly uses structural equation modeling (SEM). Accordingly, we will use GW-SEM (Verhulst et al. 2017) to estimate a two-factor latent variable model where both the anxiety and depression factors are simultaneously regressed on to each SNP, while accounting for the correlation.

We will use data from the UK biobank study to examine the comorbidity and pleiotropy for anxiety and depressive phenotypes (N ~ 375,000). In contrast with alternative GWAS methods that rely on listwise deletion and thus complete phenotypic data, the GW-SEM internally accounts for missing data. The anxiety factor will be assessed with 5 items that examine the individual's recent feelings of "annoyance or irritability", "nervousness or anxiety", "concentration", "foreboding", and "inability to stop or control worrying". The depression factor will be assessed with 4 items that assess the individual's feelings of "sadness/depression", "un-enthusiasm/disinterest", "tenseness/restlessness", and "tiredness/lethargy." Because we explicitly account for the correlation between the latent factors, it is possible to distinguish whether the SNP association contributes to the comorbidity between traits by examining the associations of each phenotype with the SNP and the change in the correlation between the phenotypes controlling for the SNP. The results for this study will be particularly useful for a more complete understanding of the comorbidity between internalizing phenotypes.

## PgmNr 2163: Prenatal alcohol exposure increases risk for ADHD after accounting for genetic liability.

### Authors:

L. Wetherill<sup>1</sup>; C. Goodlett<sup>2</sup>; F. Aliev<sup>3</sup>; J. Salvatore<sup>3</sup>; S.N. Mattson<sup>4</sup>; G. Chan<sup>5</sup>; D. Dick<sup>3</sup>; H.J. Edenberg<sup>1</sup>; V. Hesselbrock<sup>5</sup>; M. Hesselbrock<sup>5</sup>; J. Kramer<sup>6</sup>; M. Plawecki<sup>1</sup>; B. Porjesz<sup>7</sup>; E.P. Riley<sup>4</sup>; M. Schuckit<sup>8</sup>; T. Foroud<sup>1</sup>; CIFASD, COGA

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana.; 2) Psychology Dept, Indiana University Purdue University at Indianapolis, Indianapolis, IN; 3) Psychology Dept, Virginia Commonwealth University, Richmond, VA; 4) Psychology Dept, San Diego State University, San Diego, CA; 5) Psychiatry Dept, University of Connecticut, Farmington, CT; 6) Psychiatry Dept, University of Iowa, Iowa City, IA; 7) Psychiatry Dept, State University of New York (SUNY) Downstate, New York, NY; 8) Psychiatry Dept, University of California at San Diego, San Diego, CA

**INTRODUCTION:** Attention deficit hyperactivity disorder (ADHD) is a common, heritable disorder with high prevalence rates in children of alcoholics and in children with prenatal alcohol exposure (PAE). An often overlooked issue in research is the compounded risk due to the likelihood that mothers who are alcohol dependent are more likely to drink during pregnancy (and risk of prenatal alcohol exposure is increased in alcohol dependent mothers). Polygenic risk scores (PRS) for ADHD were computed using summary data from the Psychiatric Genomics Consortium (PGC) in: (1) offspring in the Collaborative Studies on the Genetics of Alcoholism (COGA, N=1,498), and (2) participants in the Collaborative Initiative on Fetal Alcohol Spectrum Disorders (CIFASD, N=366). Parental ADHD and alcohol dependence were collected as part of COGA but were not available in CIFASD.

**METHODS:** Probability of being ADHD was examined, and significant covariates were identified separately in COGA and CIFASD data. PAE was included in all analyses (COGA: # drinks during the entire pregnancy; CIFASD: heavy vs minimal/none). COGA analyses included maternal alcohol dependence (mom-AD) and mom-AD\*PAE. Family history of alcohol dependence (FHAD = yes/no) was used in CIFASD. Maternal (mom-ADHD) and paternal ADHD (dad-ADHD) were included to assess which (or both) predicted ADHD in the offspring. PRS were included in both datasets to account for genetic liability in the absence of parental diagnosis.

### RESULTS:

COGA: Dad-ADHD increased risk of ADHD, independent of PAE ( $p < 0.0001$ ). There was an increased risk if mom-AD=yes ( $p = 0.0065$ ), which increased with higher levels of PAE (PAE\*mom-AD  $p = 0.07$ ). PRS increased risk if dad-ADHD=no. If dad-ADHD=yes, the PRS increased the risk of ADHD in individuals with low levels of PAE (# drinks < 76), which was reduced at high levels of PAE (# drinks > 266).

CIFASD: Having a FHAD increased risk for ADHD ( $p = 0.045$ ). PRS ( $p = 0.088$ ) and PAE\*PRS ( $p = 0.068$ ) modestly increased the risk for ADHD. PRS increased risk for ADHD if there was no PAE, but had minimal effect on individuals with PAE.

**CONCLUSION:** In the absence of PAE: PRS, dad-ADHD and mom-AD contributed to risk for ADHD. With PAE, genetic contribution was diminished and risk associated with PAE increased. These results indicate potentially different etiologies of ADHD in individuals with and without PAE.

**ACKNOWLEDGEMENTS:** NIH U10AA008401, U01AA014809, U01AA014834, U01AA017122, U24AA014811

## **PgmNr 2164: Identification and prioritization of schizophrenia master regulators from two independent causal networks integrating genetic, transcriptomic and chromosome conformation capture data.**

### **Authors:**

N.D. Beckmann; A.W. Charney; P. Roussos; E.E. Schadt

[View Session](#) [Add to Schedule](#)

**Affiliation:** Icahn School of Medicine at Mount Sinai, New York, New York.

---

Over the last decade, genome wide association studies (GWAS) of increasing sample sizes have associated many loci to schizophrenia, with the largest study to date having found 145 loci passing genome wide significance. To better understand the etiology of this complex trait, several omics resources were recently generated. Gene expression data in large cohorts of post-mortem brain allowed generation of expression quantitative trait loci (eQTL) libraries and differential expression signatures, leading to fine-mapping for some of these loci; ChipSeq and Hi-C data have started to address epigenetic control and three dimensional (3D) structure of chromatin; and further fine-mapping is being done using gene expression imputation methods from large GWAS datasets. While these efforts have provided many results implicating central nervous system development and the immune system as components of the disease, causal chains defining disease mechanisms remain to be established.

To interpret GWAS results in the broader context of the network of gene expression and 3D structure of the genome, we used Bayesian regulatory networks (BN) to integrate genetic and gene expression datasets from two large cohorts of post-mortem samples with or without schizophrenia (Common Mind Consortium data, N = 594 and Human Brain Collection Core data, N = 386). This enabled us to model causal interactions between genes, as well as the genetic control of their expression by incorporating eQTLs and GWAS loci. To gain power to identify disease signatures, we used schizophrenia diagnosis as well as schizophrenia polygenic risk score as a continuous trait in our differential expression analyses. To focus our networks on disease-relevant signals, we used co-expression network analyses and identified modules enriched for differential expression and GWAS signatures to generate a list of input genes into our BN models. We took advantage of our flexible framework to integrate information from prior studies, including chromosome conformation capture derived topologically associated domains, to define genes and locus interactions as structure priors boosting power to infer causal relationships and increase network model accuracy. This enabled us to identify master regulators of disease signatures and of genetic loci associated to disease, and prioritize potential mechanisms of action and therapeutic targets for schizophrenia.



## PgmNr 2165: Identification of *de novo* mutations and structural variants in brain development-associated genes in Chinese schizophrenia families.

### Authors:

S. Jiang <sup>1</sup>; Y. Wang <sup>1</sup>; P. Jia <sup>1</sup>; Y. Dai <sup>1</sup>; T. Mize <sup>2,3</sup>; J. Wu <sup>2</sup>; X. Jiang <sup>4</sup>; J. Chen <sup>2</sup>; Z. Zhao <sup>1,5,6</sup>; X. Chen <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 2) Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA; 3) Department of Psychology, University of Nevada Las Vegas, Las Vegas, NV 89154, USA; 4) School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 5) MD Anderson Cancer Center UTHHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA; 6) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

---

Schizophrenia (SCZ) is a severe psychiatric disorder with a strong genetic component. It is associated with a marked reduction in fecundity, leading to the hypothesis that alleles with large effects on risk might often occur *de novo*. In this study, eight families from Taiwan, with whole-genome-sequenced SCZ-free parents, SCZ patient(s) and family-matched healthy sibling(s) and four to six family members each, were kept after kinship analysis. To reliably identify *de novo* mutations and structural variants, only *de novo* events consistently called by multiple independent software tools were retained in downstream analyses. *De novo* mutations (DNM) were confirmed by checking with Integrative Genomics Viewer and *de novo* structural variants with LowResBam2Raster visually. DNM rate per nucleotide per year was consistent with previous studies. Two stop-gain loss-of-function DNMs in genes *GJC1* and *HIST1H2AD* were identified in SCZ patients. No loss-of-function DNM was identified in healthy siblings. Because of limited occurrence of *de novo* structural variant (DNSV), only DNSVs occurred multiple times in SCZ patients but not in healthy siblings were considered detrimental DNSVs. One *de novo* deletion covering the gene *SKA3* was found in two SCZ patients but not in the matched healthy sibling in the same family. To explore if the *de novo*-implicated genes *GJC1*, *HIST1H2AD* and *SKA3* were involved in brain development, the expression data from prefrontal cortex (PFC), the most affected cerebral cortical region in SCZ, from BrainCloud (n = 269) were leveraged. We found the expression of the three genes was significantly higher in prenatal than in postnatal stages ( $P = 1.92 \times 10^{-21}$  for *GJC1*,  $P = 6.02 \times 10^{-18}$  for *HIST1H2AD* and  $P = 3.30 \times 10^{-20}$  for *SKA3*). The involvements of the three genes in brain developments were preserved across multiple cortical regions and conserved across species. To explore the functions of *de novo*-implicated genes in brain, we performed enrichment analyses by leveraging the genes coexpressed with the *de novo* target in PFC. Interestingly, the *GJC1*-coexpressed genes were found to be involved in SCZ-associated pathways and drug targets. Furthermore, the *de novo*-implicated genes were within or close to the loci enriched with GWAS signals ( $P < 1 \times 10^{-5}$ ) of SCZ, intelligence, cognitive performance and educational attainment. Our results revealed potential *de novo* mutations and structural variants in Asian population with the potential risk to schizophrenia.

## PgmNr 2166: Identification of candidate genes for alcohol use disorder by gene set enrichment and pathway analyses of RNA-seq data.

### Authors:

A.Q. Nato <sup>1</sup>; J. Denvir <sup>1</sup>; H.G. Sexton <sup>1,2</sup>; S.D. Moore <sup>3,4</sup>; D.A. Primerano <sup>1</sup>; M.-L. Risher <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Sciences, Joan C. Edwards School of Medicine, Marshall University, Huntington, WV, USA; 2) Hershel 'Woody' Williams Veteran Affairs Medical Center, Huntington WV, USA; 3) Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA; 4) Durham Veteran Affairs Medical Center, Durham, NC, USA

---

Alcohol use disorder (AUD) is a chronic psychiatric disease with a complex genetic etiology. AUD is characterized by compulsive drinking patterns that lead to damaging consequences. Excessive alcohol consumption in the form of binge drinking is common among adolescents and young adults and is associated with an increased risk of developing AUD. Using a rodent model of binge drinking (adolescent intermittent ethanol (AIE)), we identified candidate genes and pathways that may contribute to long-term changes in brain function and the development AUD. We analyzed RNA-seq data from Sprague Dawley rats (*Rattus norvegicus*) to investigate temporal variations in hippocampal gene expression. At postnatal day (PND) 30 (adolescence), rats received chronic intermittent ethanol (5g/kg intragastrically (i.g.) 10 times across 16 days). Hippocampal tissue was collected at three time points: 1) 24 hours after 4th dose (PND35), 2) 24 hours after last dose (PND46), and 3) 24 days after last dose (PND70; adult). We processed the RNA-seq data using TrimGalore and compiled gene counts using HTSeq. We employed the R/Bioconductor package DESeq2 for differential expression analysis. We performed gene set enrichment analysis (GSEA) to identify differentially regulated pathways and gene ontology terms for various comparisons based on ethanol exposure and the three time points. At PND35, genes and pathways encoding negative regulators of cellular component organization and metallopeptidases, and those involved in mitochondrial function were identified as enriched. At PND46, ethanol exposed rats showed enrichment for genes encoding potassium channels, synaptic transmission, and synaptic organization. Interestingly, immune response genes were downregulated. At PND70, ethanol exposed rats showed highly enriched gene sets involved in cytoskeletal remodeling. These results provide detailed insight into how hippocampal gene expression changes during development and in response to AIE. To identify subnetworks of genes enriched by higher interaction, we used the Markov CLustering Algorithm (MCL) as implemented by the clusterMaker app of the Cytoscape pathway analysis package on the differential expression data. We compare and contrast the results and identify genes enriched in both analyses, which may facilitate determination of the underlying mechanism for acute and long-term effects of AIE exposure that increase risk of developing AUD.

## PgmNr 2167: A trans-ethnic two-stage polygenetic scoring analysis detects genetic correlation between osteoporosis and schizophrenia.

### Authors:

F. Zhu <sup>1,2</sup>; F. Zhang <sup>3</sup>; H.W. Deng <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Translational Medicine, The First Affiliated Hospital of Xi'an Jiaotong University, Xi'an, P. R. China; 2) Stanley Center at Broad Institute, 415 Main Street, Cambridge, MA 02142, USA; 3) School of Public Health, Xi'an Jiaotong University Health Science Center, Xi'an, P. R. China; 4) Center for Bioinformatics and Genomics, Department of Global Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, LA, USA

---

The patients with schizophrenia (SCZ) have increased risk of osteoporosis. The biological mechanism remains elusive now. Our study aim to explore the genetic correlation between SCZ and osteoporosis. We conducted a trans-ethnic two-stage genetic correlation analysis of osteoporosis and SCZ, totally invoking 2,286 Caucasians subjects in discovery stage and 4124 Chinese subjects in replication stage. The BMD and bone area values of ulna & radius, hip and spine were measured using Hologic 4500W dual energy X-ray absorptiometry machine. SCZ was diagnosed according to DSM-IV criteria. Genotyping was conducted using Affymetrix SNP 6.0 array. Polygenetic risk scoring analysis was conducted by PRSice software. Multi-trait analysis of GWAS (MTAG) was performed to detect novel candidate genes for osteoporosis and SCZ. In the Caucasians discovery samples, significant genetic correlations were observed for ulna & radius BMD vs. SCZ ( $P$  value = 0.010), ulna & radius area vs. SCZ ( $P$  value = 0.031). In the Chinese replication samples, we observed significant correlation for ulna & radius area vs. SCZ ( $P$  value = 0.019). MTAG analysis identified several novel candidate genes, such as CTNNA2 (MTAG  $P$  value =  $2.24 \times 10^{-6}$ ) for SCZ and FADS2 (MTAG  $P$  value =  $2.66 \times 10^{-7}$ ) for osteoporosis. Our study results support the overlapped genetic basis for osteoporosis and SCZ, and provide novel clues for elucidating the biological mechanism of increased osteoporosis risk in SCZ patients.

## PgmNr 2168: Protective effects of N-acetylorlithine on psychiatric and neurocognitive traits.

### Authors:

J. Guo; Z. Kun; Y. Shi; Y. Tielin

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China

---

Brain disorders with extensive categories exhibit similar symptoms and share common genetic underpinnings, however, their shared biological processes remain largely unknown. Blood metabolites, which are known as functional intermediate phenotypes and capture genetic heritable traits, can be potentially used to uncover the etiology influencing brain disorders. Here, we examined the potential causal effects of 363 blood metabolites on 44 brain related traits (11 psychiatric disorders, 12 neurological disorders, 7 brain structures and 14 neurocognitive quantitative phenotypes) by using the three module of Mendelian Randomization (MR) methods: (i) The GSMR as the primary approach was performed with summary-level data from genome-wide association studies, (ii) the summary data-based MR approaches were applied for complement that comprising inverse variance-weighted (IVW) method, weighted median method, and MR-Egger, and (iii) confirmed the significant results using polygenic risk score (PRS) prediction with individual genotype data. Converging evidence showed that some metabolic exposures were causally associated with psychiatric and neurocognitive outcomes, whereas none causation appeared on neurological disorders. N-acetylorlithine, an endogenous metabolite of ornithine to participate in arginine-proline metabolism, was the most significant metabolite which had a protective effect against schizophrenia ( $P = 3.74 \times 10^{-10}$ ) with a 27% decrease in risk and 6% positively increase in intelligence ( $P = 2.48 \times 10^{-6}$ ) after adjustment of multiple testing. Further, we conducted morphology statistics of fluorescence microscopy from N-acetylorlithine supplementation in neonatal mouse cortical primary culture experiments, which indicated that N-acetylorlithine promoted neuronal differentiation with longer neurite length, more dendrite branches, higher density of spines and higher cell viability. To explore the potential molecular mechanism of N-acetylorlithine on neurodevelopment, we conducted Bayesian fine-mapping method to identify 39 causal SNPs and 3 of the 15 putative effector genes (EXOC6B, NOTO and DCTN1) that potentially act as neuron-related functional genes for N-acetylorlithine via integrating gene mapping and animal-model strategies. Our results highlight the importance of N-acetylorlithine as a predicted biomarker for brain neurodevelopmental related traits and provide a novel insight into mechanisms by which metabolites affect brain disorders.

## PgmNr 2169: Novel mutation in PPEF2 reported in a family affected with schizophrenia: Whole genome sequencing study.

### Authors:

R. Torres; S. Smieszek; E. Czeisler; C. Xiao; C. Polymeropoulos; G. Birznieks; M. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical, Vanda Pharmaceuticals, Washington, DC.

---

Schizophrenia is a complex genetic disorder with a multifactorial mode of inheritance that affects about 1% of the population. Genetic studies have implicated many different genes and pathways, but much of the genetic liability is still unaccounted for. Nonetheless, leading theories of the pathophysiology of schizophrenia include the glutamate and dopamine pathways. We performed whole genome sequencing analysis of 7 individuals diagnosed with schizophrenia and 1000 controls.

We report a novel rare stopgain in PPEF2 in a family with 3 affected individuals. The LOF PPEF2:NM\_006239:exon3:c.G135A:p.W45X has a MAF of 2.83E-05 in EXAC and has a CADD score of 37, (~1/30000 carrier).

Intriguingly, all 3 affected individuals in this pedigree share a stopgain variant in PPEF2. PPEF2 encodes a neuronally expressed protein phosphatase that binds calmodulin furthermore influencing mGluR5 levels. Specifically, it has been previously shown that HEK293 cells transfected with *PPEF2* demonstrated increased levels of mGluR5 in the presence of the mGluR5 agonist suggesting that disturbance of *PPEF2* could lower mGluR5 membrane levels. Previously a missense substitution in this gene was reported for a single affected family.

The three affected individuals are all men and developed schizophrenia early on at 9, 11, and 17 years of age. The strong genetic predisposition for schizophrenia can be seen in the early-onset of the disorder. In fact, two out of three developed schizophrenia before age 13, which it is extremely rare, with a prevalence rate of approximately 1/10,000.

Whole genome sequencing delineated a novel deleterious variant in *PPEF2* as a likely genetic risk factor for schizophrenia. Furthermore, additional genomic screens that stratify patients based on the affected genetic pathways may better guide the most effective treatments.

## PgmNr 2170: Molecular characterization of postpartum depression using genomic, methylomic, transcriptomic, and endocrine data.

### Authors:

J. Guintivano <sup>1</sup>; R. Chan <sup>2</sup>; D.R. Rubinow <sup>1</sup>; P.F. Sullivan <sup>1,3,4</sup>; K.A. Aberg <sup>2</sup>; S. Meltzer-Brody <sup>1</sup>; E.J.C.G van den Oord <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 2) Center for Biomarker Research and Precision Medicine, Virginia Commonwealth University, Richmond, VA.; 3) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

---

**INTRODUCTION:** Postpartum depression (PPD) is a severe perinatal form of major depressive disorder. It is one of the most frequent complications of childbirth (prevalence 10-15%) and leading cause of maternal death due to suicide. PPD is an under-studied health concern and its biological underpinnings are not well understood.

**METHOD:** We recruited 1,517 ancestrally diverse women (549 cases, 968 controls) at six weeks postpartum. PPD status was determined using diagnostic interview. Levels of estradiol (E2), progesterone (P4), BDNF, and oxytocin (Oxt) were assayed in serum/plasma. DNA and RNA from whole blood was used to generate data with three omic platforms. Genotyping was done on the Illumina Multi-Ethnic Genotyping Array which, after imputation and quality control (QC), left 12.5 million SNPs. DNA methylation was measured using the Illumina 450k array. A new QC pipeline was developed to prevent test static inflation across 374K CpGs. RNA-seq reads were annotated using ENSEMBL Release 92 resulting in 108,474 transcripts for association testing. Cell proportions were derived using methylation references and complete blood counts. Cell type proportion allowed for association tests within constituent populations of T-cells/B-cells/monocytes/granulocytes. Associations were determined for the outcomes case status and its interaction with assayed hormones. Analyses were performed in RAMWAS.

**RESULTS:** Associations with PPD were seen across platforms ( $P < 5e-08$ ), notably within CD8T, CD4T, and B cell populations and interaction analyses of PPD and P4. Top genotype signals showed significant overlap with both the top methylation and transcription findings (P range: 0.02-6.00E-04), and that the top methylation findings showed significant overlap with the gene expression findings (P range: 0.01-6.86E-03). Pathway analyses suggest processes linked to dysregulation of gene expression. Lastly, we used a machine learning algorithm with 10-fold cross validation to condense all association information for bulk and the cell types into a single risk score for each outcome tested. When combined, these risk scores correlate 0.3 with case status with the main predictors being methylation and gene expression in bulk.

**DISCUSSION:** Convergent evidence across three omic platforms suggests differential responses to hormones, namely P4, characterizes PPD. Biomarker risk scores suggest methylation and expression data can be used along with clinical data to better anticipate PPD onset.

## PgmNr 2171: GWAS for handedness indexes: Support a role of microtubule in brain asymmetries and neurodevelopmental disorders.

### Authors:

F. Abbondanza; J. Schmitz; C. Buenaventura; S. Paracchini

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Medicine, University of St Andrews, St Andrews, Fife, United Kingdom

---

**Background:** Worldwide, ~10% of individuals are left-handed with twin studies showing a 25% heritability. Increase of non-right handedness frequency has been reported in neurodevelopmental conditions including dyslexia and schizophrenia. These findings highlight the potential to use behavioural lateralised traits to investigate neurodevelopmental disorders. To date, very few genes have been associated with hand skill (*PCSK6*) or hand preference (*MAP2*, *MAPT*, *TUBB*), with the latter genes being linked to brain disorders and microtubule-related processes.

**Methods:** Here we performed and compare GWAS analyses of four quantitative handedness measures and we perform the first GWAS of foot and eye preference. All analyses were performed in SNPTEST while functional mapping on FUMA. Genetic correlations and SNP heritability were computed with LDSC and PRS for neurodevelopmental disorders were performed with PRSice.

**Results:** With a sample size of ~ 8,000 children, we did not find genome-wide significant hits. However, SNPs in the proximity to genes being associated with neurodevelopmental disorders (Schizophrenia, Usher, Autism), microtubule functions and axonal growth were highlighted. Pathway-analyses shows enrichment of inflammatory pathways, while locomotory behaviour and neuronal-related pathways have been highlighted.

**Discussion:** Handedness is highly influenced by social factors which, together with the heterogeneity of phenotypes in the literature, hinder genetic analysis. The behavioural correlation across the handedness traits is low, suggesting that the tests explore different cognitive skills, and this is reflected by different genetic profiles. The markers highlighted in the current study strengthen the current understanding that behavioural laterality, brain asymmetries and neurodevelopmental disorders are linked. Top hits close to genes involved in tubulin and cytoskeleton are present, in line with the literature. Despite the modest sample size, these findings highlight the importance of the phenotype definition, including for a trait like handedness intuitively straightforward



## PgmNr 2172: Gene ontology derived candidate gene analysis for reading disability.

### Authors:

A.K. Adams <sup>1</sup>; J. Frijters <sup>2</sup>; D.T. Truong <sup>3</sup>; M.M.C. DeMille <sup>3</sup>; J.M. Bosson-Heenan <sup>3</sup>; J.R. Gruen <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Yale University School of Medicine, New Haven, Connecticut.; 2) Department of Child and Youth Studies, Brock University, St Catharines, Ontario, Canada.; 3) Department of Pediatrics, Yale University School of Medicine, New Haven, Connecticut.; 4) Investigative Medicine Program, Yale University School of Medicine, New Haven, Connecticut, USA

---

While prior studies have demonstrated high heritability for reading disability (RD), commonly called dyslexia, small sample sizes have hindered identification of associated genes and variants. To balance limited sample sizes with our goal to discover novel genes, candidate genes for RD were selected based on connection to gene ontology (GO) terms. The primary sample was drawn from the Genes, Reading, and Dyslexia study (GRaD), a cohort of 1,331 Hispanic American and African American children 9 – 14 years of age. GRaD combines a comprehensive battery of reading and language measures, as well as genotyping with over 2,300,000 SNPs on the Illumina HumanOmni 2.5 array. We chose a highly-vetted endophenotype for reading and language disorders called Rapid Automatized Naming of objects (RAN objects), which taps into underlying cognitive processing speed and recently produced genome-wide significant results in GRaD (Truong et al. 2019). Subjects were separated into ancestry groups according to clustering with 1000 Genomes Project AMR superpopulation and ASW population. For the primary analysis in 865 GRaD Hispanic subjects, SNP variants were assigned to genes identified with GO terms, and gene-wise association determined with MAGMA applied to RAN-objects performance, controlling for sex, a SES composite measure derived from participation in social assistance programs, 2 ancestry derived PCs, subject age at testing, and highest year of education for the primary caretaker. *RUNX1* showed a suggestive p-value of  $4.9 \times 10^{-5}$ , though this p-value did not meet the Bonferroni adjusted threshold of  $2 \times 10^{-5}$ . *RUNX1* is a transcription factor associated with early neural development. This gene has previously not been associated with RD, supporting the value of this approach in connecting new genes to RD. While no single gene achieved statistical significance when correcting for multiple testing, the top 10 genes from the primary GRaD analysis were selected for replication in the Hispanic subset of the New Haven Lexinome Project, a longitudinal cohort of 360 (to date) students in the New Haven Public School District. Replication results and analysis of the African American subset of both datasets are forthcoming.



## PgmNr 2173: Contribution of polygenic risk score for psychiatric disorders in consanguineous Pakistani pedigrees.

### Authors:

J. John <sup>1,2,4</sup>; A. Johnson <sup>2</sup>; Q. He <sup>2</sup>; M. Christian <sup>2</sup>; L. Delisi <sup>3</sup>; R. Joobar <sup>1</sup>; G.A. Rouleau <sup>4</sup>; L. Xiong <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry and Douglas Mental Health University Institute, McGill University, Verdun, QC H4H 1R3, Canada; 2) Centre de Recherche, Institut Universitaire en Santé Mentale de Montréal et Université de Montréal, Montréal, QC, H1N 3M5, Canada; 3) VA Boston Healthcare System, Department of Psychiatry, Harvard Medical School, Brockton, MA, 02301, USA; 4) Department of Neurology and Neurosurgery and Montreal Neurological Institute and Hospital, McGill University, Montreal, QC, H3A 2B4, Canada

---

**Introduction:** Large-scale genome-wide association studies (GWAS) in Caucasian case-control cohorts have shown that schizophrenia (SCZ) and bipolar disorder (BPD) are highly polygenic in nature and numerous loci with small effects may contribute to these mental disorders. Polygenic risk scores (PRS), which aggregate the effects of variants across the genome at various thresholds, have been used to estimate heritability, to infer genetic overlap between traits and to predict phenotypes based on the genetic profile in additional samples. In this study, we aimed to estimate the contribution of PRSs from the GWASs of Caucasian samples in large consanguineous Pakistani pedigrees with SCZ and BPD.

**Samples and Methods:** 124 cases (SCZ or BPD) and 151 controls from ten large consanguineous pedigrees plus 34 unrelated healthy individuals, recruited from Sindh Province of Pakistan, were genotyped by Illumina Human OmniExpress Bead Chip. The summary statistics from the latest Psychiatric Genomics Consortium SCZ and BPD GWAS studies were used, and markers were pruned in each dataset with p-value-informed LD clumping (using LD pruning parameters of  $r^2=0.1$  over 500 kb windows) and collapsed to the most significant marker. After standard QC and imputation on the Pakistani family genotype data, PRS was calculated using PRSice-2 software, including all genetic variants with minor allele frequency  $\geq 5\%$ .

**Results:** Preliminary results showed that PRS generated from the PGC SCZ and BPD GWAS datasets were able to explain a maximum variance in our Pakistani family dataset at the p-value significance threshold of  $5e-08$  and  $0.1$ , respectively. Estimating the respective heritability and exploring PRS from GWAS summary statistics of other related psychiatric phenotypes in our family dataset are underway.

## PgmNr 2174: The genetics of addiction in a large Northern Nevada cohort of the Healthy Nevada Project.

### Authors:

K.A. Schlauch<sup>1,2</sup>; R.W. Read<sup>1,2</sup>; E.T. Cirulli<sup>3</sup>; S. White<sup>3</sup>; N.L. Washington<sup>3</sup>; J. Lu<sup>3</sup>; A. Slonim<sup>1,2</sup>; J.J. Grzymalski<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Applied Innovation Center, Desert Research Institute, Reno, Nevada.; 2) Institute for Health Innovation, Renown Health, Reno, NV; 3) Helix, San Carlos, CA, USA

---

Addiction, particularly to opioids, has reached epidemic levels in the U.S. By merging the genetics of a large population with comprehensive Electronic Health Records (EHR), socio-economic, geographic, and environmental data, the Healthy Nevada Project (HNP) aims to identify disease associations specific to local communities. Here we use Helix Exome+ sequencing of 20,000 HNP participants to examine genetic and phenotypic associations of substance addiction. The current control cohort contains 650 participants carefully selected as users of opiates, other illicit drugs, or alcohol, but who do not exhibit dependence or addiction. The case cohort consists of 1,000 participants with at least one addiction diagnosis in the last 12 years. A genome-wide association study (GWAS) was performed across cohorts using 2.5 million high-quality variant calls (GQ>20) obtained after applying additional thresholds (90% SNP call rates, 85% individual call rates, Hardy-Weinberg  $p > 1 \times 10^{-6}$ ). PLINK v. 1.9 performed logistic associations with age, gender, and principal components to control for population stratification as covariates, and the standard log-additive genetic model. Two phenome-wide analyses (PheWAS) were performed: the first PheWAS examined associations between significant SNPs identified in the GWAS and EHR phenotypes based on ICD codes; the second identified associations between incidence of addiction and ICD-based diagnoses. Of the 10 strongest associated SNPs ( $p < 1 \times 10^{-5}$ ), four are in genes reported in prior genetic studies of addiction. Two genes, *ITPR2* and *GABBR2*, have functional relevance in opioid, cocaine, alcohol and nicotine addiction pathways. A third association in the HNP cohort lies in *ITGA9*, a gene down-regulated in long-term oxycodone use in mice and linked to human drug abuse. A fourth variant identified is in *PARK2*, a gene often linked to Parkinson's Disease, supporting the notion that the molecular mechanisms of neurodegeneration of the two disorders are similar. The strongest association in our cohort is a variant in the geneless region 20q11.2 not yet affiliated to addiction. PheWAS results show strong clinical links with addiction: tobacco use disorder ( $p < 1 \times 10^{-55}$ ), mood disorders, depression, and anxiety ( $p < 1 \times 10^{-10}$ ). SNPs associated with addiction in the HNP are linked to alcoholism, tobacco use disorder, emphysema, and lupus. This study shows prior identified and possible novel genomic associations with substance addiction in a large HNP cohort.

## PgmNr 2175: Applying reverse linkage as a method in 1,589 Finnish migraine families.

### Authors:

K. Veerapen<sup>1,2,3</sup>; M.I. Kurki<sup>1,2,3</sup>; A. Mitchell<sup>5</sup>; P. Gormley<sup>11</sup>; M.E. Hiekkala<sup>4</sup>; H. Runz<sup>6</sup>; P. Häppölä<sup>7</sup>; E. Hämäläinen<sup>7</sup>; M.A. Kaunisto<sup>7</sup>; FinnGen

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital / Broad Institute, Boston, Massachusetts.; 2) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 3) The Stanley Centre for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, MA, USA; 4) Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland; 5) Eisai Center for Genetics Guided Dementia Discovery, Eisai Inc., Cambridge, MA; 6) Department of Genetics & Pharmacogenomics, Merck Research Laboratories, Merck and Co., Inc., Boston, MA; 7) Institute for Molecular Medicine Finland FIMM, HiLIFE, University of Helsinki, Helsinki, Finland; 8) Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland; 9) Program in Medical and Population Genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 10) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 11) Genomics and Pharmacogenomics, Merck Research Labs, Boston, MA

---

Migraine affects up to 15% of the population and is shown to have a heritability of up to 57% (PMID:14624726). In addition to the 38 common loci published in a genome-wide association meta-analysis (PMID:27322543), rare variants may provide a deeper understanding of underlying etiology of migraine. Previous linkage studies on rare hemiplegic migraines have identified variants in *CACNA1A*, *SCN1A*, and *ATP1A2*: however, the role of these genes in common forms of migraine has not been established. Rare variants are more enriched in isolate populations e.g. Finland. Hence, we hypothesize increased chances in identifying rare variants by studying large Finnish migraine families. In traditional linkage analysis, all families are analyzed together by summing LOD (logarithm-of-odds) scores across families in a given locus. A signal from a highly penetrant rare variant which appears in a subset of families may be missed, resulting in a loss of power to detect linkage. We propose to define a quantitative method by integrating linkage analysis and whole genome sequencing (WGS) data -- the reverse-linkage method. First, we joint called WGS data from 631 migraineurs: 184 index patients and 447 family members. We applied a strict QC criteria and annotated the filtered variants. Variant and family selection were based on the index patients with the assumption that variants identified would segregate in respective families. We filtered for variants which were shared in  $\geq 2$  families, rare, and damaging variants: missense ( $\text{gnomAD}_{\text{FIN}} < 0.05$ ,  $\text{CADD} > 10$ ) and loss-of-function (LOF) variants. Our QC and filtering resulted in 7836429 variants: 12966 missense and 745 LOF variants. Upon comparing damaging:synonymous variants ratio, results indicate a low damaging load (1:1.2). Subsequently, we ran parametric linkage analyses on markers near the variants identified in the WGS using the migraine collection of 8,319 Finns in 1,589 families. The resulting partial LOD scores for families carrying the proxy marker were then used for computing sum-of-LODs ( $\text{LOD}_v$ ). With a filter of  $\text{LOD}_v > 2.0$ , we identified 6 variants. These variants were investigated for possible phenotype etiological association in 96,499 Finns (FINNGEN project) against 1,122 phenotypes constructed from registry data. The results showed that the 6 variants identified were of low effect sizes ( $p > 10^{-4}$ )

towards constructed phenotypes. Our results suggest that highly penetrant variants are not major genetic contributors to migraine.

## PgmNr 2176: Case-only exome analysis of severe alcohol dependence using a multivariate hierarchical gene clustering approach.

### Authors:

B.T. Webb <sup>1,2</sup>; J. Alexander <sup>1,2</sup>; S.A. Bacanu <sup>1,2</sup>; K.S. Kendler <sup>1,2</sup>; B.P. Riley <sup>1,2,3</sup>; A.E. Gentry <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Virginia Institute of Psychiatric and Behavioral Genetics; 2) Department of Psychiatry; 3) Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA, USA

Background: Variation in genes involved in ethanol metabolism has been shown to influence risk for alcohol dependence (AD) including the inactivating rs671 in *ALDH2*; rs671 heterozygotes show only 20-40% of wild type enzymatic activity due to the homotetrameric structure of mature ALDH2. When ethanol is consumed, the reduced enzymatic activity leads to the accumulation of acetaldehyde, the first product of ethanol catabolism, and unpleasant symptoms including flushing. This variant is common in east Asian populations and associated with lower rates of alcohol abuse/dependence. Given the existence of protective loss of function alleles in ethanol metabolizing genes, we hypothesized that people with severe alcoholism would have less rare functional variation than expected.

Objective: Test this hypothesis using a case only design and Whole Exome Sequencing (WES) data from a collection of 195 severe alcoholic cases from Ireland.

Methods: A novel case only analysis method to compare sets of hypothesized candidate genes to matched sets of controls genes was developed and applied. Briefly, multivariate hierarchical clustering was used to match genes of interest (GOI) with sets of control genes using gene-level summary features obtained from the gnomAD database. After identifying sets of control genes for each GOI, multivariable logistic regression was used to test if GOI were different from control genes for the number of synonymous, missense, and LOF variants. This approach was applied to two separate gene sets, first using 15 *ADH* and *ALDH* alcohol-metabolizing genes and second using a published set of 357 genes (including the 15 from the previous set) shown to alter ethanol response in invertebrate models.

Results: Results from the analyses of the first set of 15 genes showed no significant difference in number of synonymous, missense, or LOF variants between GOI and control genes while analysis of the second, larger set of genes found slightly elevated counts of synonymous variants in alcohol-related genes of interest as compared to control genes. With this analysis, we have demonstrated a viable, computationally and statistically reasonable approach for genetic analyses of case-only data.

## PgmNr 2177: Transcriptomic signature of the 16p11.2 duplication in cultured neurons and brain.

### Authors:

N. Akula<sup>1</sup>; J. Xueying<sup>1</sup>; S. Liu<sup>2</sup>; L. Kassem<sup>1</sup>; S. Detera-Wadleigh<sup>1</sup>; B. England<sup>1</sup>; A. Raznahan<sup>2</sup>; F.J. McMahon<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Mood and Anxiety Section, Human Genetics Branch, National Institute of Mental Health, NIH, Bethesda, Maryland.; 2) Developmental Neurogenomics Unit, National Institute of Mental Health, NIH, Bethesda, Maryland.

---

The 650 kb duplication copy number variant (CNV) on chromosome 16p11.2 confers substantial risk for autism, schizophrenia, bipolar, and other neuropsychiatric disorders, but little is known about how resulting differences in gene dosage lead to pathogenic changes in the transcriptome, function, or structure of the developing brain. We have ascertained a proband with schizoaffective bipolar disorder who was found to carry the 16p11.2 duplication after SNP array analysis of DNA extracted from peripheral blood. Digital droplet PCR of fibroblasts obtained from the proband and 5 available relatives confirmed the duplication and revealed 2 additional carriers, along with 3 non-carriers. Fibroblasts were reprogrammed into induced pluripotent stem cells (iPSC) with a Lentivirus vector, then differentiated into neurons by use of published protocols. Neuronal identity was established through immunostaining and morphology. After 6 -7 wk of neuronal differentiation, high-quality RNA was depleted of ribosomal RNA (Ribo-Zero Gold) and used to prepare stranded libraries that were pair-end sequenced on the Illumina NovaSeq 6000. A total of 100 million reads per sample were mapped to 26,363 Ensemble genes. Differential gene expression between carriers and non-carriers was analyzed with DESeq2. As expected, many (14/27) genes within the duplicated region showed increased expression, with mean fold-change values of 1.5-2.5. Genome-wide, 152 genes were differentially expressed (DE) at FDR<5%. DE genes were enriched for pathways involved in neuronal development and migration, synaptic transmission, and neurite growth. Consistent with these results, 7-8 wk neuronal cultures from 16p11.2 carriers showed decreased dendritic arborization, increased neurite length, and decreased expression of the post-synaptic density protein, PSD95. DE genes were disproportionately expressed across 59 brain regions mapped by the Allen Brain Atlas. Comparison with published reports of brain morphologic differences in people with 16p11.2 CNVs (Martin-Brevet et al. 2018) revealed striking overlaps between anatomical and gene expression differences, especially in the right insula. This study demonstrates for the first time that patient-derived neurons carrying the 16p11.2 duplication display concordant transcriptomic and cellular phenotypes that overlap with dosage-sensitive anatomical differences in brain regions thought to underlie attention, emotion, and reward.

## PgmNr 2178: Evaluating the phenotype of the 15q13.3 deletion using clinical MRI data.

### Authors:

A. Hare-Harris <sup>1</sup>; M. Ulicny <sup>1</sup>; A. Moreno-De-Luca <sup>2</sup>; G. Mongelluzzo <sup>2</sup>; C.L. Martin <sup>2</sup>; V. Troiani <sup>2</sup>;  
Regeneron Genetics Center

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biological and Allied Health Sciences, Bloomsburg University, Bloomsburg, PA; 2) Geisinger, Lewisburg, PA

---

The 15q13.3 deletion (breakpoints 4-5; 15qdel) occurs in approximately 1 of 40,000 individuals worldwide and encompasses six known genes, including *CHRNA7*. The 15qdel exhibits variable expressivity with clinical phenotypes that range in severity. The most common clinical diagnoses associated with 15qdels include intellectual disability, epilepsy, schizophrenia, and other neuropsychiatric disorders. Previous studies of this deletion have focused on pediatric populations with severe phenotypes, while adults and individuals with milder 15qdel phenotypes have not been well characterized. In this study, we identified 55 individuals with a 15qdel (mean age = 44.96 years) from 92,455 patient-participants in the Geisinger-Regeneron DiscovEHR project, which includes exome and electronic health record (EHR) data from Geisinger. Of the 55 individuals with the deletion, 17 had at least one clinical brain MRI in their EHR as a result of epilepsy, migraine, and/or meningioma diagnoses. Therefore, the aim of this study is to evaluate clinical MRI reports to characterize the phenotype of the 15qdel as reflected in brain imaging findings. Radiologist reports that were entered into the EHR at the time of the scan were exported into a database and findings from each report were summarized. Overall, there were 2 with reported meningioma, 10 individuals with reported scattered white matter hyperintensities (WMH), 7 with reported sinus abnormalities, and 2 with brain volume loss. The incidence of meningioma in these patients is higher than expected, with an annual incidence of approximately 2 cases per 100,000 individuals. Importantly, some of these findings may reflect variation in clinical reporting terminology, particularly those that are more common, such as in the sinuses and WMH. In ongoing analyses, we will assess whether radiological findings are different using a rubric-based consensus review by neuroradiologists, as well as comparing volumetric brain measurements with a matched control group. This is the first study of its kind to assess the phenotype of the 15qdel using radiologist report summaries and provides insight into the variable neurological manifestations of 15qdel.

## PgmNr 2179: Studying the role of short tandem repeat variants in schizophrenia risk.

### Authors:

S. Saini <sup>1</sup>; M. Gymrek <sup>1,2</sup>; PGC Schizophrenia Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computer Science and Engineering, Univ California San Diego, La Jolla, California.; 2) Department of Medicine, Univ California San Diego, La Jolla, California.

---

Genome-wide association studies (GWAS) have successfully identified thousands of genetic loci associated with a range of complex traits. However, GWAS based on common single nucleotide polymorphisms (SNPs) still fail to explain the majority of heritability for most traits, even with sample sizes of up to several hundred thousand individuals. One compelling hypothesis explaining this is that complex variants, such as multi-allelic repeats not in strong linkage disequilibrium (LD) with common SNPs are important drivers of complex traits. We focus on the role of Short Tandem Repeats (STRs), a highly polymorphic class of repeats consisting of 1-6bp repeating motifs present at more than 1.5 million loci in the human genome. STRs have been implicated in more than 30 Mendelian disorders and a variety of complex traits. Intriguingly, most repeat disorders identified to date have involved neurological or psychiatric phenotypes.

Here, we present the first STR-based GWAS of schizophrenia, a highly polygenic trait which has been robustly associated with more than 100 genomic loci. We first used our published phased reference haplotype panel to impute STRs into more than 60,000 samples from Psychiatric Genomics Consortium (PGC) schizophrenia dataset. We used imputed genotypes to test more than 450,000 STRs for association with schizophrenia. We next performed simulation analyses on case-control cohorts to obtain an empirical estimate of the genome-wide significance threshold for STR association testing. Our GWAS identified 1740 STRs that passed the genome-wide p-value threshold estimate of  $1.11 \times 10^{-5}$ . We applied LD-based clumping to identify independent association signals using SNP and STR summary statistics and used CAVIAR to perform fine-mapping of each signal.

Preliminary results suggest 11 of the 108 previously associated signals are potentially driven by underlying STRs. Candidate loci will be validated using orthogonal cohorts. Overall, we anticipate that incorporation of STRs and other complex variant types into GWAS will identify additional causal variants and ultimately lead to improved understanding of the genetic architecture and biological processes driving complex traits.



## PgmNr 2180: Genome-wide study of speech-sound disorder identifies ties to regulatory elements controlling neurological development and function.

### Authors:

P. Benchek <sup>1</sup>; R.P. Igo <sup>1</sup>; H. Voss-Hoynes <sup>1</sup>; Y. Wren <sup>2</sup>; B. Truitt <sup>1</sup>; L. Freebairn <sup>1</sup>; J. Tag <sup>1</sup>; E.R. Chan <sup>1</sup>; B. Lewis <sup>1</sup>; C. Stein <sup>1</sup>; S. Iyengar <sup>1</sup>; P. Roussos <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH. 44106; 2) University of Bristol, Beacon House, Queens Road, Bristol, BS8 1QU, UK.; 3) Departments of Genetics and Genome Sciences, and Psychiatry, Icahn School of Medicine at Mount Sinai, NYC, NY

---

The sound system of language forms the basis for most human communication, and disorders that affect speech have a lifelong adverse impact. The known neural basis of speech resides in Broca's and Wernicke's areas of the brain, but the presence of speech and language dysfunction in several neurodevelopmental disorders suggests a broader genetic foundation. Speech-pathologist diagnosed families (N=435 children from 148 families) in the Cleveland Family Speech study (CFS) were assessed for 16 variables representing 6 domains (oral motor function, phonological memory and awareness, reading decoding, spelling, expressive and receptive language and vocabulary). We examined the genetic basis of non-syndromic childhood speech sound disorder (SSD) via a genome-wide association study (GWAS) using RVtests for >8.5M SNPs (MAF>5%). For replication, we obtained data from the Avon Longitudinal Study of Parents and Children (ALSPAC), a birth cohort containing children with similar endophenotypes and available genetic data (max N=6203), in which mothers of 4% of children self-reported problems with speech. Although genetic correlations between measures are high (range: 0.56-0.99), there is considerable familial discordance in inheritance of identical phenotypes between sibs, suggesting that cross-trait replication across any pair of assessments between CFS and ALSPAC may provide more consistent conclusions. We thus identified 163 significantly associated SNPs in CFS ( $p < 1 \times 10^{-5}$ ), which replicated in correlated ALSPAC traits ( $p < 0.05$ ), residing in 32 novel loci, with 18 SNPs having combined  $p < 10^{-8}$ . We annotated SNPs using HaploReg, RegulomeDB, and FUMA-GWAS as they were mostly intergenic (74%) or intronic (23%). Many lead SNPs or LD-tags (80%) either mapped to brain eQTLs or were enriched for neuronal promoter/enhancer elements. For example, rs856380 (combined  $p = 2.1 \times 10^{-6}$ ) is an eQTL for DACT1, highly expressed in the substantia nigra, and an enhancer for the language processing temporal region of the brain. In CFS alone, we replicated previously associated genes ATP2C2 ( $p = 7.7 \times 10^{-8}$ ) and CNTNAP2 (CFS  $p = 5.2 \times 10^{-6}$ ). Using gene set enrichment analysis, regulation of gene expression and molecular function were top GO pathways overrepresented by our mapped genes ( $p = 1 \times 10^{-6}$  and  $1.7 \times 10^{-7}$ ). This evidence suggests that the genetic influence on SSD is primarily in the regulation of neural and brain pathways, and that these pathways may be common with other neurological disorders.

## PgmNr 2181: Genetics of severe mental illness in a Colombian population isolate.

### Authors:

L. Olde Loohuis <sup>1</sup>; S. Service <sup>1</sup>; J. de la Hoz <sup>1</sup>; M. Castano <sup>7</sup>; J. Song <sup>1</sup>; M. Lopez Tobon <sup>2</sup>; L. Agudelo <sup>2</sup>; J. Melo <sup>2</sup>; M. Munoz Umanes <sup>1</sup>; C. Vargas Upegui <sup>2</sup>; S. Ruiz Sanchez <sup>2</sup>; C. Montoya <sup>2</sup>; T. Teshiba <sup>1</sup>; T. Moore <sup>3</sup>; A. Port <sup>3</sup>; J. Castro Navarro <sup>8</sup>; S. Jensen <sup>1</sup>; C. Gallego <sup>7</sup>; J. Palacio <sup>2</sup>; R.C. Gur <sup>3</sup>; R.E. Gur <sup>3</sup>; C. Sabatti <sup>4</sup>; J. Escobar <sup>5</sup>; V. Reus <sup>6</sup>; C. Bearden <sup>1</sup>; C. Lopez-Jaramillo <sup>2</sup>; N. Freimer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of California Los Angeles, Los Angeles, USA.; 2) University of Antioquia, Medellin, Colombia; 3) University of Pennsylvania School of Medicine; 4) Stanford University, Stanford, USA; 5) Rutgers-Robert Wood Johnson Medical School, USA; 6) Univ California San Francisco, San Francisco, USA; 7) University of Caldas, Manizales, Colombia; 8) Clinica San Juan de Dios, Manizales, Colombia

---

Our current classification system divides most severe mental illness (SMI) into dichotomous diagnostic categories; psychotic disorders and mood disorders. However, these categories overlap clinically, and are genetically correlated. We are ascertaining and phenotyping a Colombian sample of 8,000 SMI individuals across diagnostic categories, and 2,000 demographically matched controls. In this sample, we will attempt to identify symptom-level and quantitative phenotypes that better represent the biological underpinning of SMI than do the diagnostic categories alone.

Recruitment of cases takes place primarily at a single psychiatric hospital serving the state of Caldas, a large catchment area within the “Paisa” region of Colombia, characterized by a high level of cultural and genetic homogeneity. Through high quality electronic health records (EHR), patients across the full range of severe mood and psychotic disorders are ascertained in a diagnosis-naïve way. Phenotyping is performed in a deep and uniform manner, and includes diagnostic interview (NetSCID), symptom assessment (SA45) and a neurocognitive battery (PennCNB).

Here, we will provide an overview of the first 3,000 cases, representing SMI across diagnoses in Colombia, and highlight some initial results from the “Paisa project”.

The diagnostic breakdown of the SMI sample is as follows: 38% bipolar disorder (BD) type I, 13% other BD, 33% major depressive disorder and 13% schizophrenia. This surprisingly high proportion of cases with BD matches that of hospitalized patients as a whole as well as that for other hospitals in the Paisa region.

Our genetic analyses of the Paisa population highlight strong characteristics of isolation, even relative to the Finnish population. Exploring the relationship between genome-wide significant SMI-associated variants and specific symptoms of SMI in our cohort, our data provides a replication of a locus previously associated with BD, which in our sample has a much larger effect size than those observed in published psychiatric GWAS.

Finally, using NLP-based machine learning approaches, we are leveraging EHR data on >70,000 cases collected over a period of 14 years from the hospital to extract phenotypic constructs such as diagnostic features and substance use phenotypes.

Our cohort provides unique opportunities for the (genetic) characterization of SMI that may ultimately lead to novel approaches for disease classification.

## PgmNr 2182: Association between polygenic liability for schizophrenia and substance involvement: A nationwide population-based study in Taiwan.

### Authors:

S.H. Wang <sup>1</sup>; R.Y. Lai <sup>1</sup>; Y.C. Lee <sup>2</sup>; M.H. Su <sup>2</sup>; C.Y. Chen <sup>3</sup>; A.C. Yang <sup>4</sup>; Y.L. Liu <sup>5</sup>; S.J. Tsai <sup>4</sup>; P.H. Kuo <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) China Medical University, Taichung, Taiwan; 2) National Taiwan University, Taipei, Taiwan; 3) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Taipei Veterans General Hospital, Taipei, Taiwan; 5) National Health Research Institutes, Miaoli County, Taiwan

---

**Background:** Schizophrenia and substance involvement, from use to severe dependence, frequently co-occur in the same individual. A growing number of studies have investigated shared genetic liability between schizophrenia and a range of substance involvement, including alcohol drinking, tobacco smoking, and substance use disorder, but these studies provided conflicting evidence. This study investigated the genetic overlap between schizophrenia and substance involvement, including phenotypes in relation to tobacco, alcohol, and betel nut use. **Methods:** The study subjects were recruited from the Taiwan Biobank, and genome-wide genotyping data was available in 18327 individuals without schizophrenia. We calculated the Psychiatric Genomics Consortium-derived polygenic risk score (PRS) for schizophrenia, the cumulative additive effect of disease-associated variants across the genome, in each individual, and the PRS was normalized to a Z score. The significance of the schizophrenia PRS associated with a wide range of substance involvement was evaluated by regression models with adjustment for sex, age, and population stratification components. The modified effect of sex or birth decade on the association between schizophrenia PRS and substance involvement was also explored. **Results:** The schizophrenia PRS was positively associated with lifetime tobacco smoking in women (OR in per SD increase in PRS=1.12 with 95% CI 1.04-1.20, p=0.002), but not in men (OR=0.99 with 95% CI 0.95-1.04, p=0.74), and the sex-PRS interaction reached significance (p=0.006). The strength of association, OR, between PRS and lifetime tobacco smoking increased with birth decade (p of birth decade-PRS interaction=0.0002). In women, OR increased from 0.97 (p=0.85) for birth decade before 1950 to 1.21 (p=0.04) for birth decade after 1980; in men, the corresponding OR increased from 0.88 (p=0.04) to 1.13 (p=0.11). There was no association between schizophrenia PRS and alcohol/betel nut use phenotypes. **Conclusions:** This study provided evidence for the genetic overlap between schizophrenia and tobacco smoking in women, and such overlap was stronger in later birth decade. However, the shared genetics may not reflect real biological pleiotropy, further studies are required to explore the influence of potential confounders/mediators.

## PgmNr 2183: Revisiting schizophrenia from an evolutionary perspective: An association study between recent evolutionary markers and schizophrenia.

### Authors:

U. Bhattacharyya <sup>1</sup>; S. Khanna <sup>1</sup>; S.N. Deshpande <sup>2</sup>; B.K. Thelma <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of Delhi South Campus, India, New Delhi, Delhi, India; 2) Department of Psychiatry, PGIMER- Dr. RML Hospital, New Delhi, India

---

Recent investigations suggest that abnormalities that arise during central nervous system development and subsequently exacerbated during cerebral maturation in adolescence and early adulthood may underlie schizophrenia. SZ persisted in human populations despite high morbidity and reduced fertility and thus it may likely be a by-product of key evolutionary events. Human brain has evolved into much larger and highly complex organ since the divergence from Chimpanzee. These evidences suggest that genomic components, attributable to human-specific brain development may have been favoured by natural selection and also predispose to SZ. Comparative genomics has led to the discovery of genomic regions namely Human Accelerated Regions (HARs) that are highly conserved among non-human species but have experienced accelerated substitutions in the human genome, serving as a marker of human-specific evolution since our divergence from chimpanzees. 93.1% HARs are present in the non-coding region and act as enhancers and/or repressors regulating expression around them. Of note, previous studies have shown that SZ genes are closer to HARs than would be expected by chance, but these are limited to a subset of HARs present near the known SZ candidate genes. To reassess the association of HAR SNPs in a hypothesis free manner, we have systematically selected 43 SNPs from all the HARs identified till now, genotyped and tested their association with SZ, using a two-stage study design comprising discovery and replication cohorts (n=1012 cases; n=1049 controls). Significant association of two SNPs (rs764453; P-value=7.7e<sup>-06</sup>, OR=1.5; CI=1.1-1.8 & rs3801844; P-value=0.0002, OR=1.4, CI=1.1-1.5) was observed. As these SNPs are intergenic, gene mapping was done using - i) positional, ii) eQTL and iii) chromatin interactions approaches and 26 genes were mapped. Pathway analysis of these genes revealed significant (P-value=2.30e-02) overrepresentation of developmental processes. Of these, previous knock-out mouse studies of six most relevant novel genes namely *DLX5*, *DLX6*, *CHN2*, *HOXA1*, *SKAP2* and *SLC25A13* have shown abnormal brain development, behavioral anomalies and increased anxiety related response and were also reported to be differentially expressed in SZ cases vs controls in the largest meta-analysis of gene-expression microarray studies of neuropsychiatric disorders. Further investigations of these novel HAR influenced genes may provide additional insights for SZ etiology in humans.

## PgmNr 2184: Exome variation in ethanol metabolizing enzyme genes in Irish severe Alcohol Dependence (AD) cases.

### Authors:

M. Ahangari<sup>1,2</sup>; A. Gentry<sup>1</sup>; M. Kealhofer<sup>1</sup>; C. Sun<sup>1,3</sup>; B. Wormley<sup>1,3</sup>; S.A. Bacanu<sup>1,3</sup>; B.T. Webb<sup>1,3</sup>; K. Kendler<sup>1,3,4</sup>; B. Riley<sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; 2) Integrative Life Sciences Doctoral Program, Virginia Commonwealth University, Richmond, VA; 3) Department of Psychiatry, Virginia Commonwealth University, Richmond, VA; 4) Department of Human & Molecular Genetics, Virginia Commonwealth University, Richmond, VA

---

Common variants do not account for all the genetic risk for complex traits like AD. For some traits, part of this 'missing heritability' has been shown to be due to the influence of rare variation. Rare functional variants that reduce AD risk are well documented, but the question of whether rare variants that increase AD risk exist has remained largely unexplored. We sequenced the exomes of 195 Irish AD cases, and variants were selected for further study based on minor allele frequency (MAF) differences between 195 cases and 3757 UK10K Project controls and bioinformatic assessment of the putative functional impact of variants. We focused initially on variants with MAF <0.01 in 10 genes encoding enzymes with a known role in ethanol metabolism (*ADH1A-C*, *ADH4-7*, *CYP2E1*, *CAT* and *ALDH2*). We identified 23 variants with predicted functional impact in these 10 genes that were more common in case exome data than in UK10K controls. Selected variants were then directly genotyped in 906 Irish AD cases and 2000 Irish controls using SNaPshot assays, to increase overall sample size and to ensure that differences were not due to Ireland-UK population differences. An ethnically homogeneous sample from Ireland is well-suited to the study of rare exome variation because of the differences observed in rare variant frequencies and distribution across European populations. Genotyping of the first panel is complete and no significant differences in frequency between Irish cases and Irish controls were observed. Two additional panels of variants in these 10 genes (including all the prioritized loss of function variants), are being genotyped currently.

## **PgmNr 2185: De novo missense mutations revealed by whole-exome sequencing in monozygotic twins are associated with schizophrenia.**

### **Authors:**

L. Guan <sup>1,2</sup>; N. Lv <sup>1</sup>; B. Wu <sup>1</sup>; H. Ma <sup>1</sup>; X. Yu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Peking University Institute of Mental Health, Beijing, China; 2) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA

---

### **Objectives:**

The genetic risk of schizophrenia is conferred by a large number of alleles, including common alleles of small effect and rare alleles of highly penetrance. Despite high heritability, a large fraction of cases with schizophrenia do not have a family history of the disease. Recent large-scale genomic studies have revealed a polygenic burden of rare disruptive mutations in schizophrenia and de novo mutations may account for a significant fraction of sporadic cases. Availability of next-generation exome sequencing permits the study of de novo mutations, especially in the coding regions. Monozygotic (MZ) twins, as natural control of each other, have been widely studied to dissect the relative contributions of genetics and environment. This study aim to identify novel variants and genes in sporadic cases of schizophrenia.

### **Methods:**

We collected three typical MZ twins families which has no history of psychiatric disorders except for the twins themselves. Standardized questionnaires and tools were administered to make schizophrenia diagnosis according to DSM-IV criteria and record the demographic details, developmental history as well as clinical and cognitive assessment. We sequenced the whole exomes of 14 subjects on Illumina HiSeq2500 platform. 13 SNVs and 8 indels in three families were selected to undergo the validation through Sanger sequencing.

### **Results:**

We identified two de novo missense mutations in two families. One mutation was found in the affected proband in a discordant twin pair. The other mutation originating from the germline was found to exist in both of the concordant twin pair. The two mutations both induce a disruptive amino acid change and are evaluated to be damaging and evolutionary consensual using SIFT, PolyPhen-2 and GERP++. Additionally, the two loci were absent in our in-house sample of 2,516 healthy individuals. The gene indentified with the germline mutation is highly expressed from the early fetal stages in human brain, including the hippocampus, substantia nigra, brain cortex, white matter, and ventral thalamus, which are regions implicated in the pathogenesis of schizophrenia.

### **Conclusions:**

Our study implicates novel genes in the etiology of schizophrenia and supports the notion that de novo mutations may account for some of the heritability reported for schizophrenia. Follow-up studies are needed to investigate the function of mutant protein.

## PgmNr 2186: Family based whole exome sequencing identifies novel genetic susceptibility loci in Han Chinese schizophrenia patients.

### Authors:

M. Li <sup>1</sup>; S. Qin <sup>1,2</sup>; L. He <sup>1,3</sup>; C. Wan <sup>1</sup>; L. Shen <sup>1</sup>; L. Chen <sup>1</sup>; H. Huang <sup>4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bio-x Institutes, Shanghai Jiao Tong Univeristy, Shanghai, Shanghai, China; 2) Collaborative Innovation Center, Jining Medical University, Jining, Shandong, China; 3) The Third Affiliated Hospital, Guangzhou Medical University, Guangzhou, Guangdong, China; 4) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 5) Broad Institute of Harvard and MIT, Cambridge, MA, USA

---

Schizophrenia (SCZ) is a highly heritable psychiatric disorder that affects approximately 1% of population around the world, causing life-long treatment and substantial social burden. However, our knowledge of the genetic mechanisms of SCZ still remains incomplete and early relevant studies did not reach clear conclusions, suggesting that additional susceptibility loci that exert significant influence on SCZ are yet to be revealed. Family-based designs are particularly well suited for analysis of rare variants and transmission information from parents to offspring in a relatively small sample size. So, in order to identify novel susceptibility genes that account for the genetic risk of SCZ, we performed a systematic family-based study using whole exome sequencing (WES) in 65 Han Chinese families. The analysis of 51 SCZ trios with both unaffected parents identified 22 exonic and 1 splice-site *de novo* mutations (DNMs) on a total of 23 genes, and showed that 12 genes carried rare protein-altering compound heterozygous mutations in more than one trio. In addition, we identified 26 exonic or splice-site single nucleotide polymorphisms (SNPs) on 18 genes with nominal significance ( $P < 5 \times 10^{-4}$ ) using a transmission disequilibrium test (TDT) in all the families. Moreover, TDT result confirmed a susceptibility locus on 3p21.1, encompassing the multigenetic region *NEK4-ITIH1-ITIH3-ITIH4*. Through comprehensive evaluation, we revealed 4 previous discovered susceptibility genes (*TSNARE1*, *PBRM1*, *STAB1* and *OLIG2*) and 3 novel susceptibility loci (*TLR5*, *MGAT5B* and *SSPO*) in Han Chinese SCZ patients. In summary, we identified a list of putative candidate genes for SCZ using a family-based WES approach, thus improving our understanding of the pathology of SCZ and providing critical clues to future functional validation.



## PgmNr 2187: Cell-type specific expression quantitative trait loci during human neocortical differentiation.

### Authors:

N. Aygun<sup>1,2</sup>; M.J. Lafferty<sup>1,2</sup>; D. Liang<sup>1,2</sup>; A.L. Elwell<sup>1,2</sup>; K. Cheek<sup>1,2</sup>; K. Courtney<sup>1,2</sup>; L. de la Torre-Ubieta<sup>4</sup>; D.H. Geschwind<sup>3,4</sup>; J.L. Stein<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Human Genetics, David Geffen School of Medicine University of California, Los Angeles, Los Angeles, CA; 4) Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, CA

Genome wide association studies (GWAS) indicate that common non-coding variants have major impact on many human phenotypes, including neurodevelopmental disorders. One potential function of these variants is to regulate the cis expression genes. We performed an expression quantitative trait locus (cis-eQTL analysis) to map the effect of genetic variation on gene expression. As eQTLs can vary by cell type, we performed RNA-seq (rRNA depleted library preparation) on a homogeneous population of primary human neural progenitor cells (n = 89 donors) and their 8-week differentiated, virally labeled, and sorted neuronal progeny (n=75 donors). Each donor cell line was also genotyped on a dense array (Illumina Omni 2.5+Exome) and imputed to a common reference panel (1000 Genomes).

We performed proximal-eQTL analysis by evaluating association between each gene's expression and variants within  $\pm 1$  Mb window of transcription start site of each gene by implementing linear mixed model associations. We analyzed ~5.56 million variants and 15,848 protein coding, 2,219 lncRNA and 2,600 antisense RNA genes. From the eQTL analysis, 8,154 eSNPs displayed significant association with 3,666 eGenes in progenitors, and 4296 eSNPs were significantly associated with 2,196 eGenes in neurons at an FDR of 5%. Although 1,531 eQTLs were shared between progenitor and neurons, we also detected cell-type specific eQTLs. As an example of a cell-type specific eQTL, a locus previously associated with intracranial volume was found to regulate the expression of the gene *CENPW* in progenitors (rs9401883;  $p = 6.03e-07$ ), but not in neurons. To identify cell-types and genes impacted by disease GWAS loci, we further assessed coincidence of significant eQTLs with loci showing significant associations for schizophrenia. 24 eQTLs in progenitors and 28 eQTLs in neurons were coincident with schizophrenia GWAS SNPs, where the index SNP was shared with the GWAS and eQTL. For instance, index GWAS variant rs73058052 ( $p = 2.74e-08$ ) was also identified as lead eQTL SNP ( $p = 7.8e-06$ ) for schizophrenia related gene *NOSIP* in only progenitor eQTLs. Importantly, both these two eQTL examples were not significant in adult brain eQTLs from GTEx.

Here, we demonstrate a platform to explore the cell-type specific genetic basis of gene regulation during a critical time period of brain development, cortical neurogenesis, which complements existing post-mortem brain eQTL approaches.

## PgmNr 2188: Sequencing of chromosome 1 and chromosome 13 identifies damaging variants linked to verbal trait disorders.

### Authors:

D.T. Truong <sup>1</sup>; A.K. Adams <sup>2</sup>; S.D. Smith <sup>3</sup>; L.D. Shriberg <sup>4</sup>; J.R. Gruen <sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Yale School of Medicine, New Haven, CT; 2) Genetics, Yale School of Medicine, New Haven, CT; 3) Pediatrics, University of Nebraska at Omaha, Omaha, NE; 4) Waisman Center, University of Wisconsin-Madison, Madison, WI; 5) Investigative Medicine, Yale University School of Medicine, New Haven, CT

---

**Purpose:** Verbal trait disorders (VTD) describe a wide range of conditions that impair ability to communicate, and include deficits in speech, language, reading, spelling, and writing. We previously identified a 30Mb region of chromosome 13q14-q21 (LOD=4.35) that segregates with poor nonword repetition (NWR)—a robust endophenotype for VTD—in a six-generation extended family of European ancestry with a history of VTD (Truong et al., 2016). However, the chromosome 13 linkage signal does not fully explain all observed cases of VTD in this family. We hypothesize that there are additional genetic loci harboring damaging mutations segregating with other verbal trait domains in this family.

**Method:** We used a genome-wide multipoint variance component linkage analysis to identify chromosomal regions that segregate with self-reported problems in reading, spelling, speech, and/or language in school within the family. We then performed whole exome sequencing and targeted deep sequencing of candidate loci to identify damaging mutations segregating in 20 informative family members. ANNOVAR was used to predict functional variants that were likely pathogenic.

**Results:** The linkage analysis identifies a region on chromosome 1p22.3-22.1 with max LOD=3.98 between 122.5 and 123 cM, spanning approximately 900kb that segregates with a history of spelling problems in school. Whole exome sequencing of this region identifies two variants in predicted promoter regions—one for *TMED5*, the other, *FNBP1L*. Whole exome sequencing and targeted deep sequencing of the initial chromosome 13 linkage signal also identifies 17 predicted functional and deleterious variants in six genes that segregate with poor NWR. All eight genes associated with the chromosome 1 and 13 linkage and sequencing analyses are expressed in developing human brain, with 10 of the variants located in regions of predicted gene regulation. Eleven of the 19 variants are in *DIAPH3*, a gene associated with auditory processing. We also show that variants implicated in the regulation of *DIAPH3* are also associated with poor NWR performance in an independent general population cohort.

**Conclusions:** The often-overlooked non-coding genome may play a critical role in the etiology of VTD. Findings from this study highlight the importance of the non-coding genome in the regulation of genes relevant to the development of the brain, and verbal trait performance.

## PgmNr 2189: Genome-wide association study of opioid cessation.

### Authors:

J.W. Cox <sup>1</sup>; R.M. Sherva <sup>1</sup>; K.L. Lunetta <sup>2</sup>; E. Johnson <sup>3</sup>; N. Martin <sup>4</sup>; L. Degenhardt <sup>5</sup>; A. Agrawal <sup>3</sup>; E. Nelson <sup>3</sup>; H. Kranzler <sup>6</sup>; J. Gelernter <sup>7</sup>; L.A. Farrer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Section of Biomedical Genetics, Boston University, Boston, MA; 2) Dept. of Biostatistics, Boston University, Boston, MA; 3) Dept. of Psychiatry, Washington University in St. Louis, St. Louis, MO; 4) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Australia; 5) National Drug and Alcohol Research Centre, University of New South Wales, Sydney, Australia; 6) Dept. of Psychiatry, University of Pennsylvania, Philadelphia, Pennsylvania; 7) Dept. of Psychiatry, Yale School of Medicine, West Haven, CT

---

In the U.S, opioid use disorder (OUD) related deaths are rising. While studies have identified several loci that are associated with OUD risk, the genetic basis for the ability to discontinue opioid use has not been investigated. We performed a genome-wide association study (GWAS) of opioid cessation in the Yale-Penn study, including 1,130 African Americans (AAs) and 1,859 European Americans (EAs) recruited for genetic studies of opioid, cocaine, or alcohol use disorders and who met DSM-5 criteria for mild to severe OUD. Opioid cessation was defined as abstinence from opioids for at least one year before the interview date in the context of lifetime OUD. We examined the association of opioid cessation status genome-wide (GW) using a regression model that included terms for imputed single nucleotide polymorphism (SNP) dosage, sex, age and the first five principal components of ancestry that used generalized estimating equations to correct for correlations among related individuals. Association tests performed separately within each population were combined by GW meta-analysis with summarized results obtained from the Comorbidity and Trauma Study, which included a European-ancestry only Australian sample (N=1,060). Although there were no genome-wide significant associations, we found suggestive associations with three independent loci: rs1535677 in *PTPRD* ( $p=6.44 \times 10^{-8}$  in AAs and EAs combined), rs2865688 located 0.7kb downstream of *MINAR1* ( $p=5.96 \times 10^{-7}$  in EAs), and rs72330536 in *KCNQ5* ( $p=7.77 \times 10^{-7}$  in AAs and EAs combined). Pathway analysis identified significant functions related to glutamate signaling, cell-mediated immunity and cell cycle. We also found evidence ( $p<0.005$ ) of shared genetic underpinnings between opioid cessation and other substance use disorder such as cocaine, cessation of smoking and drinking, and chronic back pain in the UK Biobank by evaluating the association of polygenic risk scores constructed from summary statistics of these phenotypes with opioid cessation. These results provide evidence of genetic influences on opioid cessation and suggest genetic overlap with other relevant traits.

## PgmNr 2190: Whole genome sequencing determines contributions of rare and common genetic variants to risk for schizophrenia.

### Authors:

B. AdhamiMojarad <sup>1</sup>; Y. Yin <sup>1</sup>; R. Manshaei <sup>2</sup>; G. Costain <sup>3,4</sup>; T. Heung <sup>3</sup>; D. Merico <sup>5,6</sup>; C.R. Marshall <sup>7</sup>; A.S. Bassett <sup>3,8</sup>; R.K.C. Yuen <sup>1,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario; 2) Ted Rogers Centre for Heart Research, Cardiac Genome Clinic, The Hospital for Sick Children, Toronto, Ontario.; 3) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario; 4) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario.; 5) Deep Genomics Inc., Toronto, Ontario.; 6) The Centre for Applied Genomics (TCAG), The Hospital for Sick Children, Toronto, Ontario.; 7) Paediatric Laboratory Medicine, Genome Diagnostics, The Hospital for Sick Children, Toronto, Ontario.; 8) The Dalglish Family 22q Clinic and Toronto General Hospital Research Institute, University Health Network, Campbell Family Mental Health Research Institute, and Department of Psychiatry, University of Toronto, Toronto, Ontario; 9) Department of Molecular Genetics, University of Toronto, Toronto, Ontario.

**Objective:** Schizophrenia (SCZ) is a highly heritable mental disorder affecting ~1% of the population. Heterogeneity of SCZ challenges the identification of its underlying genetic risk. Rare (with the predominant role of copy number variants) and common genetic variants have both been shown to contribute to the etiology of SCZ. We aimed to determine the contribution of rare/common variation to SCZ and provide a more comprehensive estimation of genetic underpinnings in this disorder through the application of advanced genomic technologies to capture various genetic variants.

**Methods:** Here, we used whole genome sequencing (WGS) to study 259 unrelated adults with SCZ, a subset of a well-characterized community-based cohort previously assessed with chromosomal microarray, and by design enriched for those with rare CNVs. We assessed rare genetic variants for clinical relevance, including single nucleotide variants (SNVs), small insertions and deletions (indels), copy number variants (CNVs), and structural variants affecting genes or previously reported to be associated with schizophrenia and related disorders. In addition, we estimated SCZ liability from common genetic variants (single nucleotide polymorphisms, SNPs) using polygenic risk scores (PRSs). We also identified genes that were recurrently impacted by ultra-rare loss of function (LoF) variants.

**Results:** WGS detected all 28 pathogenic CNVs previously reported by microarray in this cohort, accounting for a likely genetic diagnosis in 10% (n=26) of the 259 cases. Rare clinically relevant SNVs and indels alone, likely contributing to the phenotype, constituted 3.5% (n=9) of individuals with no other identified relevant variant, e.g. CNVs. Common variants were responsible for 9.4% of the variance in SCZ liability, and significantly enriched in individuals with a family history of SCZ.

However, analysis of PRS quantile indicated these confer relatively small overall risk for pathogenicity even considered the top PRS quantile (odds ratio: 2.7, 95% confidence interval: 1.1-6.6]. Analysis of ultra-rare LoF variants suggested *ZMYM2*, a gene associated with transcription repression, as a novel SCZ-relevant gene candidate.

**Conclusions:** These results support the power of WGS to delineate the contributions of a variety of

genetic variants to SCZ liability. Our findings support the feasibility of using WGS as a potential single comprehensive genetic test for SCZ.

## PgmNr 2191: Inverse pleiotropic effects of the nicotine metabolizing gene *CYP2A6*.

### Authors:

M.J. Bray <sup>1</sup>; L.S. Chen <sup>1,2</sup>; L.J. Bierut <sup>1,2</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry, Washington University in St. Louis, St. Louis, Missouri.; 2) The Alvin J. Siteman Cancer Center, Washington University School of Medicine, St. Louis, MO.

**Introduction:** Smoking behaviors are contributors to lung cancer. Various genome-wide association studies (GWAS) have implicated shared and unique genetic loci with either smoking-related behaviors or lung cancer. For example, the chromosomal region 19q13.2 encompassing *CYP2A6*, which metabolizes nicotine, has been associated with cigarettes smoked per day, smoking cessation, and lung cancer. In this study, we examine the genetic correlations and interconnectedness between smoking-related phenotypes.

**Methods:** Genetic data consisted of GWAS summary results from previously published studies. SNP heritability and genetic correlations were estimated using linkage disequilibrium (LD) score regression software. Z-score-z-score (ZZ) plots were generated by comparing z-scores of each SNP between pairwise comparisons of each outcome. Chromosomal region 19q13.2 including *CYP2A6*, the primary nicotine metabolizing gene, is highlighted.

**Results:** The overall genetic correlations between increased cigarettes smoked per day, failed smoking cessation, and increased lung cancer risk are strong (cigarettes smoked per day – lung cancer:  $r = 0.59$ ; failed smoking cessation – lung cancer:  $r = 0.49$ ; cigarettes smoked per day – failed smoking cessation:  $r = 0.43$ ). The chromosomal region 19q13.2 shows a positive correlation between increased cigarettes smoked per day and lung cancer. The effect allele-C of the lead SNP (rs56113850) within *CYP2A6* is associated with an increased risk of lung cancer (z-score = 8.91;  $p = 5.02 \times 10^{-19}$ ) and increased cigarettes smoked per day (z-score = 19.2;  $p = 1.10 \times 10^{-81}$ ). Surprisingly, this variant is associated with successful smoking cessation (z-score = -8.17;  $p = 2.52 \times 10^{-26}$ ).

**Conclusion:** Genome-wide, there are strong, positive genetic correlations between all smoking-related phenotypes. Interestingly, the chromosomal region 19q13.2 encompassing the gene *CYP2A6* is not only associated with increased cigarettes smoked per day and elevated risk of lung cancer but also with successful smoking cessation. This inverse relationship highlights the need for additional analyses to determine how *CYP2A6*, which metabolizes nicotine, could increase smoking cessation yet increase risk of lung cancer likely through increased cigarettes smoked per day.

## PgmNr 2192: Association between a genetic variant in the *PDE9A* gene and the risk of idiopathic hypersomnia.

### Authors:

T. Miyagawa <sup>1,2</sup>; K. Tanida <sup>2</sup>; M. Shimada <sup>1,2</sup>; S.S. Khor <sup>2</sup>; K. Mishima <sup>3</sup>; M. Honda <sup>1,4</sup>; K. Tokunaga <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sleep Disorders Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Department of Psychophysiology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan; 4) Seiwa Hospital, Neuropsychiatric Research Institute, Tokyo, Japan

---

Idiopathic hypersomnia (IHS) is a rare sleep disorder characterized by excessive daytime sleepiness, great difficulty in awakening, and prolonged sleep time. The exact prevalence of IHS is unknown, but it has been estimated to be approximately 0.005%. In contrast to narcolepsy type 1 which is a well-recognized hypersomnia, the etiology of IHS has been poorly understood. No susceptibility loci for IHS have also been identified, although familial aggregations have been observed among patients with IHS. Narcolepsy type 1 is tightly associated with *HLA-DQB1\*06:02*. However, no significant associations were observed between IHS and HLA alleles. We performed a genome-wide association study and replication studies involving a total of 412 Japanese patients with IHS and 5,345 Japanese healthy individuals to identify genetic variants influencing IHS susceptibility. Genome-wide SNP genotyping was performed with the Affymetrix Genome-Wide Human SNP Array 6.0. The genomic inflation factor of the GWAS was 1.01, suggesting that the effect of population stratification was negligible. Although no SNPs reached the genome-wide significance threshold in the present study, a genetic variant (rs2250870) within an intron of *PDE9A* was suggestively associated with IHS ( $P = 3.0E-05$ ; odds ratio = 1.5). rs2250870 showed a significant association with expression levels of *PDE9A* ( $P = 7.8E-09$ ). The leading SNP in the *PDE9A* region was the same in associations with both IHS and *PDE9A* expression levels. *PDE9A* is a potential target for treatment of diseases, such as depression or stress-induced heart disease. Further, inhibitors of *PDE9A* have been developed as a new potential drug of Alzheimer's disease. Since higher expression levels of *PDE9A* were found in the risk allele of rs2250870 for IHS, *PDE9A* inhibitors and the related compounds might contribute to develop new treatment for IHS.



## PgmNr 2193: Genome-wide association meta-analysis of nicotine dependence reveals novel loci and shared genetic influences with multiple traits: Findings from the iNDiGO (Nicotine Dependence GenOmics) consortium.

### Authors:

B.C. Quach<sup>1</sup>; M. Bray<sup>2</sup>; N.C. Gaddis<sup>1</sup>; M. Liu<sup>3</sup>; T. Palviainen<sup>4</sup>; G.W. Reginsson<sup>5</sup>; S. Zellers<sup>3</sup>; R. Sherva<sup>6</sup>; C.C. Minica<sup>7</sup>; F. Aliev<sup>8</sup>; M. Nothnagel<sup>9</sup>; K.A. Young<sup>10</sup>; H. Young<sup>3</sup>; J. Marks<sup>1</sup>; Y. Guo<sup>1</sup>; A. Waldrop<sup>1</sup>; M.T. Landi<sup>11</sup>; D.W. McNeil<sup>12</sup>; L.A. Farrer<sup>6</sup>; C.A. Markunas<sup>1</sup>; J. Vink<sup>13</sup>; W.G. Iacono<sup>3</sup>; H.R. Kranzler<sup>14,15</sup>; N.L. Saccone<sup>2</sup>; M.C. Neale<sup>8</sup>; P. Madden<sup>2</sup>; M. Rietschel<sup>16</sup>; M.L. Marazita<sup>17</sup>; M. McGue<sup>3</sup>; T.B. Baker<sup>18</sup>; G. Winterer<sup>19</sup>; R. Grucza<sup>2</sup>; D.I. Boomsma<sup>7</sup>; D. Dick<sup>8</sup>; J. Gelernter<sup>20,21</sup>; T.E. Thorgeirsson<sup>5</sup>; N.E. Caporaso<sup>11</sup>; J. Kaprio<sup>4</sup>; K. Stefansson<sup>5</sup>; J.E. Hokanson<sup>10</sup>; S. Vrieze<sup>3</sup>; L.J. Bierut<sup>2</sup>; E.O. Johnson<sup>1</sup>; D.B. Hancock<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) RTI International, Research Triangle Park, NC, USA; 2) Washington University, St. Louis, MO, USA; 3) University of Minnesota, Minneapolis, MN, USA; 4) University of Helsinki, Helsinki, Finland; 5) deCODE Genetics/Amgen, Reykjavik, Iceland; 6) Boston University, Boston, MA, USA; 7) Vrije Universiteit, Amsterdam, The Netherlands; 8) Virginia Commonwealth University, Richmond, VA, USA; 9) Christian-Albrechts University, Kiel, Germany; 10) University of Colorado, Aurora, CO, USA; 11) National Cancer Institute, National Institutes of Health, United States Department of Health and Human Services, Bethesda, MD, USA; 12) West Virginia University, Morgantown, WV, USA; 13) Radboud University, Nijmegen, The Netherlands; 14) University of Pennsylvania, Philadelphia, PA, USA; 15) Crescenz VA Medical Center, Philadelphia, PA, USA; 16) University of Heidelberg, Mannheim, Germany; 17) University of Pittsburgh, Pittsburgh, PA, USA; 18) University of Wisconsin, Madison, WI, USA; 19) Charité – University Medicine Berlin, Berlin, Germany; 20) Yale University, New Haven, CT, USA; 21) VA CT Healthcare Center, West Haven, CT, USA

Cigarette smoking is a leading cause of morbidity and mortality worldwide. Each phase—initiation, regular smoking, nicotine dependence (ND), and cessation—is a heritable trait, but knowledge of the specific genetic loci is incomplete. We performed a 1000 Genomes-imputed genome-wide association study (GWAS) meta-analysis for ND, defined by the Fagerström Test for Nicotine Dependence (FTND), using the largest sample size to date: 58,000 smokers of European (EUR) or African ancestry from 23 cohorts. We identified two novel genome-wide significant associations with ND, *TENM2* (chr5, lead SNP rs1862416,  $P=1.5 \times 10^{-8}$ ) and *MAGI2/GNAI1* (chr7, lead SNP rs2714700,  $P=2.3 \times 10^{-8}$ ), while replicating associations at three known loci. Rs1862416-T and rs2714700-C were associated with increased ND risk across ancestries. For independent testing, we used 33,791 smokers from the UK Biobank with the heaviness of smoking index (HSI), which is based on two FTND items that explain much of the variance in the full 6-item FTND. Rs2714700-C was associated with higher HSI ( $P=0.014$ ), but rs1862416 was not associated ( $P=0.39$ ). Testing specific FTND items in our GWAS cohorts showed rs1862416 associations with items not captured by the HSI, suggesting that the composite ND phenotype enhanced power for discovery but limited its detection in the UK Biobank. Next, using LD score regression on EUR-specific GWAS results, we estimated the common SNP-based heritability of ND (8.6%) and found 13 traits genetically correlated with ND at  $P < 0.0012$  ( $\alpha=0.05/43$  tests): other smoking traits ( $r_g=0.40$ – $0.95$ ), lung cancer ( $r_g=0.66$ ), alcohol dependence ( $r_g=0.56$ ), depression



symptoms ( $r_g=0.40$ ), years of schooling ( $r_g=0.33$ ), coronary artery disease ( $r_g=0.32$ ), neuroticism ( $r_g=0.28$ ), and schizophrenia ( $r_g=0.16$ ). We also applied stratified LD score regression to test genetic correlations with 205 gene expression datasets from brain and other tissues/cell types. We found that genes spanning ND-associated SNPs were significantly enriched for expression in cerebellum ( $P < 1.4 \times 10^{-4}$ ,  $\alpha = 0.05/205$  tests). Our results show the utility of the FTND as a composite measure to expand genetic knowledge of smoking, as the novel loci were not detected in much larger GWAS of single component smoking traits (e.g., ever vs. never smoking with  $N > 1$  million). Also, the genetic correlation of ND with cerebellum-specific expression highlights a potential role for this brain tissue, which is often overlooked in studying the etiology of ND.

## PgmNr 2194: Rigorously decomposing the phenotype of complex psychiatric traits by structured causal modeling: Results from the AGRE/iHART Collection.

### Authors:

N. Stockham <sup>1</sup>; K. Paskov <sup>2</sup>; D.P. Wall <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurosciences, Stanford University, Stanford, CA; 2) Biomedical Informatics, Stanford University, Stanford, CA; 3) Pediatrics and Biomedical Data Science, Stanford University, Stanford, CA

---

Autism Spectrum Disorder (ASD) is a neurodevelopment disorder with consistently high heritability estimates. However, the genetic architecture of inherited ASD risk remains elusive despite many studies with several thousand individuals. A plausible explanation for this lack of discovery is that ASD encompasses several disorders, each with distinct genetic etiology. This work presents a novel statistical framework and computational method (Structured Causal Model PCA; SCMPCA) to rigorously evaluate and further decompose the ASD phenotype into distinct dimensions that are consistent measures of genetic liability. We evaluated current clinical ASD behavioral scoring algorithms under this framework, and fit a set of novel phenotype scoring algorithms (termed Synthetic Genetic Liability (SGL) scores) directly from behavioral data using SCMPCA. Notably, this method can be applied to any complex genetic trait with recurrence in families.

ASD is typically defined by three sub-phenotypes: social interaction deficits, communication deficits, and restricted and repetitive behavioral patterns. Using a Structured Causal Model (SCM), we derived a criterion to assess if a given sub-phenotype scoring algorithm gives a consistent measure of genetic influence. We applied this criteria to the Autism Diagnostic Interview Revised (ADIR) behavioral instrument data in the AGRE/iHART collection (largest WGS multiplex autism dataset: 4242 full siblings, 172 DZ, 192 MZ twins) of 1004 families with multiple children with ASD. We found that the 'Reciprocal Social Interaction' (RSI) score is a consistent measure of genetic influence, whereas the 'Restricted, Repetitive, and Stereotyped Patterns' (RRSP) score is not a consistent measure of genetic influence. The 'Communication' (C) score is a consistent measure of genetic and non-genetic influences. Supporting this finding, genetic similarity between dizygotic twins is very highly correlated with RSI similarity, but not correlated with C or RRSP scores. Strikingly, this finding is supported by recent human biomarker studies from the Karen Parker lab (Oztan, O. 2018) that show high correlation of CSF vasopressin with ADIR RSI, but no correlation with RRSP. Finally, we used the SGL scores fit by SCMPCA to partition the iHART families into phenotypically distinct sub-groups that may reflect shared genetic etiology.

Oztan, O., et al. (2018) 'Cerebrospinal fluid vasopressin and symptom severity in children with autism', *Annals of Neurology*

## PgmNr 2195: The effects of natural selection on the persistence of common genetic risk for physiologic and pathologic traits of the human brain.

### Authors:

F.R. Wendt <sup>1</sup>; C.M. Carvalho <sup>2</sup>; J. Gelernter <sup>1</sup>; R. Polimanti <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Yale School of Medicine, New Haven, CT.; 2) Universidade Federal de São Paulo (UNIFESP), São Paulo, SP, Brazil

---

Genome-wide association studies (GWAS) have identified common genetic variants underlying many brain-related phenotypes, including physiologic and pathologic traits. Different functional categories of genomic variation contribute disproportionately to trait heritability. Common genetic variation underlying several neuropsychiatric phenotypes, such as schizophrenia and autism spectrum disorder, has demonstrated influence by natural selective pressures. It is unclear how universally these pressures influence the heritability across brain-related phenotypes. Investigating 73 brain-related traits analyzed in large-scale GWAS including up to 766,345 individuals of European descent, we detected widespread enrichment of variants in mutation-intolerant regions of the genome and variants under background selection in psychiatric disorders (maximum loss-of-function (LoF) intolerance enrichment: *Tourette syndrome*, 2.11-fold,  $p=3.08 \times 10^{-5}$ ; maximum background selection enrichment: *bipolar disorder*, B-statistic (a measure of nucleotide diversity reduction due to purifying selection), 3.03-fold,  $p=2.87 \times 10^{-5}$ ), brain imaging phenotypes (maximum LoF enrichment: *caudate volume*, 1.87-fold,  $p=0.007$ ; maximum background selection enrichment: *caudate volume*, B-statistic, 3.64-fold,  $p=0.039$ ), and internalizing behavioral traits (maximum LoF enrichment: *recent easy annoyance*, 2.25-fold,  $p=1.25 \times 10^{-5}$ ; maximum background selection enrichment: *recent restlessness*, B-statistic, 5.79-fold,  $p=0.007$ ) and externalizing behavioral traits (maximum LoF enrichment: *recent low energy*, 2.08-fold,  $p=4.81 \times 10^{-6}$ ; maximum background selection enrichment: *frequency of drinking alcohol*, B-statistic, 3.92-fold,  $p=0.002$ ) after correction for multiple testing. Several caudate nucleus imaging phenotypes (i.e., caudate nucleus measured in five out of the six brain imaging phenotypes available for heritability partitioning) robustly demonstrated these signatures at multiple thresholds of background selection dichotomization (i.e., top 2%, top 1%, and top 0.5% of genome-wide scores), supporting its pathological role in development of schizophrenia, bipolar disorder, and alcohol use, and its physiological contribution to educational attainment and cognitive performance. These findings reveal the widespread influence of background selection to remove mutations with large deleterious effects on fitness thereby enriching for variants with small effects now observed as common risk loci for neuropsychiatric traits.

## PgmNr 2196: The brain transcriptomic profiles of immunogenic system in major neuropsychiatric disorders: A cross-disorder study.

### Authors:

Y. Chen <sup>1</sup>; J. Dai <sup>1</sup>; J. Qiu <sup>1</sup>; C. Chen <sup>1</sup>; C. Liu <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Central South University, Changsha, Hunan, China; 2) Department of Psychiatry, SUNY Upstate Medical University, NY, USA

---

Accumulating evidence suggests that inflammation plays a part in several neuropsychiatric disorders. Yet the underlying transcriptomic relationships of the individual diseases to neuroimmunity remains clouded. To understand the role of immunologic changes to these disorders, we assessed gene expression changes related to neuroimmunogenic systems in post-mortem brain samples of individuals with schizophrenia (SCZ), bipolar disorder (BD), autism spectrum disorder (ASD), alcoholism (AAD), major depressive disorder (MDD), Alzheimer's disease (AD) and Parkinson's disease (PD).

Using data from 2540 brain samples, we calculated differential gene expression (DGE) using a linear mixed-effects model. In 1256 immune-related genes collected from databases, we identified differential expressed immune genes to each disorder (AD: 631, ASD: 273, BD: 56, MDD: 26, AAD: 51, PD: 95, SCZ: 216). Meanwhile, we also found immune-related genes fold changes showing significant sex differences in ASD (FDR  $q < 0.05$ ) and MDD (FDR  $q < 0.05$ ). In differential pathway analysis, we found specific immune pathways to each disorder. In one instance, we found highly enriched virus-response related genes only in the AD samples (GO:0009615 Response to virus, FDR  $q = 1.88e-20$ ). To test for immune transcriptome similarity between disorders, we used correlation analysis of immune gene changes across disorder pairs. Results showed significant positive correlation between BD and SCZ ( $\rho = 0.77$ , FDR  $q < 0.05$ ). Whereas, results showed negative correlation between AD and other disorders: SCZ ( $\rho = -0.27$ , FDR  $q < 0.05$ ), BD ( $\rho = -0.33$ , FDR  $q < 0.05$ ), ASD ( $\rho = -0.37$ , FDR  $q < 0.05$ ). Meanwhile, we found that immune gene fold changes were significantly correlated with genetic correlations across the same disease pairs (Pearson'  $r = 0.54$ ,  $p = 0.014$ ). Lastly, we placed individual genes into networks using Weighted Gene Co-Expression Network Analysis (WGCNA), identifying 11 co-expression modules and 3 modules enriched for immune genes. These three modules were also enriched for neuronal development functions, with one enriched for schizophrenia and intelligence GWAS signals.

Our study identified shared changes and disease-specific changes in the immune process across different neuropsychiatric disorders at the transcriptome level. These findings provide new insights into immunogenic systems alterations at transcriptome level in major neuropsychiatric disorders.

## PgmNr 2197: A pilot study of the association between the gut microbiota, sleep quality, and the risk of depression among elderly Taiwanese.

### Authors:

L. Chuang <sup>1</sup>; S. Lin <sup>2</sup>; C. Chung <sup>5</sup>; C. Chiang <sup>2</sup>; M. Yu <sup>3</sup>; Y. Huang <sup>3</sup>; J. Liu <sup>3</sup>; J. Hsieh <sup>4</sup>; P. Kuo <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Nursing, Cardinal Tien Junior College of Healthcare & Management, Yilan County, Taiwan; 2) Department of Medical Research and Education, Taipei Veterans General Hospital, Yuan Shan/Su Ao Branch, Yi-Lan, Taiwan; 3) Department of Nursing, Taipei Veterans General Hospital, Yuan Shan/Su Ao Branch, Yi-Lan, Taiwan; 4) Department of Surgery, Taipei Veterans General Hospital, Yuan Shan/Su Ao Branch, Yi-Lan, Taiwan; 5) Department of Public Health & Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan

---

Among the trend of aging globalization, the major depressive disorder has become one of the most important topics of public health. A community-based study revealed that association between the score of the geriatric depression test and the feature of microbiota composition. The gut microbiota associated with the risk of geriatric depression via the mechanism of the gut-brain axis. The prevalence of the geriatric depression was around 16-25% in Taiwan; however, there was no study in the field of the mechanism between the gut microbiota and the geriatric depression in Taiwan. The present study aimed to explore the association analysis between the risk of geriatric depression and gut microbiota. We recruited 14 participants over 65 years old from Day Care Center, Nursing home and Intermediate Care Center. The questionnaires 15-item Geriatric Depression Scale (GDS) and the Pittsburgh Sleep Quality Index were used to understand the depressive status and subjective sleep parameters, respectively. The fecal samples were examined by Illumina Miniseq or Miseq platform for 16s rRNA sequencing of microbiota features. Linear discriminant analysis effect size (LEfSe) and random forest (RF) procedures were applied to examine the feature of the microbiota composition for geriatric depression. There were 4 (28.6%) participants with depressive status, and the participant who cannot get to sleep within 30 minutes more than once a week was significantly increased the risk for depressive status, showing odds ratio (95% confidence interval) was 4.45 (1.16-17.10). After the RF procedure, Families Mogibacteriaceae and Turicibacteriaceae were the top two feature ranks in contributions for classification accuracy of depressive participants and control. The LEfSe revealed that the relative abundance of Family Prevotellaceae significantly increased in the participants who had trouble pain sleeping more than once a week. Our study indicated the association between the gut microbiota features, sleep quality, and the risk of depression among elderly Taiwanese.

\* Dr. Li-Chung Chuang and Ms. Su-Chu Lin are the joint first authors and contributed equally to the work in the present study.

## **PgmNr 2198: *OXTR* and *AVPR1A* polymorphisms are associated with brain substrates involved in feeling the spirit.**

### **Authors:**

L. Dai <sup>1</sup>; J. King <sup>2</sup>; M. Ferguson <sup>2</sup>; J. Nielson <sup>2</sup>; H. Nazarloo <sup>3</sup>; C. Carter <sup>3</sup>; J. Anderson <sup>2</sup>; J. Korenberg <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Pediatrics, Univ of Utah, Salt Lake City, Utah.; 2) Radiology, Univ of Utah, Salt Lake City, Utah; 3) Kinsey Institute

---

Religious experience is a transformative force that shapes the positive and negative social interactions of individuals and societies. The relationship between self and other, whether beings or dieties, shapes social science and religious philosophy but has common neural circuitry has been elusive. Therefore, the brain systems mediating these may overlap, suggesting the social neuropeptides oxytocin (OT), arginine vasopressin (AVP) and their receptors, the *OXTR* and the *V1a*, respectively, were also involved in the brain networks of religious experience. In our prior studies, we have shown that endogenous OT, but not AVP, was associated with BOLD activation in anterior and posterior cingulate cortex, ventromedial prefrontal cortex and anterior insula, the hubs of default mode and paralimbic networks that are involved in self-awareness and self-other, when individuals felt the spirit. In this report, we explored the hypothesis that the genetic variance of the receptors, *OXTR* and *AVPR1A*, might be independently associated with specific brain substrates in response to three religiously evocative tasks (Quotations Initial, Scripture Reading and Quotations Final), in a cohort of individuals trained to report spiritual responses. Our results showed that DNA polymorphisms of *OXTR* (rs2254298, rs2268498) were associated with left dorsolateral prefrontal cortex, right inferior parietal and left posterior cingulate cortex. The *AVPR1A* promoter-region microsatellite repeat length of RS3 was associated with right ventrolateral orbitofrontal cortex. These brain substrates were also involved in default mode network, self-other and high cortical processes, suggesting the genetic variance, present for birth, possibly acting throughout development, and the other, reflecting the variable interaction and results of a spectrum of developmental and experiential events. Putting together with our prior studies, these results indicate that both the original state of the receptors, and the endogenous peptide, are involved in religious experience, most strongly in regions involved in early developing brain systems but also with the later appearing brain systems for the more complex aspects of self involving others.

## PgmNr 2199: Polygenic risk scores for intelligence and schizophrenia jointly contribute to community functioning in patients with a psychotic disorder.

### Authors:

A.V. Ahola-Olli <sup>1,2,3</sup>; Z. Misiewicz <sup>2</sup>; N. Mars <sup>2</sup>; M. Lähtenvuo <sup>4</sup>; B.M. Neale <sup>1,5</sup>; T. Männynsalo <sup>6</sup>; E. Isometsä <sup>7</sup>; A. Wegelius <sup>7</sup>; J. Hietala <sup>8,9</sup>; W. Haaki <sup>8</sup>; O. Kampman <sup>10,11</sup>; J. Veijola <sup>12,13,14</sup>; J. Tiihonen <sup>4,15,16</sup>; T. Kieseppä <sup>17</sup>; J. Suvisaari <sup>17</sup>; S.E. Hyman <sup>1,18</sup>; M.J. Daly <sup>1,2,19</sup>; A. Palotie <sup>1,2,3,5,20</sup>; SUPER-study

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 2) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 3) Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; 4) Department of Forensic Psychiatry, University of Eastern Finland, Niuvanniemi Hospital, Kuopio, Finland; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 6) Department of Psychiatry, Helsinki City Health Centre, Helsinki, Finland; 7) Department of Psychiatry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 8) Department of Psychiatry, Turku University Hospital, Turku, Finland; 9) Department of Psychiatry, University of Turku, Turku, Finland; 10) Faculty of Medicine and Life Sciences, Tampere University, Tampere, Finland; 11) Department of Psychiatry, South Ostrobothnia Hospital District, Seinäjoki, Finland; 12) Department of Psychiatry, Research Unit of Clinical Neuroscience, University of Oulu, Oulu, Finland; 13) Department of Psychiatry, Oulu University Hospital, Oulu, Finland; 14) Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Finland; 15) Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; 16) Center for Psychiatry Research, Stockholm City Council, Stockholm, Sweden; 17) Department of Public Health Solutions, Mental Health Unit, National Institute for Health and Welfare, Helsinki, Finland; 18) Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts; 19) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 20) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

---

SUPER is a nationwide study conducted in Finland, where health care system is publicly funded and supported housing facilities are accessible to every citizen. National health care registers cover every hospital visit since 1969 and prescription drug purchase since 1995. Data on education, employment and cause of death can be also drawn from registers and linked to other information based on an individual national social security number.

Subtle changes in cognitive performance and social behavior can be observed years before onset of psychosis. Different aspects of functional capacity are poorly correlated, suggesting that they might have different etiological backgrounds. Functional deficiencies are refractory to antipsychotic therapy, further suggesting that they are independent from mechanisms causing psychotic behavior.

To evaluate biological basis of functional deficiencies in patients with a psychosis, we used polygenic risk scores (PRSs) for various neurodevelopmental disorders. Risk scores were derived using LDpred software and summary statistics from recent genome-wide association studies. SUPER recruited

10,382 individuals (age  $46.7 \pm 14.9$  years, 49.5% females) with a history of at least one psychotic episode. Functional capacity was assessed with a structured interview, which contained questions regarding past school difficulties, current employment, housing and relationship status. Genotyping with Illumina Global Screening Array has so far been performed for 7,310 subjects. Imputation was performed using the population-specific SISu panel as a reference panel.

PRS for schizophrenia (SCZ) was associated with need for supported housing at the time of study visit (OR 1.08, 95% CI 1.02-1.14,  $P=4.9 \times 10^{-3}$ ), whereas it had no association with subjective difficulties at school. However, subjective difficulties at school were associated with PRS for intelligence (OR 0.87, 95% CI 0.83-0.91,  $P=4.7 \times 10^{-8}$ ). Subjects who had both subjective difficulties at school in childhood and needed supported housing in adulthood had lower PRS for intelligence and higher PRS for SCZ compared to those with constantly good community functioning.

In conclusion, the poor lifelong functional capacity in subjects with a history of psychosis is associated with an unfavorable polygenic profile for two distinct neuropsychiatric traits. We will further confirm these findings using registry data.



## PgmNr 2200: Biobank-based genome-wide analysis for post-traumatic stress disorder.

### Authors:

C. Chen<sup>1,2,3,4</sup>; T. Ge<sup>2</sup>; A. Zheutlin<sup>2</sup>; K. Choi<sup>2,4,5</sup>; J. Smoller<sup>2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Translational Biology, Biogen, Cambridge, MA, 02142; 2) Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, 02114; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, 02114; 4) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, 02142; 5) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, 02115

---

Biobanks that integrate electronic health records (EHR) with genome-wide data create vast growing resources for genetic research. Here, we leveraged Partners HealthCare Biobank, which links EHR with genome-wide data, to perform genome-wide analyses for posttraumatic stress disorder (PTSD). We identified 637 PTSD cases ( $\geq 2$  PTSD ICD9 codes) and 18,295 controls (0 PTSD codes) of European ancestry. We first genetically validated EHR-based PTSD against 2 external PTSD GWAS from UK Biobank ( $N=123,794$ ) and Psychiatric Genomic Consortium (PGC;  $N=9,954$ ). Using LD-score regression, we showed that EHR-based PTSD has SNP-based heritability ( $h^2_g$ ) similar to conventionally-ascertained PTSD ( $h^2_g=0.11$ ,  $p=1.04 \times 10^{-6}$ ), and substantial genetic correlation ( $r_g$ ) with external PTSD samples ascertained by diagnosis/systematic interview ( $r_g=0.53$  for PGC PTSD GWAS,  $p=0.03$ ), but lower  $r_g$  with PTSD identified through self-report survey ( $r_g=0.37$  for UKBB PTSD GWAS,  $p=0.0004$ ). Using external GWAS summary statistics, we also confirmed significant  $r_g$  between EHR-based PTSD and other psychiatric traits, such as ADHD ( $r_g=0.40$ ,  $p=2.63 \times 10^{-6}$ ), depression ( $r_g=0.39$ ,  $p=2.01 \times 10^{-5}$ ), schizophrenia ( $r_g=0.22$ ,  $p=0.0004$ ), educational attainment ( $r_g=-0.30$ ,  $p=1.29 \times 10^{-7}$ ), and cognitive performance ( $r_g=-0.30$ ,  $3.88 \times 10^{-7}$ ). Second, we used UK Biobank and PGC PTSD GWAS meta-analysis to generate polygenic risk scores for PTSD ( $PRS_{PTSD}$ ) in Partners Biobank with 2 methods: 1) LD-clumping with association p-value thresholding and 2) Bayesian regression with continuous shrinkage priors (PRS-CS). We showed that PRS-CS with causal variant proportion parameter set to 10% gives the best prediction ( $R^2=0.009$ ,  $p=5.21 \times 10^{-11}$ ). We also showed that patients with top 5%  $PRS_{PTSD}$  have a 1.69-fold increased risk for PTSD (95% CI=1.25-2.29), compared with the rest of the patient population. Finally, we performed a phenome-wide association study (PheWAS) to explore the capacity of  $PRS_{PTSD}$  to predict a range of EHR-based health outcomes using 853 PheWAS codes derived from ICD codes.  $PRS_{PTSD}$  significantly predicts the risk of mood disorders, depression, anxiety disorders, substance addiction, and chest pain. For the first time, we showed that PTSD patients identified through diagnostic codes from EHR are genetically comparable to conventionally-ascertained PTSD patients. Furthermore, we showed potential applications of PRS in a real-world clinical system to identify patients with higher risk of PTSD and other comorbidities of PTSD.

## PgmNr 2201: Insights into the genetics of the mood disorder spectrum.

### Authors:

J.R.I. Coleman <sup>1,2</sup>; G. Breen <sup>1,2</sup>; H. Gaspar <sup>1,2</sup>; J. Bryois <sup>3</sup>; Major Depressive Disorder and Bipolar Disorder Working Groups of the Psychiatric Genomics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology, and Neuroscience, King's College London, London, United Kingdom; 2) National Institute for Health Research Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust and King's College London, London, United Kingdom; 3) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

---

Mood disorders affect 10-20% of the population, ranging from brief, mild episodes to severe, incapacitating conditions that markedly impact lives. Multiple approaches have shown considerable sharing of genetic risk factors between unipolar and bipolar mood disorders.

We used genome-wide association study results for major depression and bipolar disorder to investigate the molecular basis of the shared genetic liability to mood disorders. We meta-analysed the latest results from the Psychiatric Genomics Consortium (PGC) major depression and bipolar disorder cohorts, along with an additional major depressive disorder cohort from UK Biobank (185,285 cases, 439,741 controls, non-overlapping N = 609,424). 73 loci reached genome-wide significance in the meta-analysis, with additional loci significant in subtype and depression-only analyses. More genome-wide significant loci from the PGC analysis of major depression (39/44, 89% of the loci) than from the PGC analysis of bipolar disorder (4/19, 21%) reached genome-wide significance in the meta-analysis. Genetic correlations calculated between major depression and bipolar disorder subtypes revealed that type II bipolar disorder correlates strongly with major depression. Integrating the results with systems biology information, we implicated pathways and neuronal subtypes that highlight similarities but also potential differences between major depression and bipolar disorder.

Our results reflected major depression more than bipolar disorder, perhaps due to the larger sample size for major depression, but also perhaps because depression is the predominant common feature of mood disorders. Overall, these results provide evidence for a genetic mood disorders spectrum.

## PgmNr 2202: Common and rare damaging variants both contribute to familial form of bipolar disorder and explain difference in clinical manifestations in multiplex families.

### Authors:

E. Courtois <sup>1,2,3</sup>; A. Henrion <sup>1,2,3</sup>; C. Barau <sup>4</sup>; P. Le Corvoisier <sup>5</sup>; M. Leboyer <sup>1,2,3,6</sup>; B. Etain <sup>3,7,8,9</sup>; S. Jamain <sup>1,2,3</sup>; FACE-BD group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) INSERM U955, Psychiatrie Translationnelle, Créteil, France; 2) Université Paris Est, Faculté de Médecine, Créteil, France; 3) Fondation FondaMental, Créteil, France; 4) AP- HP, Hôpital H.Mondor - A. Chenevier, Plateforme des Ressources Biologiques, Créteil, France; 5) INSERM, Centre d'Investigation Clinique 1430 and APHP, Créteil, France; 6) AP-HP, DHU PePSY, Pôle de Psychiatrie et d'Addictologie des Hôpitaux Universitaire Henri Mondor, Créteil, France; 7) AP-HP, GH Saint-Louis - Lariboisière - F. Widal, Département de Psychiatrie et de Médecine Addictologique, Paris, France; 8) Université Paris Diderot, Sorbonne Paris Cité, Paris, France; 9) INSERM, UMR-S1144, Paris, France

---

Bipolar disorder (BD) is a severe psychiatric disorder with a complex pattern of inheritance. Genome wide association studies have revealed an additive polygenic contribution of common low risk alleles, explaining up to 25% of the overall genetic variance in the heritability of BD. Rare high penetrant variants might partly explain this missing heritability and family studies with a high density of illness are of particular relevance for such study.

To estimate the contribution of common and rare variants in BD vulnerability, we combined both whole exome sequencing and genotyping arrays in 8 multiplex families with BD (39 subjects, of whom 22 were affected). We showed a higher BD polygenic risk scores (PRS) in multiplex families when compared with 1,636 individuals of general population, but no difference was observed between affected and unaffected subjects. When extending this analysis to a cohort of 445 BD patients, no difference in BD PRS was observed between individuals with or without family history of BD. However, we reported a higher schizophrenia (SZ) PRS in familial cases of BD than in unfamilial ones.

The analysis of rare damaging variations in constraint genes shared by affected subjects in multiplex families revealed a single interaction network enriched in neuronal and developmental biological pathways, as well as in the regulation of gene expression. We sequenced 241 additional patients with BD and compared the rare mutation frequency for these genes with the 21,071 individuals from the non-Finnish and non-psychiatric European subjects of the Exome Aggregation Consortium (ExAC) cohort. Four genes showed a higher mutation rate in patients with BD than in the ExAC population after correction for multiple testing, including three involved in epigenetic regulations. Interestingly, these patients showed a specific clinical manifestation. Furthermore, we computed a score combining information from BD and SZ PRS and showed a significant negative correlation of this score with rare damaging variations specifically in unaffected individuals of multiplex families.

Altogether, our results suggest that common and rare genetic variants both contribute to the familial aggregation of BD and that this genetic architecture may explain the heterogeneity in clinical manifestations in multiplex families.

## PgmNr 2203: Genes regulated by BCL11B during T-cell development are enriched for *de novo* mutations found in schizophrenia patients.

### Authors:

L. Fahey <sup>1,2</sup>; P. Ó Broin <sup>2</sup>; G. Donohoe <sup>1</sup>; D. Morris <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cognitive Genetics and Cognitive Therapy Group, Centre for Neuroimaging & Cognitive Genomics, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Galway, Ireland.; 2) School of Mathematics, Statistics and Applied Mathematics, National University of Ireland Galway, Galway, Ireland.

---

Schizophrenia (SCZ) is a common but severely debilitating adult-onset mental illness characterized by hallucinations, delusions, and a lack of desire to accomplish goals or form social relationships. Cognitive deficits (affecting memory, IQ or attention) are also a core feature of SCZ. However, these deficits are not effectively targeted by current treatments, primarily owing to a lack of understanding of the shared biology of SCZ and cognition. Abnormal neurodevelopment contributes to SCZ risk but there is also evidence to support a role for immune dysfunction in SCZ. *BCL11B* is associated with SCZ in genome-wide association studies (GWAS) and is a transcription factor involved in regulating the differentiation and development of cells in both the central nervous system and the immune system. Here, we use functional genomics data from studies of *BCL11B* to investigate the contribution of neuronal and immune processes to SCZ pathophysiology. We generated three gene-sets that contain the targets of BCL11B in (i) brain striatal cells (n=220 genes), (ii) Thy3 developing T-cells (n=74 genes) and (iii) Thy4 developing T-cells (n=560 genes). For each gene-set, the BCL11B targets were identified using an integrated analysis of differential gene expression data and ChIP-seq binding data. We tested each gene-set for enrichment of genes associated with SCZ or cognition using MAGMA gene-set analysis (GSA) and summary statistics from GWAS. Enrichment of SCZ *de novo* mutations was tested in our gene-sets with denovolyzeR using data from multiple studies that have identified genes containing *de novo* mutations through exome sequencing of SCZ probands and their parents (n=1,024 trios). GSA of GWAS did not identify evidence of enrichment of SCZ or cognition genes in our gene-sets. Analysis of *de novo* mutations did identify that the Thy4 gene-set was enriched for genes containing protein altering mutations (p=0.0007). When this gene-set was divided up into genes that were either up- or down-regulated upon *BCL11B* knockout, the enrichment signal was coming from the up-regulated genes (p=0.0002). Pathway analysis of these up-regulated genes identified 'Interferon alpha/beta signaling' and 'Cytokine signaling in immune system' as biological pathways that are enriched for these genes. These analyses, leveraging a GWAS-identified SCZ risk gene and functional genomics datasets, indicate that *de novo* mutations in immune pathways contribute to SCZ risk.

## PgmNr 2204: Dissection of genetic findings of psychiatric disorders using three-dimensional regulatory chromatin interactions from adult and fetal cortex.

### Authors:

P. Giusti-Rodríguez <sup>1</sup>; L. Lu <sup>2</sup>; Y. Yang <sup>3</sup>; C.A. Crowley <sup>3</sup>; X. Liu <sup>2</sup>; J. Bryois <sup>4</sup>; I. Juric <sup>5</sup>; J.S. Martin <sup>3</sup>; A. Abnoui <sup>6</sup>; N. Ancalade <sup>1</sup>; N.J. Bray <sup>7</sup>; G. Breen <sup>8,9</sup>; C.M. Bulik <sup>4,10</sup>; P.R. Jansen <sup>11,12</sup>; Y. Li <sup>2</sup>; S. Marzi <sup>13,14</sup>; J. Mill <sup>13,14</sup>; M.C. O'Donovan <sup>7</sup>; M.J. Owen <sup>7</sup>; A.F. Pardiñas <sup>7</sup>; D. Posthuma <sup>12</sup>; J.E. Savage <sup>12</sup>; C.A. Stockmeier <sup>15</sup>; J. TR Walters <sup>7</sup>; S. Yao <sup>4</sup>; G.E. Crawford <sup>16</sup>; F. Jin <sup>2</sup>; M. Hu <sup>6</sup>; Y. Li <sup>1,3</sup>; P.F. Sullivan <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, UNC-Chapel Hill, Chapel Hill, NC.; 2) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, US; 3) Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 5) Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA; 6) Cleveland Clinic, Department of Quantitative Health Sciences, Cleveland, OH, US; 7) Cardiff University, MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff, Wales, UK; 8) King's College London, Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, London, UK; 9) South London and Maudsley Hospital, UK National Institute for Health Research Biomedical Research Centre, London, UK; 10) University of North Carolina, Department of Psychiatry, Chapel Hill, NC, US; 11) Erasmus University Medical Center, Department of Child and Adolescent Psychiatry, Rotterdam, NL; 12) Vrije Universiteit Amsterdam, Department of Complex Trait Genetics, Amsterdam, NL; 13) King's College London, London, UK; 14) University of Exeter, University of Exeter Medical School, Exeter, UK; 15) University of Mississippi Medical Center, Department of Psychiatry and Human Behavior, Jackson, MS, US; 16) Duke University, Department of Pediatrics, Division of Medical Genetics, Durham, NC, US

---

Genome-wide association studies (GWAS) for psychiatric disorders have been extremely successful. Interpretation of most GWAS findings is complicated by the presence of many significant and highly correlated associations, mainly located in non-coding regions. Chromatin conformation capture methods enable the identification of 3D chromatin interactions *in vivo* and can clarify GWAS findings. We have generated high-resolution 3D chromatin interactome of adult and fetal cortex using “easy Hi-C” (eHi-C), a variation of Hi-C that yields high quality Hi-C libraries using a much lower input. To aid interpretation of the eHi-C readouts, we generated or assembled RNA-seq, ATAC-seq, and ChIP-seq data from adult and fetal cortex. We used in-house pipelines for the analysis of compartment A/B, to call frequently interacting regions (FIREs), and to identify topologically associated domain (TAD) boundaries and chromatin interactions. Our Hi-C readouts agreed with external datasets and captured key biological processes of adult and fetal brain. We focused our analysis on high-confidence regulatory chromatin interactions (HCRCI), which involve enhancers or promoters, and we used these 3D data to functionally connect genetic results for multiple psychiatric disorders and brain traits to specific genes. Finally, we systematically compared how HCRCI fared against other established methods for connecting single nucleotide polymorphisms (SNP) to specific genes. Our data suggests that location- based approaches provide a limited view of the complexity of psychiatric GWAS

findings, and provide support for the idea that, following genetic identification of a locus for a complex disease like schizophrenia, it is essential to incorporate knowledge of chromatin interactions in a disease-relevant tissue.

## PgmNr 2205: The genetic overlap between body composition, glycemic, and psychiatric traits is age- and sex-dependent.

### Authors:

C. Hübel <sup>1,2</sup>; H. Gaspar <sup>1</sup>; J. Coleman <sup>1</sup>; K. Hanscombe <sup>1</sup>; K. Purves <sup>1</sup>; I. Prokopenko <sup>2</sup>; the. MAGIC investigators <sup>2</sup>; the. Psychiatric Genomic Consortium <sup>3</sup>; P. O'Reilly <sup>1</sup>; C. Bulik <sup>3</sup>; G. Breen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) King's College London, London, United Kingdom; 2) Imperial College London, London, UK; 3) Karolinska Institutet, Stockholm, Sweden

---

Alterations in appetite regulation, energy intake, and physical activity are key symptoms of psychiatric disorders. Negative and positive genetic correlations between anorexia nervosa and attention-deficit/hyperactivity disorder (ADHD) with body mass index implicate a shared genetic etiology. Additionally, psychiatric disorders show sex differences with more females being affected by major depressive disorder and anorexia nervosa and more males by schizophrenia and autism. The body mass compartments, such as fat mass and fat-free mass, also differ between the sexes. We generated sex-specific GWAS of fat-free mass and fat mass in a healthy and medication-free subsample of the UK Biobank (all N = 155,961), the MAGIC consortium supplied GWAS of glycemic traits (N = 140,583), and the Psychiatric Genomics Consortium and others of 17 psychiatric disorders and behavioural traits (up to N = 217,568). We then calculated sex-specific genetic correlations across all these traits using bivariate linkage disequilibrium score regression. Two notable patterns emerged: (1) anorexia nervosa, schizophrenia, obsessive-compulsive disorder, and education years were negatively associated with body fat % and fat-free mass, and, (2) ADHD, alcohol dependence, insomnia, and heavy smoking were positively associated with body fat % and fat-free mass. Anorexia nervosa showed a significantly stronger genetic correlation with body fat % in females than in males, whereas education years showed a stronger genetic correlation with fat mass in males than in females. Most genetic correlations were significant with GWAS of body composition traits in late adulthood, whereas only ADHD showed a genetic correlation with obesity in childhood. Mendelian randomization analyses indicated evidence consistent with schizophrenia, anorexia nervosa and more years of education causing decreased fat mass while, in contrast, that higher body fat % may be a causal risk factor for ADHD and heavy smoking. These findings generate new hypotheses for targeted preventive strategies.

## PgmNr 2206: GWAS of anxiety disorders in pediatric-enriched cases.

### Authors:

H. Hakonarson <sup>1,2</sup>; H.S. Hain <sup>1</sup>; X. Chang <sup>1</sup>; J.J. Connolly <sup>1</sup>; F.D. Mentch <sup>1</sup>; P.M.A Sleiman <sup>1,2</sup>; The electronic medical records and genomics consortium (eMERGE)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.; 2) Department of Pediatrics, the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Anxiety disorders as a group are the most prevalent mental health condition worldwide, with approximately one in nine people meeting relevant diagnostic across a 12-month period. Relevant disorders include agoraphobia, generalized anxiety disorder, panic disorder, phobias, selective mutism, separation anxiety disorder, and social anxiety disorder. While heritability estimates for anxiety disorders broadly range from ~30–50%, susceptibility loci have not been reliably validated. In collaboration with our colleagues at the electronic medical records and genomics (eMERGE) consortium, we developed an electronic algorithm for identifying anxiety cases and controls across the pediatric-adult spectrum. We conducted a genome-wide association study on 2,529 European individuals with anxiety and 3,414 ancestry-matched controls selected from eMERGE. Whole genome imputation was performed by the Michigan Imputation Server, while association between single nucleotide polymorphisms (SNPs) and anxiety was tested by SNPTTEST using linear regression, adjusting for age, gender, site, and ancestry principal components (PC).

Although no tested SNP surpassed the traditional genome-wide significance threshold, we report that 1q23.3 (rs7526815,  $3.5 \times 10^{-6}$ ), 7q36.1 (rs855738,  $5.4 \times 10^{-7}$ ), 8p23.2 (rs10503346,  $1.5 \times 10^{-6}$ ), 16p23.2 (rs73574279,  $1.3 \times 10^{-6}$ ) were nominally associated with the risk of anxiety. Further analysis of eQTL data from GTEx indicated that genotypes of rs7526815 were significantly associated the expression level of *VANGL2* in tissue from cerebellum ( $6.5 \times 10^{-9}$ ). It has been reported that products of *VANGL2* may be involved in the development of the neural plate, and associated with several neurological disorders.



## PgmNr 2207: Schizophrenia polygenic risk score analysis in 22q11.2 deletion syndrome.

### Authors:

A.M. Holleman<sup>1</sup>; T. Monfeuga<sup>2</sup>; H.R. Johnston<sup>1</sup>; I. Cleynen<sup>3</sup>; Y. Zhao<sup>4</sup>; D.M. McDonald-McGinn<sup>5,6</sup>; R.E. Gur<sup>6</sup>; S.T. Warren<sup>1</sup>; J. Vermeesch<sup>3</sup>; B.S. Emanuel<sup>5</sup>; B.E. Morrow<sup>4</sup>; C.E. Bearden<sup>7</sup>; M. Van den Bree<sup>2</sup>; M. Owen<sup>2</sup>; J.A.S. Vortsman<sup>8</sup>; E. Chow<sup>8</sup>; A. Swillen<sup>3</sup>; J.G. Mulle<sup>1</sup>; A.S. Bassett<sup>8</sup>; N. Williams<sup>2</sup>; M.P. Epstein<sup>1</sup>; International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Emory University, Atlanta, Georgia, USA.; 2) Cardiff University, Cardiff, Wales, UK.; 3) KU Leuven, Leuven, Belgium.; 4) Albert Einstein College of Medicine, New York City, New York, USA.; 5) Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.; 6) University of Pennsylvania, Philadelphia, Pennsylvania, USA.; 7) University of California-Los Angeles, Los Angeles, California, USA.; 8) University of Toronto, Toronto, Canada

---

**Background:** ~25% of individuals with 22q11.2DS develop schizophrenia (SZ), which is >20 times the SZ risk in the general population. As for other genetic conditions that are not fully penetrant, there may be a contribution by other genetic variants. Polygenic risk score (PRS) analysis examines the collective contribution by common genetic variants to risk, and has shown that SZ in the general population has a polygenic component explaining ~7% of variance on the liability scale. The present study used PRS methods to investigate a role for common variants in SZ among patients with 22q11.2DS.

**Methods:** Whole genome sequencing (WGS) data were available from 432 patients with 22q11.2DS of European descent (212 SZ cases; 220 non-psychotic controls). Additional European data were obtained from a general population SZ case-control study (10,791 CLOZUK cases; 24,391 WTCCC controls) to compare PRS between 22q11.2DS and general population samples (utilizing genotypes imputed from array data). PRS were constructed using the PGC SZ GWAS (SCZ2) summary results as training data. When CLOZUK/WTCCC samples were included among the target samples, the PGC SCZ2 dataset excluding CLOZUK/WTCCC samples was used as training data. Associations between phenotype and PRS were examined using logistic regression adjusting for sex and ancestry.

**Results:** PRS analysis of the 22q11.2DS WGS data showed that PRS explained 7.7% of variance in SZ on the observed scale ( $p=6.73 \times 10^{-6}$ ), with one standard deviation increase in PRS yielding a 1.77 increased odds of SZ (95% CI: 1.41-2.22). Comparison of idiopathic SZ cases with 22q11.2DS SZ cases revealed idiopathic cases to have higher PRS (OR=1.31; 95% CI: 1.10-1.55;  $p=0.012$ ).

**Discussion:** 22q11.2DS patients with SZ have significantly greater PRS, consisting of common SZ risk alleles, as compared with 22q11.2DS non-psychotic controls, suggesting variants outside the 22q11.2 deleted region contribute to SZ risk in 22q11.2DS, much like in the general population. In addition, lower PRS among patients with 22q11.2DS and SZ compared with idiopathic cases indicates that individuals with this large SZ-associated copy number variant may require lesser contribution by common genetic variants for SZ to develop. These findings contribute to understanding the

mechanisms underlying SZ, both among individuals with 22q11.2DS and within the general population.

## PgmNr 2208: Genome-wide admixture mapping of DSM-IV alcohol dependence in African American population.

### Authors:

D. Lai <sup>1</sup>; A. Agrawal <sup>2</sup>; T. Foroud <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical and Molecular Genetics, Indiana Univ School of Medicine, Indianapolis, IN; 2) Department of Psychiatry, Washington University School of Medicine, St. Louis, MO

---

The prevalence of alcohol dependence is significantly lower in African Americans (AA) as compared with European Americans (EA). In addition to sociocultural and contextual factors that might contribute to this population difference, there is known genetic heterogeneity as well. For example, rs2066702 in *ADH1B* gene has been linked to AD in AA only, while rs1229984, which is amongst the single strongest contributors to AD in EAs, is not found in AA. However, studies searching for AA-specific AD genes are limited. Furthermore, due to their African Ancestry, AA have more variants and faster decay of linkage disequilibrium across physical distance, resulting in the need for more tag variants to achieve genomic coverage. Principle components (PCs) while accounting for global ancestry may not fully account for local ancestry in such recently admixed populations. Admixture mapping has been successfully applied to test the association between local ancestry and ancestry-specific disease etiology. Here, using AA samples from the Collaborative Study on the Genetics of Alcoholism (COGA, 875 AD cases and 940 controls) as a discovery sample, we performed genome wide (632,882 SNPs) admixture mapping of AD. Two independent AA samples, Study of Addiction: Genetics and Environment (SAGE, 400 cases and 341 controls), and Alcohol Dependence GWAS in European and African Americans (Yale-Penn, 1552 cases and 491 controls), were used for replication. SHAPEIT2 was used to phase haplotypes, then RFmix (v1.5.4) was used to infer local ancestry. At each locus, association between AD and African ancestry allelic count was tested with study specific covariates (sex and birth cohorts for COGA and SAGE, sex and age for Yale-Penn). Genome wide significant threshold was determined as 1.83E-04 after adjusting 273 effective ancestry blocks estimated using autocorrelation method. One region on chromosome 11 (84,844,920-85,623,607 bp,  $p=1.2E-04$ ) was significant and replicated in SAGE ( $p=0.03$ ) but not in Yale-Penn ( $p=0.42$ ). In this region, having African ancestry increases the risk of AD. Studies have linked this region to Schizophrenia, Attention Deficit/Hyperactivity Disorder. Multiple genes (e.g. *DLG2*, *SYTL2*, etc.) are located in this region and further study is needed to pinpoint the causal gene. The finding of this study wasn't nominated by any previous genetic studies of AD in AA and highlighted the importance of performing admixture mapping in admixed populations.

## PgmNr 2209: Identifying cognitive genes for nootropic drug repurposing.

### Authors:

M. Lam<sup>1,2,3,4</sup>; T. Lencz<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Feinstein Institute of Medical Research, New York, NY.; 2) The Zucker Hillside Hospital, New York, NY; 3) Stanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, MA; 4) Analytic and Translational Genetics Unit, MGH, Boston, MA

---

**Background:** Cognitive traits have been demonstrated to be strongly genetic correlated with psychiatric disorders and physical traits. In the last 2 years, several cognitive GWASs have been reported, identifying more than 200 loci across 300,000 individuals in the population. Here, we aim to consolidate GWAS evidence for cognitive performance, and perform novel downstream analysis that aims to identify candidate genes for downstream drug re-purposing.

**Methodology:** We leverage on two large-scale GWASs in the recent literature (PMID 29844566 & 29942086), which reported 148 and 205 genome-wide significant loci, respectively. Summary statistics were harmonized and meta-analyzed using Multi-Trait Analysis of GWAS [MTAG] and METAL 2.0, which is optimized for handling sample overlaps. Independent loci were identified, and downstream analysis was carried out using MAGMA pathway analysis and S-Predixcan/SMultixcan Transcriptomic Wide Analysis (with GTEx 7.0 brain samples as reference). Significant TWAS and MAGMA genes were looked up in multiplem chemoinformatic databases (DGI, KI, and a review on 'druggability' (PMID 28356508).

**Results:** A total of 264 genome-wide significant loci were identified after harmonizing across the two input GWAS; meta-analysis and QC filtering procedures identified 200 loci that could be followed up for downstream analysis. After Bonferroni correction, 558 genes and 444 genes were significant in the MAGMA and TWAS analysis respectively. The intersection of significant GWAS loci, MAGMA and TWAS genes yielded 81 high-confidence genes associated with cognition that could be searched in drug repurposing databases. Of these, 24 genes (~30%) were found to be promising candidates for drug repurposing.

**Discussion:** Preliminary data analysis suggest that the following genes -- CYP2D6, MST1, IL27, MST1R, SULT1A2, GPX1, SEMA3F, NEK4, BTN3A2, SULT1A1, CHST10, ITIH3, SREBF2, DAG1, ITIH4, NAGA, ATP2A1, ITIH1, CAMKV, ALMS1, NDUFS3, RBL2, PBRM1, PSMC3 -- identified in large-scale GWAS meta-analysis for cognition may be candidates for nootropic drug repurposing. Further work is necessary to further identify biological properties of these candidate genes.

## PgmNr 2210: Sex-stratified genome-wide association study of smoking behavior identified sex-specific genetic loci and functional regions in the brain.

### Authors:

B. Li <sup>1</sup>; K. Xu <sup>2,3</sup>; H. Zhao <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Yale School of Public Health, New Haven, Connecticut.; 2) Yale School of Medicine, New Haven, CT;  
3) Veterans Affairs Connecticut Healthcare System, West Haven, CT

---

With great efforts made by large research consortia, recent genome-wide association studies (GWASs) of smoking have identified hundreds of associated loci. However, the distinct genetic architecture of smoking behavior between women and men remains understudied. We analyzed sex-stratified GWASs of smoking initiation (smkInit) and smoking cessation (smkQuit) in the UK Biobank (women N~180,000, men N~156,000). We identified 81 genetic loci significantly associated with smkInit in the combined-sex GWAS, 22 loci in the female GWAS, and 21 loci in the male GWAS. Twelve loci were sex-specific, including 5 female-specific loci (5q14.3, 10q26.13, 15q26.1, 16p11.2, and 18q21.2) mapped to *TMEM161B-AS1*, *TACC2*, *RGMA*, *ZNF689*, and *DCC* by the nearest location, and 7 male-specific loci (2p16.3, 2q37.3, 4q34.3, 10p12.31, 12q24.13, 1p21.3, and 13q31.1) mapped to *NRXN1*, *OTOS*, *TENM3*, *CASC10*, *RBM19*, *LOC101928241*, and *LINC01068*. These sex-specific loci were not identified to be significant in the combined-sex or the opposite sex, hinting at sex-specific genetic variants of smoking initiation for women and men. Four sex-specific loci were novel in comparison to recent GWAS results reported in Liu et al. (2019) and Linnér et al. (2019). To better understand the biological mechanism underlying different genetic risks of smoking behavior, we estimated the overall and partitioned SNP-based heritability of smkInit for female and male. The estimated heritability was comparable among combined-sex ( $h^2=0.09$ ,  $se=0.003$ ), female-specific ( $h^2=0.10$ ,  $se=0.005$ ), and male-specific ( $h^2=0.11$ ,  $se=0.005$ ) GWASs. Partitioned heritability analysis identified brain cingulate gyrus ( $p=2.5E-04$ ) as a female-specific region significantly enriched for the heritability of smkInit. Brain dorsolateral prefrontal cortex ( $p=5.2E-06$ ), angular gyrus ( $p=1.4E-04$ ), and hippocampus middle ( $p=6.4E-04$ ) were identified as male-specific regions. Brain anterior caudate was significantly enriched for both women ( $p=2.2E-04$ ) and men ( $p=2.9E-04$ ). These regions are functionally involved in decision-making. Combined-sex GWAS identified 5 loci for smkQuit and none was sex-specific. Results from the sex-stratified genetic study highlighted possibly distinct genetic architecture of smoking initiation between women and men. While we appreciate the power gain of meta-analysis, we stressed the importance of uncovering biologically interpretable associations and distinct mechanisms accounting for heterogeneity within the population.

## PgmNr 2211: Examining mtDNA variants and sleep disturbances in major depressive disorder.

### Authors:

L. Melhuish Beaupre<sup>1,2</sup>; V. Goncalves<sup>1,3</sup>; J. Kennedy<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Brain Sciences Research Department, Centre for Addiction and Mental Health, Toronto, Canada; 2) Institute of Medical Sciences, University of Toronto, Toronto, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, Canada

---

**Background:** Major depressive disorder (MDD) is a debilitating mood disorder that affects 11-14% of the global population. A core symptom of MDD is sleep disturbance, mainly insomnia and hypersomnia, affecting approximately 75% of individuals. Sleep disturbances might constitute a biotype of MDD. Unfortunately, MDD with accompanied sleep disturbances poses a substantial problem associated with a poorer course of the disease, and a substantial economic burden. Although it is a heritable feature, the genetics of these sleep disturbances is poorly understood. Thus far, only a select number of core circadian genes in the nuclear genome have been investigated, with positive findings for *CLOCK* and *TIMELESS*.

There is compelling evidence that mitochondria play a role in sleep-wake regulation, probably adjusting energy metabolism for day/night periods. To the best of our knowledge, nobody has investigated variants in the mitochondrial genome (mtDNA) thus far. Based on the above, we hypothesize that variants in mtDNA are associated with sleep disturbances in MDD.

**Method:** We selected a subset of depressed individuals from the CAMH IMPACT study, which had sleep disturbance measures available. We had N=135 mtDNA variants on the Illumina Omni microarray genotyped in this sample. All participants had a Beck Depression Inventory (BDI) score greater than 19. Sleep disturbances were measured using the BDI #1.6 (change in sleep), and Udvalg for Kliniske Undersogelser Side Effect Rating Scale (UKU) #1.7 (increased sleep) and #1.8 (decreased sleep). We performed a binary logistic regression for association testing. All individuals (N=293) were genetically of European descent.

**Results:** The top findings from our data were for SNPs from *MT-ND4* ( $p=0.004$ ) and *MT-ND6* ( $p=0.006$ ) and altered sleep as well as a variant found in the mtDNA control region with altered sleep ( $p=0.01$ ) and increased sleep ( $p=0.01$ ).

**Discussion:** We were unable to detect any mtDNA variant that survive multiple hypothesis testing. However, our data is suggestive that variants in mtDNA may play a role in sleep disturbances in MDD. Overall, given the risks associated with MDD and concurrent sleep disturbances, it is vital we gain a better understanding of the genetics underlying them so that we can improve diagnosis and treatment regimens for these individuals. As such, further research in a larger sample is warranted.

## PgmNr 2212: Presence of copy-number variants in the metabotropic glutamate receptor gene network and psychiatric comorbidities of attention deficit hyperactivity disorder.

### Authors:

A.I. Nesbitt<sup>1</sup>; H.S. Hain<sup>1</sup>; I. Slaby<sup>1</sup>; F. Mentch<sup>1</sup>; J.T. Glessner<sup>1</sup>; H. Hakonarson<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Center of Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA.; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

---

Psychiatric comorbidities are commonly observed in attention deficit hyperactivity disorder (ADHD) and autism spectrum disorders (ASD) including anxiety, depression, impulse control, and mood disorders. This poses a challenge regarding treatment regimens addressing all co-occurring disorders as stimulant drugs, the first-line ADHD therapy, often worsen comorbid symptoms. Several studies have linked metabotropic glutamate receptors (mGluR) with psychiatric disorders. We previously demonstrated that copy-number variants (CNVs) in the mGluR network of 273 genes were associated with both ADHD and ASD (Elia *et al*, 2011; Hadley *et al*, 2014). This led to the study of fasoracetam (NFC-1), a non-stimulating mGluR modulator, demonstrating its effectiveness at treating symptoms of ADHD and its comorbid traits in individuals with CNVs in the mGluR genes (Elia *et al*, 2018). In an effort to tie together mGluR gene association with several psychiatric disorders, we evaluated the comorbidities of these disorders in individuals with both ADHD and mGluR CNVs, as these individuals may benefit the most from mGluR modulating drugs. From the Center of Applied Genomics' pediatric biobank, over 50,000 individuals were genotyped on SNP-arrays that were amenable to CNV analysis. Phenotypic information was obtained from electronic medical records with manual record validation. Variant calls were analyzed using ParseCNV. As part of our initial review of findings, we evaluated CNV calls in mGluR genes with previous ADHD associations and observed response to NFC-1: *GRM1*, *GRM5*, *GRM7*, *GRM8* and the GRM interactor gene, *CNTN4*. When compared to individuals with no CNVs in the 273 mGluR network gene set, we uncovered significant ( $p < 0.001$ ) enrichment of these 5 mGluR genes for ADHD, aggressive behavior (including conduct disorder) and mood disorders. Further, in individuals with either ADHD or ASD, greater comorbidity with aggression was observed in the presence of *CNTN4* CNVs ( $p = 0.014$ ). This study extends our previously published results demonstrating that rare recurrent CNVs in mGluR genes that associate with ADHD and autism are also associated with their comorbid traits. Studies to determine the optimal patient cohort responsive to NFC-1 treatment are in the planning, including a basket study to explore the potential benefit on ADHD, autism, anxiety, mood and depression both individually and in ADHD patients with these comorbid symptoms where there is significant unmet medical need.

## PgmNr 2213: Nightmares: Genetic variants and risk of psychiatric disorders.

### Authors:

H.M. Ollila <sup>1,2,3,4</sup>; N. Sinnott-Armstrong <sup>3</sup>; K. Kantojarvi <sup>5</sup>; T. Palviainen <sup>4</sup>; C. Hublin <sup>6</sup>; K. Valli <sup>7</sup>; N. Sandman <sup>7</sup>; E. Kronholm <sup>7</sup>; K. Kaprio <sup>4,5</sup>; T. Paunio <sup>5</sup>; R. Saxena <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center of Genomic Medicine, Massachusetts General Hospital, Boston, MA.; 2) Broad Institute, Cambridge, MA; 3) Stanford University, School of Medicine, Palo Alto, CA; 4) Finnish Institute for Molecular Medicine Helsinki, Finland; 5) University of Helsinki, Helsinki, Finland; 6) Institute for Occupational Health, Finland; 7) University of Turku, Finland

---

Nightmares are dreams that can inflict intense fear, sadness or anxiety. However, why we experience nightmares has remained unknown. To address this question, we examined the phenotypic and genetic underpinnings of nightmares in 28,596 individuals from Finland and from the United States. Here we show that nightmares associate strongly with sleep and psychiatric traits at epidemiological level; awakenings during the night ( $p=1.93e-11$ ), depression ( $p<2e-16$ ) and alcohol use ( $p=6.32e-12$ ). Similarly, we describe a strong genetic correlation between the frequency of nightmares and personality, psychiatric disorders and sleep-related traits; neuroticism ( $rg=0.59$ ,  $p=8e-7$ ), major depressive disorder ( $rg=0.68$ ,  $p=7e-4$ ), schizophrenia ( $rg=0.23$ ,  $p=0.02$ ), insomnia ( $rg=0.50$ ,  $p=1.87e-5$ ) and tiredness ( $rg=0.62$ ,  $p=1e-6$ ). Analysis of directionality using mendelian randomization showed a significant effect from psychiatric traits to nightmares ( $p<0.01$ ) with no evidence of pleiotropy and no evidence of nightmares predisposing to psychiatric problems. Finally, we identified individual genetic variants that predispose to nightmares near *MYOF* ( $rs701873$ ,  $p=2.18e-8$ ) and *PTPRJ* ( $rs11039471$ ,  $p=3.7e-8$ ) - the latter representing a locus previously associated with sleep duration. These findings are the first evidence showing individual genetic associations with nightmares. Furthermore, while nightmares are caused by unique genetic risk factors, a substantial effect in nightmares is conveyed through underlying psychiatric traits.



## PgmNr 2214: Enhancing access to data at the NIMH Repository and Genomics Resource.

### Authors:

B.G. Ruocco <sup>1</sup>; R. Mayani <sup>2</sup>; S. Sharma <sup>2</sup>; S. Wilson <sup>1</sup>; K. Vahi <sup>2</sup>; S. Voinea <sup>2</sup>; G. Davis <sup>1</sup>; J. Valentine-Cooper <sup>3</sup>; W. Valentine-Cooper <sup>3</sup>; J. Mathew <sup>2</sup>; M. Romanella <sup>1</sup>; Y. Arens <sup>2</sup>; E. Deelman <sup>2</sup>; M.A. Azaro <sup>1</sup>; V.J. Vieland <sup>3</sup>; J.L. Ambite <sup>2</sup>; L.M. Brzustowicz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, Rutgers University, Piscataway, NJ; 2) Information Sciences Institute, University of Southern California, Los Angeles, CA; 3) Battelle Center for Mathematical Medicine, The Research Institute at Nationwide Children's Hospital, Columbus, OH

---

The NIMH Center for Collaborative Genomic Studies on Mental Disorders, now known as the NIMH Repository and Genomics Resource (NRGR), was established through the NIMH Human Genetics Initiative in 1998 to leverage and increase the value of human genetic samples and data produced through NIMH funded research. NRGR plays a key role in facilitating psychiatric genetic research by providing collections of well characterized, high-quality patient and control samples from a wide-range of mental disorders. The NRGR maintains clinical data, genetic data, and biomaterial samples from over 200,000 well characterized individuals with a range of psychiatric illnesses, their family members, and unaffected controls. Clinical data include overall diagnoses and, in many cases, item-level data from assessment instruments. Biomaterials include DNA on virtually all subjects and lymphoblastoid cells lines or cryopreserved lymphocytes on most subjects; plasma, RNA, fibroblasts, and iPSCs are available on subsets of the collection. Genome-wide SNP or sequence data is available for many NRGR subjects. For many subjects, additional data is available through dbGaP and NDA, collaborating data repositories. Given the size of the NRGR collection and the diversity of data available, new tools were required to facilitate research using this resource. To this end, the NRGR recently partnered with a user experience design company and developed an enhanced website that streamlines access to data, provides more powerful search features, and improves the overall user experience. The new website was developed in concert with ongoing processes of enhanced curation of the underlying data to improve uniformity and harmonization across individual studies, including harmonization of diagnoses for comparison across studies and collections, and improved linkage of subject IDs at collaborating repositories. Highlights of the improvements to the website and data will be presented to assist interested researchers in using this resource, including a new query interface for identifying individuals satisfying desired demographic, phenotypic, genetic, and biosample characteristics, and a new web-based interactive data submission pipeline with numerous quality control checks to enhance data curation. Please visit the improved website at [www.nimhgenetics.org](http://www.nimhgenetics.org) to explore our available data and resources, and contact [help@nrgr.on.spiceworks.com](mailto:help@nrgr.on.spiceworks.com) with feedback and/or questions.

## PgmNr 2215: Association between maternal depression symptoms and child telomere length: Evidence from growing up in New Zealand.

### Authors:

C. Walker <sup>1</sup>; Z. Thayer <sup>2</sup>; K. Ly <sup>1</sup>; A. Pillai <sup>1</sup>; E. Marks <sup>1</sup>; K. Waldie <sup>1,4</sup>; R. Snell <sup>3</sup>; S. Berry <sup>5</sup>; S. Morton <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Longitudinal Research, School of Population Health, University of Auckland, Auckland, New Zealand; 2) Department of Anthropology, Dartmouth College, Hanover, New Hampshire; 3) School of Biological Sciences, Centre For Brain Research University of Auckland, Auckland, New Zealand; 4) School of Psychology, University of Auckland, Auckland, New Zealand; 5) Auckland War Memorial Museum, Auckland, New Zealand

---

The biological mechanisms that explain how early life events influence adult disease risk are poorly understood. One proposed mechanism is via accelerated biological ageing, for which telomere length is understood to be a useful and valid biomarker. We aimed to determine if maternal depression pre and post-partum was associated with telomere length in children at 4 years of age (n~ 4,000). Mothers completed questionnaires during pregnancy, at nine months and at 54 months that included an assessment of their depression symptoms. Linear regression was used to investigate the relationship between telomere length and maternal depression at each time-point. Child telomere length was longer if their mothers depression score was higher at nine months of age (coefficient 0.003, SE 0.001, P=0.01) or 54 months of age ((coefficient 0.003, SE 0.002, P=0.02). Although these findings seem paradoxical, it suggests that early life stressors may result in adaptive responses in the child, leading to increased telomere length. We propose several testable hypotheses to explain these results and suggest further research is required to determine if the positive association between depression and telomere length is a developmental adaptation or an indirect consequence of environmental factors.

## PgmNr 2216: Integrative analysis of omics summary data prioritize susceptibility genes underlying alcohol consumption.

### Authors:

H. Zhang; S.P. Chou

[View Session](#) [Add to Schedule](#)

**Affiliation:** NIH/NIAAA, Rockville, Maryland.

---

Access and use of alcohol are associated with risk of abuse and alcohol use disorder, both are prevalent and serious mental health conditions. A genome-wide association study (GWAS) of up to 1.2 million individuals discovered hundreds of genetic variants for alcohol consumption (Nat Genet. 2019;51(2):237-244). In order to prioritize genes identified in this study for the following functional studies, we used summary-data-based Mendelian randomization method (SMR) to re-analyze the summary GWAS results underlying alcohol consumption. Among the genetic variants (SNPs) reaching the genome-wide significant threshold ( $P < 5 \times 10^{-8}$ ), we discovered 12 genomic loci (*FUT2*, *GCKR*, *IZUMO1*, *KANSL1*, *MAPT*, *NSF*, *PLEKHM1*, *RASIP1*, *SPPL2C*, *STH*, and two intergenic loci (rs1260333, rs4788084)) pleiotropically associated with both alcohol consumption (volume of alcohol consumption per week) and Methylation QTL(mQTL) in brain tissue (SMR P value  $< 1 \times 10^{-10}$ ). In addition, there are 11 genomic loci (*CRHR1*, *FUT2*, *KLB*, *NSF*, *PLEKHM1*, *RASIP1*, *RFC1*, *SH2B1*, *SLC39A8*, intergenic rs2726033, and intergenic rs4788084) discovered in the same method using mQTL expression data in blood lymphocytes and 1 gene (*CRHR1*) using eQTL expression data in peripheral blood tissue (SMR P value  $< 1 \times 10^{-10}$ ). The results indicated that the effects of these susceptibility genes on alcohol consumption are mediated by the expression level of the corresponding gene. It also suggested that the Mendelian randomization method using summary GWAS is effective and efficient in identifying and prioritizing candidate loci for the downstream functional studies.

## PgmNr 2217: A large schizophrenia family with a highly penetrant, rare, non-synonymous functional variant in the NRXN1 interactant CASKIN1.

### Authors:

M. Wahbeh<sup>1</sup>; A. Hatzimanolis<sup>2</sup>; S. Bacharaki<sup>3</sup>; S. Dimitrakopoulos<sup>2</sup>; R. Wang<sup>1</sup>; X. Yang<sup>1</sup>; K. Perzel Mandell<sup>1</sup>; N. Stefanis<sup>2</sup>; D. Avramopoulos<sup>1</sup>; Baylor-Hopkins Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.; 2) Department of Psychiatry, School of Medicine, National and Kapodistrian University of Athens, Eginition Hospital, Athens, Greece.; 3) Department of Psychiatry, General Hospital of Syros, Cyclades, Greece.

---

Schizophrenia is a highly heritable (estimated at 60-80%), common, and severe psychiatric disorder that affects ~1% of the population worldwide. The first episode of psychosis typically occurs in early adulthood, and the course of disease varies among individuals. In most cases, disease risk is conferred by many common genetic variants with additive effects, where each variant makes small contributions to the risk. However, rare variants of high penetrance also contribute to the disease prevalence. These variants include both copy number variants (CNVs) and single nucleotide variants (SNVs). We have identified a 3 generational Greek family where schizophrenia (SZ) appears to be inherited in an autosomal dominant fashion, suggesting the presence of a causative rare variant of high penetrance. The pedigree includes 4 individuals with SZ, one with schizoaffective disorder and one with bipolar disorder as well as 8 unaffected individuals (for which we have DNA). The presence of two obligate carriers in the family suggests high but incomplete penetrance. We sequenced the exome of the 4 affected individuals and found only two rare, non-synonymous variants (<1% in dbSNP, ExAc and gnomAD databases) that were present in all and are in genes that are expressed in the brain. Both variants were bioinformatically predicted to be damaging and were completely absent from all variation databases. Genotyping of other individuals in the family by Sanger sequencing showed that one of the variants (D1204N in CASKIN1) was segregating with psychosis and linkage analysis gave the highest possible LOD score of 2.2 for the family at a penetrance of 0.75. D1204N has high intolerance scores and is in the Proline rich region of the protein CASKIN1. CASKIN1 interacts directly with NRXN1 at the synapse through tripartite complexes. Neurexins interact across the synapse with neuroligins promoting synapse formation. The *NRXN1* gene encoding neurexin 1 in particular is disrupted by the only known single-gene SZ CNV, making CASKIN1 an excellent candidate as a causative highly penetrant variant in this family. We are currently introducing this variant into induced pluripotent stem cells using CRISPR Cas9 to identify phenotypic and transcriptomic changes after differentiation into neuronal cells. Identifying and characterizing a high penetrance SZ variant in the neurexin-neuroligin signaling at the synapse will improve our understanding of the role of the synapse in SZ development.

## PgmNr 2218: Transcriptome analysis to identify co-expressed gene networks as a molecular signature for childhood trauma-related mood disorders.

### Authors:

A. Delahaye-Duriez <sup>1,2,3</sup>; A. Mokhtari <sup>1</sup>; B. Etain <sup>4</sup>; I. Yalcin <sup>5</sup>; C. Marie-Claire <sup>4</sup>; E.C. Ibrahim <sup>6</sup>; R. Belzeaux <sup>6</sup>; P.-E. Lutz <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NeuroDiderot, INSERM UMR1141, Universite de Paris, Paris, France; 2) UFR SMBH, Universite Paris 13, Bobigny, France; 3) Service Histologie-Embryologie-Cytogenetique-Biologie de la reproduction-CECOS, Hopital Jean Verdier, AP-HP, Bondy, France; 4) INSERM UMR-S1144, Universite de Paris, Paris, France; 5) Institut des Neurosciences Cellulaires et Integratives UPR3212, CNRS, Universite de Strasbourg, Federation de Medecine translationnelle de Strasbourg, Strasbourg, France; 6) Universite Aix-Marseille, CNRS, Institut de Neurosciences de la Timone, Marseille, France

---

Bipolar Disorder (BD) and Major Depressive Disorder (MDD) are common and severe psychiatric diseases which can be devastating for the patients, resulting in higher risks of suicide, drug abuse, and shorter life expectancy. Despite the heterogeneous etiology of these illnesses, ranging from genetic predisposition to environmental factors, recent studies showed significant associations between childhood trauma (CT), the severity of symptoms and early onset of BD and MDD. Furthermore, patients who experienced adversity during childhood have been found less responsive to Lithium, the standard mood-stabilizer. Acknowledging that the prognosis of both BD and MDD strongly depends on thymic relapses, understanding their risk factors and underlying mechanisms is, hence, a major challenge to target patients requiring more intensive care.

In this study, we aim to identify a transcriptomic signature of CT in both BD and MDD patients whose experiences during youth have been carefully assessed using the standardized childhood trauma questionnaire (CTQ). First, at the gene level, we conducted a differential expression analysis on: 1) RNA-seq data of lymphoblastoid immortalized cell lines generated from 37 BD patients and 20 healthy controls; and 2) RNA-seq data of peripheral blood leukocytes, generated from 30 MDD patients and 34 controls. Then, at a system level, we performed a weighted gene correlation network analysis (WGCNA), to detect specific or common modules and hub genes amongst both disorders. The expression of all genes within each module was summarized as eigengene to test association between module expression and CT.

Our preliminary results indicate interesting enrichments, with up-regulated genes involved in immune response in both BD and MDD, and down-regulated genes implicated in neuronal development among BD patients. Further characterization of genes whose expression is modified by the exposure to CT might provide a more comprehensive pathophysiological pathway leading from CT exposure to a higher severity in BD and MDD.

## PgmNr 2219: RNA sequencing reveals transcriptomic changes in individuals with insomnia.

### Authors:

S. Mithani <sup>1,2</sup>; H. Kim <sup>1</sup>; S. Yun <sup>3</sup>; P. Gerhman <sup>4</sup>; A. Fink <sup>2</sup>; J. Gill <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Institutes of Health, Rockville, Maryland.; 2) University of Illinois, College of Nursing, Chicago, Illinois 60612; 3) Yotta Biomed, LLC., 8908 Ewing Drive, Bethesda, MD 20817; 4) University of Pennsylvania, Department of Psychiatry, Pennsylvania, PA 19104

---

**Purpose:** Insomnia affects 10-20% of the population and is associated with negative health and psycho-social sequelae. In addition, disruption in sleep homeostatic processes results in higher stress vulnerability. In this study, we examined differentially expressed genes in 15 individuals with chronic insomnia and age- and sex-matched good sleepers (n=15).

**Methods:** We performed RNA-sequencing on total of 150 whole blood samples (five samples per participant) by using Illumina's NovaSeq-6000 sequencer. We used STAR for alignment to hg38 reference genome. Differential gene expression analysis was performed using DESeq2.

**Results:** Based on the RNA-sequencing of samples withdrawn at time point 1 (0 hours) and timepoint 5 (42-44 hours), insomnia patients showed 2516 dysregulated genes compared to controls. Those dysregulated genes include CSMD1, IMPA1, and MIR1244-2. Particularly, CSMD1 gene is 4.4 fold down-regulated in insomnia patients, which is also reported to be related to Schizophrenia.

**Conclusion:** This first study of understanding transcriptome changes in individuals who suffer from insomnia, may provide future direction for understanding pathways and identifying targets for drug and intervention studies.

## PgmNr 2220: Genome-wide association study in multiplex consanguineous Pakistani pedigrees with schizophrenia and bipolar disorder.

### Authors:

L. Xiong<sup>1,2,3</sup>; J. John<sup>1,2,3</sup>; A.M. Johnson<sup>3</sup>; Q. He<sup>3</sup>; M. Christian<sup>3</sup>; L.E. Delisi<sup>4</sup>; R. Joobar<sup>2</sup>; M.-P.G.A. Rouleau<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology and Neurosurgery and Montreal Neurological Institute and Hospital, McGill University, Montreal, QC, H3A 2B4, Canada; 2) Department of Psychiatry and Douglas Mental Health University Institute, McGill University, Verdun, QC H4H 1R3, Canada; 3) Centre de Recherche, Institut Universitaire en Santé Mentale de Montréal et Université de Montréal, Montréal, QC, H1N 3M5, Canada; 4) VA Boston Healthcare System, Department of Psychiatry, Harvard Medical School, Brockton, MA, 02301, USA

---

**Introduction:** Genome wide association studies (GWASs) in additional diverse populations are essential to test the generalizability of the previous GWAS findings, to extend our understanding of the disease etiologies, and to identify the population-specific genetic risk in addition to common factors. Particularly no such studies have been performed in South Asian populations, one of the largest parts of the world population with distinctive population history and genomic characteristics.

**Samples and Methods:** Ten large consanguineous pedigrees (124 affected with SCZ or BPD and 151 unaffected individuals) and 34 unrelated healthy individuals were recruited from Sindh Province of Pakistan. Genotyping was performed using Illumina HumanOmniExpress BeadChip. After the standard QC and genotype phasing and imputation, we performed family-based GWAS using DFAM implemented in PLINK1.07 and FUMA for post-GWAS annotations.

**Results:** We've identified one locus (chr2:11,237,350, rs541513649;  $P$  value  $5.52e^{-08}$ ) associated with the broad psychiatric phenotype and 14 other independent loci showing suggestive statistical evidence of association ( $P$  value  $< 10^{-5}$ ). Out of 15 loci, 10 loci also showed nominal association ( $P$  value  $< 1.00e^{-3}$ - $2.31e^{-11}$ ) in the previously reported psychiatry and/or related GWASs, e.g., a locus on chr7: 68,977,045-69,825,163, (rs12698811,  $P$  value  $8.75e^{-07}$ ) was previously reported associated with cognition (rs12112638,  $P$  value  $2.31e^{-11}$ ; Lee et al., 2018) and intelligence (rs12698891,  $P$  value  $1.22e^{-09}$ , Savage et al., 2018). Functional annotation of these loci using FUMA have also identified a number of interesting disease relevant genes, including *KCNF1*, *ROCK2*, *GABRA4*, *GABRB2*, *AUTS2*, *IGF1* etc.; which showed supporting evidence from the previous genetics and functional studies.



## PgmNr 2221: Structural variant in the RNA binding motif protein, X-linked 2 (*RBMX2*) gene found to be linked to bipolar disorder.

### Authors:

J.M. Ekholm <sup>2</sup>; W.J. Rowell <sup>2</sup>; L. Paulin <sup>1</sup>; S. Kujawa <sup>2</sup>; D. Ratterman <sup>3</sup>; G. Mayhew <sup>3</sup>; J. Wendt <sup>3</sup>; D.L. Burgess <sup>3</sup>; T. Partonen <sup>4</sup>; T. Paunio <sup>4,5</sup>; P. Laine <sup>1</sup>; P. Auvinen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Biotechnology, University of Helsinki, Helsinki, Finland; 2) Pacific Biosciences PacBio, Menlo Park, CA , USA; 3) Roche NimbleGen, Inc, Madison, WI 53719, USA; 4) National Institute for Health and Welfare (THL), Department of Health, Helsinki, Finland; 5) Department of Psychiatry, University of Helsinki

Bipolar disorder (BD) is a phenotypically and genetically complex neurological disorder that affects 1% of the worldwide population. There is compelling evidence from family, twin and adoption studies supporting the involvement of a genetic predisposition with estimated heritability up to ~ 80%. The risk in first-degree relatives is ten times higher than in the general population. Linkage and association studies have implicated multiple putative chromosomal loci for BD susceptibility, however no disease genes have yet to be identified.

Here, we have fully characterized a ~12 Mb significantly linked (lod score=3.54) genomic region on chromosome Xq24-q27 in an extended family from a genetic isolate that was using long-read single molecule, real-time (SMRT) sequencing. The family segregates BD in at least 4 generations with 16 individuals out of 61 affected. Thus, this family portrays a highly elevated reoccurrence risk compared to the general population. It is expected that the genetic complexity would be reduced in isolated populations, even in genetically complex disorders such as BD, as in the case of this extended family.

We selected 16 key individuals from the X-chromosomally linked family to be sequenced. These selected individuals either carried the disease haplotype, were non-carriers of the disease haplotype, or served as married-in controls. We designed a Nimblegen capture array enriching for 5-9 kb fragments spanning the entire 12 Mb region that were then sequenced using long-read SMRT sequencing to screen for causative structural variants (SVs) explaining the increased risk for BD in this extended family. Altogether, 192 SVs were detected in the critically linked region however most of these represented common variants that could be seen across many of the family members regardless of the disease status. One SV stood out that showed perfect segregation among all affected individuals that were carriers of the disease haplotype. This was a 330bp Alu deletion in intron 4 of the RNA Binding Motif Protein, X-Linked 2 (*RBMX2*) gene that has previously been shown to play a central role in brain development and function. Moreover, Alu elements in general have also previously been associated with at least 37 neurological and neurodegenerative disorders. In order to validate the finding and the functionality of the identified SV further studies like isoform characterization are warranted.



## PgmNr 2222: Sex differences and tissue-specific gene expression across psychiatric disorders.

### Authors:

I.O. Fedko <sup>1</sup>; A. Abdellaoui <sup>2</sup>; K.J.H. Verweij <sup>2</sup>; M.G. Nivard <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands;

2) Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

---

Prevalences of psychiatric disorders differ between males and females; for example autism spectrum disorders and substance use are more prevalent in males, whereas anxiety and depression are more prevalent in females. Twin and family studies showed that vulnerability to psychiatric disorders are partially explained by genetic factors. Genetic correlations between sexes are generally very high ( $>.9$ ), which indicates that the genetic factors are largely the same for males and females.

Differences in gene expression are a plausible mechanism underlying sex-differences in psychiatric disorders. In this project we will investigate sex differences in a large sample of gene expression data. By mining Gene Expression Omnibus (GEO) data repository we will extract two Affymetrix chip human gene expression data (GPL96 and GPL570). We will re-analyze these pooled data to extract sex specific patterns of gene expression, separately for brain ( $N \approx 300$ ) and blood ( $N \approx 2000$ ) samples. We will focus on autosomes as well as on the X chromosome. We will then integrate our findings with summary statistics obtained in large genome-wide association studies. This will allow us to facilitate our understanding of mechanisms that underlie sex differences in psychiatric disorders.

## PgmNr 2223: Discovery of novel genetic risk loci for acute central serous chorioretinopathy and genetic pleiotropic effect with age-related macular degeneration.

### Authors:

Y. Ren <sup>1</sup>; L. Feng <sup>2</sup>; S. Chen <sup>3</sup>; D. Xue <sup>4</sup>; R. Dorajoo <sup>5</sup>; J. Liu <sup>5,6</sup>; X. Yin <sup>7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of dermatology, The Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, China; 2) Eye Center, the Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou, Zhejiang, China; 3) Department of ophthalmology, Longsai Hospital, Ningbo Zhejiang, China; 4) Department of plastic surgery, the Second Affiliated Hospital of Zhejiang University School of Medicine, Hangzhou Zhejiang, China; 5) Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore; 6) Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.; 7) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA

---

Central serous chorioretinopathy (CSC) is a severe and heterogeneous chorioretinal disorder and shares certain clinical manifestations with age-related macular degeneration (AMD), suggesting common pathophysiologic mechanisms between two diseases. To advance the understanding of genetic susceptibility of CSC and further investigate genetic pleiotropy between CSC and AMD, we performed genetic association analysis of 38 AMD-associated single nucleotide polymorphisms (SNPs) in a Chinese CSC cohort, consisting 464 patients and 548 matched healthy controls. Twelve SNPs were found to be associated with CSC at nominal significance ( $p < 0.05$ ), and four SNPs on chromosomes 1 (rs1410996 and rs1329428), 4 (rs4698775), and 15 (rs2043085) showed strong associations whose evidences surpassed Bonferroni-corrected significance. While the genetic risk effects of rs1410996 and rs1329428 (within the well-established locus *CFH*) are correlated (due to high LD), rs4698775 on Chr4 and rs2043085 on Chr15 are novel risk loci for CSC. Three independent SNPs were also found to be associated with clinic manifestations of CSC patients. PRS constructed by using three SNPs showed highly significant association with CSC ( $p = 2.10 \times 10^{-7}$ ), with the top 10% of subjects with high PRS showing 6.39 times higher risk than the bottom 10% of subjects with lowest PRS. In addition, the comparison of the genetic effects (ORs) of these 38 SNPs revealed significant, but complex genetic pleiotropic effect between CSC and AMD. By discovering two novel genetic risk loci and revealing significant genetic pleiotropic effect between CSC and AMD, the current study has greatly advanced the biological understanding of CSC.

## PgmNr 2224: Resequencing of candidate genes for Keratoconus reveals a role for Ehlers-Danlos syndrome genes.

### Authors:

E. Fransen <sup>1,2</sup>; H. Valgaeren <sup>1</sup>; M. Beyens <sup>1</sup>; M. Sommen <sup>1</sup>; G. Vandeweyer <sup>1</sup>; L. Bisceglia <sup>3</sup>; V. Soler <sup>4</sup>; A. Hoischen <sup>5,6</sup>; G. Mortier <sup>1</sup>; F. Malecaze <sup>4</sup>; C. Koppen <sup>7,8</sup>; G. Van Camp <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) StatUa Center for Statistics, University of Antwerp, Belgium; 3) IRCCS Casa Sollievo della Sofferenza, UOC Genetica Medica, San Giovanni Rotondo, Italy; 4) Department of Ophthalmology, Pierre-Paul Riquet Hospital, Toulouse University Hospital, Toulouse, France; 5) Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Geert Grooteplein 10, 6525 GA Nijmegen, the Netherlands; 6) Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, 6500 HB Nijmegen, the Netherlands; 7) Department of Ophthalmology, Antwerp University Hospital, Wilrijkstraat 10, 2650 Edegem (Antwerp), Belgium; 8) Department Ophthalmology, Visual optics and Visual rehabilitation, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium

---

The involvement of genetic factors in the pathogenesis of KC has long been recognized but the identification of pathogenic variants has been challenging. In this study, we selected 34 candidate genes for KC based on previous Whole Exome Sequencing (WES) and literature, an resequenced them in 745 KC patients and 810 ethnically matched controls. Data analysis was performed using a single-SNP association test on the variant level, as well as gene-based mutation burden and variance components tests.

In our study, we detected enrichment of genetic variation across multiple gene-based tests for the genes *COL2A1*, *COL5A1*, *TNXB*, *ZEB1* and *ZNF469*. The top hit in the single-SNP association test was obtained for a common variant in the *COL12A1* gene. Interestingly, *COL5A1*, *TNXB*, *ZNF469* and *COL12A1* are all known Ehlers-Danlos Syndrome (EDS) genes. Though the co-occurrence of KC and EDS has been reported previously, this study is the first to demonstrate a consistent role of genetic variants in EDS genes in the etiology of KC. In conclusion, our data show a shared genetic etiology between KC and EDS, and clearly confirm the currently disputed role of *ZNF469* in disease susceptibility for KC. Besides the role for EDS genes, we also confirm the reported previously involvement of genetic variation in *ZEB1* in KC

## PgmNr 2225: Whole genome sequencing study identifies novel variants associated with intrinsic circadian period in humans.

### Authors:

C. Polymeropoulos; S. Smieszek; D. Xiao; G. Birznieks; M. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals, Washington DC, District of Columbia.

---

Non-24 is a circadian rhythm disorder in which the master body clock runs either slightly earlier or, more commonly in the disorder, longer than 24 hours. We conducted the first whole-genome sequencing study of a non-24 population of 174 individuals that we identified as being totally blind with Non-24 Disorder. We have directly tested the association between SNPs and circadian period length ( $\tau$ ) as calculated from the measurements of urinary 6-sulphatoxymelatonin (aMT6s) rhythms ( $n=69$ ). Linear regression corrected for PCs and covariates identified a strong signal in *HCN1*, Brain Cyclic Nucleotide-Gated Channel 1, *HCN1*. *HCN1* channel is responsible for the feedback on the rods regulating the dynamic range of light reactivity under dim or intermediate light conditions. Minor allele rs72762058 associates with longer  $\tau$ , a difference of 12 minutes, and mean  $\tau$  of 24.71. In *Drosophila* there is only one HCN channel encoding gene, *Dm1h*. Interestingly, *Dm1h* mutant flies display alterations in the rest: activity pattern, and altered circadian rhythms, specifically, arrhythmic behavior or a shorter period in constant darkness. Hyperpolarization-activated cyclic nucleotide-gated channels control neuronal excitability and their dysfunction has been previously linked to epileptogenesis.

In addition, we identify others variants that strongly associate with  $\tau$ , such as a missense variant (rs16989535), (minor allele associated with longer  $\tau$ ), within *DEPDC5*, GATOR Complex Protein; found in Familial focal and Rolandic Epilepsy). Subjects carrying the rare allele have a period  $> 25.2$ . *DEPDC5* is part of GATOR1 complex, together with NPRL2 and NPRL3 acts to inhibit the mTORC1 pathway. The mTOR signaling is part of the photic entrainment pathway in the SCN, it regulates autonomous clock properties in a variety of circadian oscillators and regulates network properties of coupled circadian oscillators, such as the SCN neurons. Light-induced mTORC1 activation appears to be important for photic entrainment of the SCN clock.

Our study illustrates the role of *HCN1* in circadian clock biology in humans as shown by an effect on  $\tau$ . We identify variants in *DEPDC5*, previously implicated in epilepsy, as exhibiting significantly longer  $\tau$ . Genetic determinants of period length are not only essential for the understanding of the basic clockwork mechanisms but could also provide insights into mechanistic links between circadian dysfunctions and human diseases such as epilepsy.

## PgmNr 2226: Expression and alternative splicing QTLs integrated with GWAS reveal novel genetic associations and causal genes for glaucoma.

### Authors:

A.R. Hamel<sup>1,2,3</sup>; J.M. Rouhana<sup>1,2,3</sup>; F. Aguet<sup>3</sup>; A.P. Khawaja<sup>4,5</sup>; M. Simcoe<sup>6,7</sup>; P.G. Hysi<sup>6,7</sup>; C.J. Hammond<sup>6</sup>; H. Kyung Im<sup>8</sup>; K.G. Ardlie<sup>3</sup>; J.L. Wiggs<sup>1,2,3</sup>; A.V. Segre<sup>1,2,3</sup>; GTEx Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ocular Genomics Institute and Department of Ophthalmology, Massachusetts Eye and Ear, Boston, MA; 2) Department of Ophthalmology, Harvard Medical School, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge, MA; 4) NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; 5) Department of Public Health and Primary Care, Institute of Public Health, University of Cambridge School of Clinical Medicine, Cambridge, UK; 6) Department of Ophthalmology, King's College London, St. Thomas' Hospital, London, UK; 7) Department of Twin Research & Genetic Epidemiology, King's College London, St. Thomas' Hospital, London, UK; 8) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, Illinois, USA

---

Genome-wide association studies (GWAS) have detected 24 genomic loci associated with Primary Open Angle Glaucoma (POAG) and 112 loci associated with intraocular pressure (IOP), a major risk factor for POAG. Yet, the implicated causal genes for these associations, explaining 5-17% of the traits' heritability, are largely unknown, as most associated variants lie in noncoding regions and tag multigenic regions. We thus tested whether *cis*-acting expression quantitative trait loci (eQTLs) or splicing QTLs (sQTLs) in 49 tissues from the Genotype-Tissue Expression (GTEx) Project (release v8) could prioritize causal genes in known GWAS loci and detect new genetic associations with POAG and IOP.

We found that 70% of POAG and IOP associated-variants were in LD with  $\geq 1$  eQTL in any of the tissues. Applying colocalization analysis to the co-occurring GWAS and eQTL signals, considering allelic heterogeneity, proposed one to several high confidence causal genes per GWAS locus. One example is *ANGPT2*, whose increased expression specifically in artery is associated with decrease in IOP (eCAVIAR Posterior Prob.=1). We next tested whether genome-wide significant to subthreshold ( $p < 0.05$ ) POAG or IOP associations were enriched among eQTLs and sQTLs using *eQTLEnrich.v2* that corrects for MAF, distance to TSS, and LD. Using the NEIGHBORHOOD POAG GWAS meta-analysis and UK Biobank IOP GWAS, we found significant enrichment of POAG and IOP associations among eQTLs and sQTLs ( $P < 2.5E-04$ ) in relevant tissues: artery, a potential proxy for Schlemm's canal involved in fluid removal from the eye; brain, a potential proxy for retinal ganglion cells; and adipose, suggesting lipid-related processes. While the adjusted fold-enrichment was higher for sQTLs (1.2-1.8) than eQTLs (1.16-1.6), a larger number of eQTLs was estimated to underlie POAG and IOP associations than sQTLs (170-1300 eQTLs and 90-800 sQTL per tissue). Top POAG-ranked eQTLs and sQTLs replicated in an independent POAG GWAS (UK Biobank; adj. fold-enrich: 2-2.5). Using our gene set enrichment method *eGeneEnrich.v2* that adjusts for expression levels in a given tissue, we found that target genes of top ranked eQTLs were enriched in specific biological processes, e.g. mitochondria organization,

morphology of retinal vasculature and anterior chamber of the eye, and regulation of neuron death.

This work suggests an important role for regulatory mechanisms in POAG risk and proposes new causal genes and pathogenic processes for glaucoma.

## PgmNr 2227: Whole genome sequencing study identifies a novel region in HCN1 affecting human chronotypes, circadian period length and sleep consolidation.

### Authors:

S. Smieszek; D. Xiao; C. Polymeropoulos; G. Birznieks; M. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals

Circadian rhythms affect most aspects of human physiology including a wide range of molecular and behavioral processes. Chronotype known as morningness/eveningness preference describes one's propensity for earlier/later sleep timing and is indicative of individuals' circadian rhythms. Variation thereof is coupled with the propensity for sleeping. To ascertain the genetic risk factors for morningness/eveningness (ME) phenotypes, we conducted a large whole-genome sequencing association analysis using 316 healthy sleeping participants.

We have directly tested the association between SNPs and morning-evening questionnaire (MEQ). We detect a large region on chromosome 5 (more than ~400 adjacent SNPs in LD, spanning ~2mb) centered within HCN1, Brain Cyclic Nucleotide-Gated Channel 1. Regional enrichment yields a p-value of  $< 10^{-99}$ . It is highly expressed in the brain and potentially modulates excitability in the brain (The protein is expressed in frontal cortex and the retina.) and responding to regulation by cyclic nucleotides, playing a critical role in shaping the autonomous activity of single neurons and the periodicity of network oscillations. It has been shown previously in a double mutant mouse model that lack of HNC1 mediated feedback in rod photoreceptor cells prolongs rod responses and saturates the downstream retinal network during bright light stimulation. The risk allele (rs12188518) is effectively correlated with lower MEQ score hence evening phenotype. The OR between morning and evening is 5.78. The locus has been shown to be a significant ( $1.6e^{-09}$ ) eQTL for HCN1 in GTEX.

HCN1 channel is responsible for the feedback on the rods regulating the dynamic range of light reactivity under dim or intermediate light conditions. We hypothesize that if this feedback is not functioning properly an individual may get saturated with even dim light resulting in misperception of the light conditions resulting in a circadian delay. Furthermore, we identify an association between the HCN1 variants and intrinsic circadian period length as calculated from the measurements of urinary 6- sulphatoxymelatonin (aMT6s) rhythms in 69 individuals (for which whole genome sequencing was performed). Our work delineates a novel role of HCN1 in circadian biology, specifically its effect upon human chronotypes, tau regulation and sleep consolidation.

## PgmNr 2228: Accumulation of RB1 LOF variants in a Non-24 whole genome sequencing study.

### Authors:

G. Birznieks; C. Polymeropoulos; S. Smieszek; D. Xiao; M. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals, DC, None selected.

---

Non-24 is a circadian rhythm disorder in which the master body clock runs either slightly earlier or, more commonly in the disorder, longer than 24 hours. It is a serious rare, circadian rhythm disorder, highly prevalent in blind individuals (Lockley). We conducted the first whole-genome sequencing study of a non-24 population of 174 individuals that we identified as being totally blind with Non-24 Disorder. Intrinsic circadian period length as calculated from the measurements of urinary 6-sulphatoxymelatonin (aMT6s) rhythms. We explored the LOF landscape of the Non-24 population. We find an accumulation of rare RB1 LOF variants. Altogether we detect 11 rare LOF mutations (7 stopgains, 1 splicing, 2 frameshift, 1 non-frameshift deletion). The OR, when compared to with a control population, is 47. All of the 15 people have a confirmed history of retinoblastoma diagnosis. The average tau is 24.4. In addition we report a single case with RB1:NM\_000321:exon1:c.110delG:p.S37fs and a *CRY1* c.1657+3A>C allele. For that one subject retinoblastoma resulted in bilateral enucleation and subsequently Non-24. Furthermore, we detect the accumulation of other variants in *MYOC*, causative of juvenile-onset open angle glaucoma, *GPER*, estrogen receptors expressed in the suprachiasmatic nucleus, which regulates circadian rhythm, *CYP11B1*, a risk factor for Glaucoma 3, and *ARMS2*, pathogenic in macular degeneration. This report underscores the genetic characterization of patients with a history of retinoblastoma, and may assist in further understanding the association of RB1 mutations and comorbidities, increased risk for non-24 as seen in patients with germline RB1 loss of function variants. Furthermore, we delineate some of the underlying germline variants that may manifest with the non-24 disorder.



## PgmNr 2229: Identification of genetic markers associated with variable response in total sleep time in the first two-thirds of the night of tasimelteon-treated subjects.

### Authors:

J. Brzezynski; S.P. Smieszek; C. Xiao; C. Polymeropoulos; G. Birznieks; M.H. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals Inc., Washington, DC.

---

The JET8 study was a randomized, double-blind, placebo-controlled, multicenter phase III study that randomized 318 healthy subjects. Whole-genome sequencing (WGS) was performed on a population of 158 subjects treated with tasimelteon and on 158 subjects treated with placebo. Tasimelteon is a MT1 and MT2 melatonin receptor agonist.

Subjects received 20 mg of tasimelteon or placebo orally prior to an 8 hour phase advance, relative to their habitual bedtime, and were assessed on nighttime sleep parameters including total sleep time in the first two-thirds of the night ( $TST_{2/3}$ ), as measured by polysomnography (PSG). Genetic markers associated with  $TST_{2/3}$  response in tasimelteon-treated subjects were analyzed. We computed association test results by linear regression assuming additive allelic effects. We used covariates of age, gender, and the top PCs to account for residual population structure. This study design induced the circadian challenge experienced by transmeridian travelers who experience symptoms of jet lag disorder.

A significant association was identified between rs11735877, a highly regulatory variant within *CPLX1*, and  $TST_{2/3}$  ( $p=10e-8$ ). *CPLX1* encodes complexin 1, a neuronal protein, which belongs to the highly conserved complexin protein family. Presumably, *CPLX1* regulates vesicle fusogenicity by interfering with the C-terminal stability of the SNARE complex (soluble N-ethylmaleimide-sensitive factor (NSF) attachment protein receptor), which is an essential effector of the synaptic vesicle fusion process and the neurotransmitter release cascade. The locus is a significant eQTL ( $p=2.6e-10$ ) for UV Stimulated Scaffold Protein A (UVSSA). Variation in UVSSA may cause increased cellular sensitivity to UV light and cutaneous photosensitivity. Based on enrichments and pathway analysis of the top scoring variants, we have identified two related categories, circadian clock and melatonin, with variants in *LRP6*, *WNT5B*, *SEMA3C*, and *GABARAPI3*, among other contributors.

We identified a locus highly associated with response to a melatonin agonist. This can be relevant to understanding the action of melatonin agonists involved in a wide array of responses.

## PgmNr 2230: A meta-analysis of 44,039 individuals doubles the number of genetic loci associated with central corneal thickness and provides important biological insights.

### Authors:

H. Choquet<sup>1</sup>; J. Yin<sup>1</sup>; T.J. Hoffmann<sup>2,3</sup>; K.K. Thai<sup>1</sup>; M.N. Kvale<sup>2</sup>; Y. Banda<sup>2</sup>; C. Schaefer<sup>1</sup>; N. Risch<sup>1,2,3</sup>; K.S. Nair<sup>4</sup>; R.B. Melles<sup>5</sup>; E. Jorgenson<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Kaiser Permanente Northern California (KPNC), Division of Research, Oakland, CA 94612, USA; 2) Institute for Human Genetics, University of California San Francisco (UCSF), San Francisco, CA 94143, USA; 3) Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA 94158, USA; 4) Departments of Ophthalmology and Anatomy, School of Medicine, UCSF, San Francisco, CA 94143, USA; 5) KPNC, Department of Ophthalmology, Redwood City, CA 94063, USA

**Purpose:** Central corneal thickness (CCT) is associated with common vision disorders and has a strong genetic component, with heritability estimates ranging between 0.68 to 0.95. To date, 45 CCT-associated loci have been reported, explaining approximately 8.5% of CCT heritability. Here, we report a combined multiethnic meta-analysis of participants from the Genetic Epidemiology Research in Adult Health and Aging (GERA) cohort and previously reported data from the International Glaucoma Genetics Consortium (IGGC).

**Methods:** The GERA cohort is an unselected cohort of adult members of the Kaiser Permanente Northern California integrated health care delivery system, with ongoing longitudinal records from vision examinations. For this study, our GERA sample consisted of 18,129 individuals from 4 ethnic groups (79.9% non-Hispanic white, 7.5% Hispanic/Latino, 8.4% East Asian, and 4.2% African American) having CCT measured. Genome-wide genotype data were generated on Affymetrix Axiom arrays and imputed to the 1000 Genomes reference panel. Genetic association analyses with CCT were performed using linear regression adjusted for age, sex, and ancestry principal components. Associations between the identified CCT-associated loci and common vision disorders (i.e., glaucoma and keratoconus) were assessed in GERA and UK Biobank.

**Results:** We identified 74 genome-wide significant loci ( $P < 5 \times 10^{-8}$ ) for CCT, including 31 novel. The effect estimates of the 31 lead variants at novel loci were consistent across the 2 studies ( $R^2 = 0.72$ ,  $P = 3.70 \times 10^{-13}$ ), and no significant heterogeneity was observed between GERA and IGGC. While conditional analyses revealed additional 11 novel independent variants within the identified loci, gene-based analysis identified a novel gene (*CIART*) associated with CCT ( $P = 2.10 \times 10^{-6}$ ). Among the lead CCT-associated variants identified, only variant rs3740685 in *RAPSN* was associated with glaucoma in GERA after multiple testing correction ( $P = 1.9 \times 10^{-4}$ ) and nominally associated in UK Biobank ( $P = 0.0017$ ). We also confirmed the negative correlation of effect estimates between CCT and keratoconus ( $R^2 = -0.26$ ,  $P = 9.53 \times 10^{-3}$ ), and the associations between keratoconus and CCT-associated loci ( $P < 0.05$ ), including *COL4A3*, *FOXO1*, *SMAD3*, and *COL5A1* as previously reported.

**Conclusions:** This largest study conducted to date on CCT not only doubles the number of CCT-associated loci reported, but also explains up to 14.2% of the CCT heritability, and provides novel biological insights.

## **PgmNr 2231: Method for predicting disease onset risk by genetic variants associated with primary open-angle glaucoma.**

### **Authors:**

F. Mabuchi <sup>1</sup>; N. Mabuchi <sup>1</sup>; Y. Sakurada <sup>1</sup>; S. Yoneyama <sup>1</sup>; K. Kashiwagi <sup>1</sup>; Z. Yamagata <sup>2</sup>; H. Iijima <sup>1</sup>; for the Japan Glaucoma Society Omics Group (JGS-OG)

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Dept Ophthalmology, University of Yamanashi, Chuo, Yamanashi, Japan; 2) Dept Health Sciences, University of Yamanashi, Chuo, Yamanashi, Japan

---

**Purpose:** To predict the onset risk on primary open-angle glaucoma (POAG) by the additive effects of genetic variants associated with POAG.

**Methods:** Twenty genetic variants associated with POAG were genotyped for 516 Japanese patients with POAG and 246 control subjects. The total number of risk alleles of these genetic variants was calculated for each participant as a polygenic risk score (PRS), and the association between the PRS and POAG was evaluated. The ratio of POAG patients to control subjects was compared with respect to the PRS.

**Results:** The PRS ( $23.8 \pm 2.6$ , mean  $\pm$  standard deviation) in patients with POAG was significantly higher ( $P < 0.0001$ , Student t-test) than that ( $22.7 \pm 2.6$ ) in control subjects. An almost 1.2-fold increase in POAG onset risk was found for the PRS. ( $P < 0.0001$ , odds ratio: 1.17 per risk allele, 95% confidence interval: 1.10 to 1.25, logistic regression analysis adjusted for age and gender) The ratio (1 POAG patient / 7 control subjects) of POAG patients to control subjects in patients with  $PRS \leq 17$  was significantly smaller ( $P = 0.0009$ , Chi-square test) than that (516 POAG / 246 control) in all patients. In contrast, the ratio (126 POAG / 29 control) in patients with  $PRS \geq 26$  was significantly larger ( $P = 0.0008$ ) than that in all patients, and the patients with  $PRS \geq 26$  had a 2.1 times higher onset risk on POAG compared with all patients.

**Conclusion:** This evaluation method by the PRS of genetic variants associated with POAG may be clinically useful for predicting the onset risk on POAG.

## PgmNr 2232: A Scandinavian whole genome sequencing study of familial asthma.

### Authors:

S. Cameron-Christie <sup>1</sup>; G. Povysil <sup>2</sup>; G. Lassi <sup>1</sup>; J. Lindgren <sup>3</sup>; G. Belfield <sup>3</sup>; I. Dillmann <sup>3</sup>; D.B. Goldstein <sup>1,2</sup>; A. Platt <sup>1</sup>; S. Young <sup>4</sup>; S. Petrovski <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Genomics Research, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Cambridge, UK; 2) Institute for Genomic Medicine, Columbia University Medical Center, New York, New York; 3) Translational Genomics, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Cambridge, UK; 4) Precision Medicine, R&D Respiratory, Inflammation & Autoimmunity, AstraZeneca, Cambridge, UK

---

Asthma risk is influenced by numerous biological factors in conjunction with environment. Although many SNPs significantly associated with asthma have been identified in genome-wide association studies, there are relatively few reports of individual loci that contribute to a high risk, and only a handful of purported Mendelian variants. To perform an unbiased search for high-risk/Mendelian variants involved in paediatric familial asthma, we present a rare variant analysis of asthma pedigrees ascertained with multiple affected members.

The Scandinavian Asthma Genetic Study (SAGA) study is an AstraZeneca study of familial, childhood asthma, where inclusion criteria required at least two children per family to be affected. We performed initial analyses on virtual exomes from 389 unrelated probands where whole genome sequence (WGS) was available and compared them to 6,624 European control exomes.

We investigated whether any genes were case-enriched for rare, protein-damaging variants. We also applied a novel method, MegaGene, to detect signals of enrichment among ~10K predefined gene-sets, including KEGG pathways and gene ontology sets. Neither gene-based collapsing analysis nor MegaGene burden analysis detected a significant enrichment of deleterious rare variants among our familial asthma cohort of 389 unrelated probands.

We selected the underlying variants in nine top ranked genes from collapsing analysis for segregation analysis. First degree affected family members were genotyped for those variants contributing to the gene-based enrichment to determine if the rate of allele sharing was significantly greater than the expected 50%. No tested gene showed significant enrichment.

With our currently limited sample size, our unbiased survey of the exome for asthma risk did not identify individual genes or gene-sets with significant enrichment of rare, deleterious variants. We now aim to: a) study structural and rare, noncoding variation in the WGS data, b) expand the study to selected cases from other available cohorts with asthma and a family history of respiratory disease, c) apply to our WGS data existing respiratory polygenic risk scores derived from large-scale Scandinavian studies such as FinnGen, and d) expand the WGS for selected pedigrees with the greatest asthma burden or where Linkage analysis will be most informative. We will present the available updates from these extended analyses.

## PgmNr 2233: Human leukocyte antigen (HLA) region class II alleles influence asthma risk through differential gene expression and differences in protein coding regions.

### Authors:

S. Clay <sup>1</sup>; M. Pividori <sup>2</sup>; H.K. Im <sup>1,2</sup>; D.L. Nicolae <sup>1,2,3</sup>; C. Ober <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, The University of Chicago, Chicago, IL.; 2) Department of Medicine, The University of Chicago, Chicago, IL.; 3) Department of Statistics, The University of Chicago, Chicago, IL.

---

Variation in the HLA region is associated with more diseases in GWAS than any other genomic region, including strong and consistent associations with asthma. Recently, independent associations were shown for childhood-onset and adult-onset asthma with rs28407950 ( $P=1.3 \times 10^{-59}$ ) and rs17843580 ( $P=3.5 \times 10^{-50}$ ), respectively, in the UK Biobank (UKB) (Pividori, Schoettler et al. 2019; *Lancet Respir Med* 7:509). To determine the relative contributions of regulatory and coding variation in HLA genes to asthma risk, we used imputed 4-digit resolution of classical HLA alleles and asthma diagnoses from the UKB. Class II HLA alleles were tested for association in each group, revealing 15 childhood-onset and 13 adult-onset class II 4-digit alleles at the *DRB1*, *DQB1*, and *DQA1* loci associated with asthma protection or risk ( $P < 5 \times 10^{-8}$ ). We then included the genotype at the lead class II GWAS SNPs as covariates. The GWAS SNPs remained significant in all analyses ( $P \leq 2.3 \times 10^{-22}$ ). Of the HLA alleles, only *DRB1\*0401* remained significant for childhood-onset asthma and *DRB1\*0401*, *DQB1\*0302* and *DQA1\*0301*, which often occur on a single haplotype, remained significant for adult onset asthma. To assess whether haplotypes tagged by the GWAS risk/protective alleles are associated with increased or decreased expression of the HLA genes, we analyzed allele specific expression in peripheral blood leukocyte RNA-seq data from 132 Hutterites after aligning RNA-seq reads to reference sequences for each person's known HLA type and phasing the GWAS SNPs and the HLA alleles. We conducted these studies in individuals heterozygous for both the GWAS SNP and HLA alleles at each gene. The patterns of expression were similar for childhood-onset and adult-onset GWAS risk alleles, with increased expression of *DQA1* transcripts (+27%) and decreased expression of *DQB1* transcripts (-35%) on the GWAS risk haplotype compared to transcripts on the non-risk haplotype. No differences in *DRB1* expression were observed between the risk and non-risk haplotypes. Independent effects of the childhood-onset and adult-onset SNPs could not be determined due to the strong LD in the Hutterites. Taken together, these data suggest that associations of class II alleles with both childhood- and adult-onset asthma are due to HLA regulatory variation acting on the HLA-DQ genes and protein coding variation in the *DRB1* gene, with *HLA-DRB1\*0401* associated with asthma risk.

## PgmNr 2234: COPD-associated miRNA-145-5p is differentially expressed in children with asthma on low lung function trajectories.

### Authors:

A. Tiwari <sup>1</sup>; J. Li <sup>1</sup>; A. T Kho <sup>1,2</sup>; M. Sun <sup>3</sup>; Q. Lu <sup>3</sup>; S. T. Weiss <sup>1</sup>; K. G Tantisira <sup>1</sup>; M. J McGeachie <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, United States of America; 2) Computational Health Informatics Program, Boston Children's Hospital, Boston, Massachusetts, United States of America; 3) Molecular and Integrative Physiological Sciences, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America

**Introduction:** MicroRNAs (miRNAs) are small non-coding RNAs that regulate their target mRNAs post-transcriptionally through degradation or translational repression. We have previously characterized lung function growth of childhood asthmatics into four patterns: normal growth (NG); reduced growth (RG); early decline (ED); and reduced growth with early decline (RGED). We have shown that reduced growth patterns can lead to COPD diagnosis prior to age 30.

**Objective:** The purpose of this study was to determine if circulating miRNAs from serum taken from an early age in asthmatic children would be prognostic of reduced lung function growth patterns over the next 16 years.

**Methods:** We performed small RNA sequencing on serums from 492 children aged 5-12 from the Childhood Asthma Management Program (CAMP) clinical trial with mild-to-moderate asthma, who were subsequently followed up for 12-16 years. We had previously assigned each CAMP participant to a longitudinal lung function pattern (NG, ED, RG, or RGED) based on their FEV1 trajectory over the course of observation. Small RNA-seq libraries were prepared using the Norgen Biotek Small RNA Library Prep Kit and sequenced on the Illumina NextSeq 500 platform. The ExceRpt pipeline was employed for QC. Mapped read counts <5 were filtered and miRNAs with coverage <50% of all subjects were removed. Using DESeq2 we normalized reads by relative log expression and identified differentially expressed miRs between lung function patterns. A significance threshold of 10% FDR was used. The top microRNA was then assessed *in vitro* for effect on human airway smooth muscle cell hypertrophy.

**Results:** After quality control, filtering and normalization, we had 448 samples and 259 miRNAs for differential analysis between the four growth patterns. In the comparison of the normal vs. the most severe group, NG vs. RGED, we found one strongly dysregulated miR, hsa-miR-145-5p (NG, n = 137; RGED, n = 89; p < 8.01E-05, FDR p = 0.021). miR-145-5p was strongly associated with airway smooth muscle cell growth *in vitro*.

**Conclusion:** Our results show that miR-145-5p, previously associated with COPD, is also associated with patterns of lung function growth leading to COPD in children with asthma and additionally increases airway smooth muscle cell proliferation. This may indicate a prophylactic role for miR-145-5p in long-term lung health and may represent a potential druggable target.

Funding: R01 HL139634, R01 HL127332, R01 HL129935, P01 HL132825



## PgmNr 2235: 11,364 whole-genome sequences in the cloud for autism spectrum disorder research.

### Authors:

S.W. Scherer<sup>1,2,3</sup>; D.M. Hartley<sup>4</sup>; B. Thiruvahindrapuram<sup>1</sup>; J. Whitney<sup>1</sup>; R.V. Patel<sup>1</sup>; J.L. Howe<sup>1</sup>; M. Bookman<sup>5,6</sup>; M. Fiume<sup>7</sup>; R.K.C. Yuen<sup>1,2</sup>; M. Zarrei<sup>1</sup>; J. Sebat<sup>8</sup>; D. Glazer<sup>5,6</sup>; T. Frazier<sup>4</sup>; B. Trost<sup>1</sup>; Autism Speaks MSSNG Collaboration

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Centre for Applied Genomics and Department of Genetics & Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 3) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 4) Autism Speaks, New York, NY, USA; 5) Verily Life Sciences, South San Francisco, CA, USA; 6) Google, Mountain View, CA, US; 7) DNASTack, Toronto, ON, Canada; 8) Beyster Center for Psychiatric Genomics, Department of Psychiatry, University of California at San Diego, San Diego, CA, USA

---

Autism spectrum disorders (ASDs) are neurodevelopmental conditions characterized by social-communication deficits and restricted, repetitive behaviours. ASDs are highly heritable and have a complex genetic etiology, with more than 100 genes implicated to date. The Autism Speaks MSSNG resource is a database of whole-genome sequencing (WGS) and phenotypic data from families affected with ASDs, and recently achieved the milestone of 11,364 individuals sequenced. These include 5,134 ASD-affected individuals from 4,231 families, of which 2,941 include both parents. Of the families with both parental sequences available, 741 are multiplex and 2,200 are simplex. Anyone with an ASD-related research question can apply for access to MSSNG, with 69 principal investigators and 179 users from 59 institutions in 14 countries having been approved for access to date. MSSNG data can be analyzed via either a cloud-based environment or a web-based portal, and is also part of the Beacon Network. MSSNG currently includes sequence data aligned to GRCh38/hg38, single nucleotide variants (SNVs), short insertions/deletions (indels), and copy number variants (CNVs); the next iteration will include several enhancements, including improved phenotype querying, epigenetic data, and structural variants (SVs).

Given their phenotypic and genetic heterogeneity, furthering our understanding of the genetic basis of ASDs requires all categories of variants to be considered, including those of differing frequency (common and rare), size (from SNVs and indels to CNVs and SVs), and location (coding and non-coding). Unlike chromosomal microarrays and whole-exome sequencing, WGS allows almost every type of variant to be detected; thus, we are currently examining all variant categories, with a focus on those that are refractory to other technologies, such as non-coding variants and SVs. We are applying state-of-the-art machine learning methods to identify regulatory regions disrupted in individuals with ASDs. To enable functional analysis of potentially clinically-relevant variants, we have generated 63 iPS-derived neuronal cell lines from individuals with ASDs and familial controls, and another 25 lines using CRISPR modelling in an isogenic line. Overall, the Autism Speaks MSSNG project combines high-quality WGS data with extensive phenotype information to allow researchers of all backgrounds to help further our understanding of the genetic architecture of ASDs.

## PgmNr 2236: Genomic prediction of objective sleep behavior in sleep-deprived and shift-work population.

### Authors:

Y. Fang <sup>1</sup>; D. Forger <sup>2</sup>; S. Sen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular & Behavioral Neuroscience Institute, University of Michigan, Ann Arbor, MI; 2) Department of Mathematics, University of Michigan, Ann Arbor, MI

---

Medical internship is characterized of sleep deprivation and shift work. Maintaining enough sleep and stable sleep pattern is crucial to medical interns' wellness and performance. Recent GWAS of morningness and sleep duration and derived polygenic score have elevated the perspective to study the biological basis of sleep behavior. Meanwhile, the emerging of wearable devices have provided inexpensive tools to objectively measure sleep in a more precise and convenient way than self-report. These progress have made it possible to conduct large-scale study on the relationship of biological basis and sleep behavior in sleep-deprived and shift-work populations like medical interns.

Intern Health Study is a longitudinal cohort study that assesses multiple aspects of medical interns around US. For 2017 cohort, whose medical internship lasted from July 2017 to June 2018, we followed 316 medical interns since baseline (one to two months prior to internship) until the end of internship. We collected DNA samples from the subjects and generated polygenic scores of morningness and sleep duration based on a recent GWAS from UK Biobank (Jones et al., 2016). We also distributed *Fitbit* to the subjects, and collected daily measures of sleep duration and timing.

With linear regressions adjusted by age, gender and top 10 principal components, we found that higher morningness polygenic score predicted smaller deviations in internship sleep timing, especially wake time, from baseline routine. One SD higher in morningness polygenic score was related to 13 less minutes in the change ( $p=0.007$ ). Also, higher sleep duration polygenic score predicted longer average daily sleep time both at baseline ( $\beta=7$  mins,  $p=0.019$ ) and during internship ( $\beta=14$  mins,  $p=2.35 \times 10^{-4}$ ), and the association during internship was marginally stronger ( $\beta=6$  mins,  $p=0.055$ ).

Our results indicated that genomically predisposed morning person was more prone to maintain a similar schedule to baseline during internship. And difference in sleep duration predicted by genomic predisposition was even wider during internship than usual. Previously we discovered that longer sleep and less deviation from baseline sleep routine were associated with better mood of medical interns (Kalmbach et al., 2018). Therefore, a deeper understanding of biological basis of sleep behavior may have potential to guide personalized intervention in improving sleep and mental health within medical interns and other sleep-disturbed professionals.



## **PgmNr 2237: Genetic correlations between brain imaging traits and dietary choices reveal complex interactions between diet and brain morphology.**

### **Authors:**

E.J. Grzeszkowiak <sup>1</sup>; J.R.B. Perry <sup>2</sup>; J.F. Wilson <sup>1,3</sup>; N. Pirastu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland; 2) MRC Epidemiology Unit, Cambridge, CB2 0SL, UK; 3) MRC Human Genetics Unit, Institute of Genetic and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland

---

In the developed world, the genetics of food and drink consumption plays a significant role in population health by influencing the incidence of numerous diet-related diseases such as obesity, type II diabetes, or coronary heart disease – it is estimated that in the US alone unhealthy diet contributes to 678,000 deaths each year. It has been demonstrated previously that genes significantly associated with dietary choices are differentially expressed in the brain. In this study we aim to understand the genetic overlap between food choices and several brain characteristics ranging from brain area volume to activation during rest and activity. To do this we estimated the genetic correlations between genome-wide association of several specific food consumption traits such as beef and cheese and brain imaging traits from UK Biobank. The brain phenotypes were obtained using T1-weighted structural imaging, resting-state functional MRI, task functional MRI, T2-weighted FLAIR structural imaging, diffusion imaging, and susceptibility-weighted structural imaging and include traits such as white matter lesions, total brain volume and volumes of specific structures. Preliminary results show that dietary choices that fall into the cluster of “unhealthy diet” (i.e. meat/fatty foods and alcohol) tend to be negatively correlated with brain volume traits, while foods usually considered “healthy” (i.e. fruit and vegetables) show an opposite pattern. These correlations remain significant after normalisation for head size, implying that the effects are not caused by body size. Looking at specific brain areas we find that the volume of hippocampus, which has been shown to play an important role in the regulation of human food intake, has a positive genetic correlation with the consumption of fortified wine and salt and a negative correlation with the consumption of beer and cider, decaffeinated and instant coffee as well as bread. Optic chiasm volume is positively genetically correlated with the consumption of lamb and processed meat. The findings give us a better understanding of the complex interactions between dietary choices and brain morphology that in the future might allow for a more personalised approach to nutrition and prevention of diet-related diseases.

## **PgmNr 2238: A genetic landscape of human behavioral traits: A cross-domain analysis of heritability, polygenicity, and biological pathways.**

**Authors:**

S.K. Jang; M.Z. Liu; S. Vrieze

[View Session](#) [Add to Schedule](#)

**Affiliation:** Psychology, University of Minnesota, Twin Cities, Minneapolis, Minnesota.

---

It is a fundamental observation in behavioral science that behaviors are heritable and correlated. Theories on the etiology of behaviors will benefit from knowledge on whether differential heritability exists, and whether and how the genetic correlations are systematically structured and what gives rise to this structure. The current study examines whether narrow-sense heritability specifically estimated by *ldsc* is different across behavioral, non-behavioral, and brain phenotypes as well as across different behavioral domains. It also examines the structure of genetic correlations among behavioral and related- physical traits using dimension reduction techniques and inspects whether this might correspond to the shared biological pathways of behaviors. A total of 61 GWAS summary statistics were collected, including 19 physical, 35 behavioral and 7 brain volume phenotypes. Physical phenotypes included medical disease, anthropometric traits, and biomarkers, while behavioral traits included psychiatric disorders, substance use, sleep, lifestyle, diet, and social relationship. A random-effects meta-analysis on the heritability produced a weighted mean of heritability for each phenotype domain. Behavioral phenotypes showed relatively smaller heritability ( $h^2=0.09$ , 95%CI=0.071-0.109) than that of physical ( $h^2=0.18$ , 95%CI= 0.13-0.22) and brain phenotypes ( $h^2=0.19$ , 95%CI=0.13-0.24). The analysis of within-domain genetic correlations indicated that brain volume and behavioral phenotypes tend to be more highly genetically correlated to themselves than physical phenotypes with the median absolute value of correlation 0.24, 0.13, and 0.05 for each domain, respectively. The graph analysis of genetic correlations categorized behavioral traits into two clusters with one associated with most psychiatric disorders and reduced fertility, and the other associated with externalizing behaviors and cognitive ability. We will conduct a sensitivity analysis on our differential heritability results with a recently developed summary-statistics based heritability estimates using *SbayesR* and discuss potential sources of differential heritability including measurement errors of psychological constructs. Further, results of cross-domain analyses of polygenicity and biological pathways will be reported.

## PgmNr 2239: Cathepsins play a major role in alpha-synuclein accumulation in Lewy body disorders.

### Authors:

N. Tayebi <sup>1</sup>; R. McGlinchey <sup>2</sup>; S. Lacy <sup>2</sup>; K. Huffer <sup>2</sup>; E. Sidransky <sup>1</sup>; J. Lee <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics branch, NIH/NHGRI, Bethesda, Maryland.; 2) laboratory of Protein Conformation and Dynamics, NHLBI/NIH, Maryland

---

Synucleinopathies including Parkinson disease (PD) are characterized by the presence of Lewy bodies (LBs) in neuronal cells of substantia nigra, which contain accumulations of the protein  $\alpha$ -synuclein (SNCA). The biophysical properties and proteolytic processing of SNCA are poorly understood. Recently, there has been evidence that cathepsins (Cts), endolysosomal cysteine proteases, are involved in the cleavage of SNCA. It has also been shown that truncations in the C-terminal of SNCA promote fibrillization and are enriched in LBs. To identify the origin of the truncated forms of SNCA, we used a *SNCA*<sup>A53T</sup> mouse model, where human transgenic A53T mutated SNCA is overexpressed. The mice were monitored for weight loss and neurological symptoms, including impaired axial rotation and limb paralysis. Symptomatic mice were euthanized, and lysosomes were isolated from brain samples along with samples from age matched wildtype and non-symptomatic *SNCA*<sup>A53T</sup> mice by density gradient. Using antibodies specific for different epitopes, including C- and N-terminal SNCA, we compared levels of the different SNCA species in purified brain lysosomes by immunoblotting. The data revealed enrichment of C-terminal truncated SNCA in symptomatic mice, which results from cleavage by CtsB and CtsL. HPLC and mass spectrometry were used to identify the specific amino acids sequence in the C-terminal truncated forms. No changes were observed using different substrates to measure Cts activity. The results were confirmed in rat dopaminergic N27 cells treated with exogenous SNCA fibrils, indicating that the terminal truncated forms of SNCA originated from incomplete degradation of fibrils in lysosomes. This work demonstrates the essential role of brain lysosomal function in PD. The increased levels of C-terminal truncated SNCA in LBs appears to result from the burden of aggregated SNCA inside the lysosome and not inactivation of Cts. Most likely, the C-terminal truncated forms of SNCA migrate to the cytosol, where they initiate SNCA aggregation, resulting in LBs formation.

## PgmNr 2240: Determining genetic correlation and causality of clinical biomarkers in human disease using population biobanks.

### Authors:

E.M. Wigmore <sup>1</sup>; M. Kurki <sup>2</sup>; P.D. Briotta Parolo <sup>2</sup>; A. Palotie <sup>2,3,4</sup>; M.J. Daly <sup>2,3,4</sup>; A. Matakidou <sup>1</sup>; FinnGen

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Genomics Research, Discovery Sciences, Biopharmaceutical R&D, AstraZeneca, Royston, Hertfordshire,, United Kingdom; 2) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Finland; 3) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

---

Genetic correlation is frequently reported between intermediate traits, such as biomarkers, and human diseases. These correlations could either represent causal associations of biomarkers in human disease, genetic subgrouping or non-causal relationships such as secondary effects of the disease/treatment biology or pleiotropy. In this study, we used summary statistic data from genome-wide association studies in two large population biobanks, FinnGen (n=~100,000) and UK Biobank (UKB, n=~500,000), to elucidate the relationship between blood count measures and human disease. Linkage disequilibrium score regression was applied to assess genetic correlation ( $r_g$ ) between disease (FinnGen) and blood counts (UK Biobank). Results indicated both well-known correlations; such as eosinophils and asthma ( $r_g = 0.35$ ,  $se = 0.033$ ,  $P = 9.2 \times 10^{-27}$ ) and reticulocytes and diabetes ( $r_g = 0.33$ ,  $se = 0.026$ ,  $P = 8.3 \times 10^{-36}$ ), as well as less well described associations such as reticulocytes and nerve and plexus disorders ( $r_g = 0.18$ ,  $se = 0.030$ ,  $P = 9.7 \times 10^{-10}$ ). To assess whether significant associations represent causal relationships, subgrouping or pleiotropy, Mendelian Randomisation and BUHMBOX were applied. This analysis highlights the utility of large biobanks in assessing the role of intermediate traits using a disease-agnostic approach that allows for unbiased assessment and the identification of novel causal associations and disease biology.

# PgmNr 2241: Characterization of CFTR mutations in chronic obstructive pulmonary disease.

**Authors:**

A. Saferali; A.A. Diaz; M.H. Cho; C.P. Hersh

[View Session](#) [Add to Schedule](#)

**Affiliation:** Brigham and Women's Hospital

---

**Introduction**

Chronic obstructive pulmonary disease (COPD) is a complex disease caused by cigarette smoking and influenced by genetic contributors. COPD is phenotypically heterogeneous, with varying manifestations of emphysema, chronic bronchitis, and airway wall thickening and bronchiectasis despite similar degrees of lung function impairment. While mutation in the cystic fibrosis transmembrane conductance regulator (*CFTR*) is associated with cystic fibrosis (CF), data suggests that cigarette smoke results in impaired *CFTR* function in smokers with and without COPD. Here, we hypothesize that heterozygous mutations in *CFTR* are associated with COPD and related phenotypes.

**Methods:**

Whole genome sequencing was performed in 5773 non-Hispanic white subjects, including 3150 moderate COPD (GOLD stage 2-4) cases and 3629 controls, from the COPDGene study through the TOPMed program. Variants within the *CFTR* gene boundary were annotated using the WGS Annotator pipeline. We examined the occurrence of previously described clinically annotated *CFTR* variants, and performed single and grouped variant testing for COPD related phenotypes. All association analyses were adjusted for principal components of genetic ancestry, age, gender, pack-years, and current smoking status.

**Results**

Within the *CFTR* gene boundary we identified 11,567 polymorphic variants, including 10,577 SNPs and 990 indels. Of these, 10,442 have been previously described. The majority of these variants are located in intergenic (1831) or intronic (7611) regions; 199 are nonsynonymous coding variants. Out of the 400 variants in the CFTR2 database of CF mutations, 75 were polymorphic in our data; of these 29 are known to be CF-causing. We identified 177 individuals who are heterozygous for deltaF508 mutation, and 80 subjects heterozygous for the 28 additional known CF-causing mutations. This combination of 29 *CFTR* mutations was nominally associated with chronic bronchitis (0.015) but not COPD or emphysema. Furthermore, burden testing weighting with percent pancreatic insufficiency as a measure of mutation severity revealed that the CF causing mutations were additionally associated with bronchodilator response ( $p=0.018$ ).

**Conclusions**

We found that a subset of 29 *CFTR* variants may be associated with chronic bronchitis in non-Hispanic white smokers. This suggests that heterozygous *CFTR* mutations together with cigarette smoking could result in pulmonary disease through lack of functional *CFTR* activity.

# PgmNr 2242: Common and low-frequency coding variants at the *ERN1* and *XBP1* loci are associated with dental caries in the primary dentition.

## Authors:

Y. Zhou <sup>1</sup>; D. McNeil <sup>2</sup>; G. Wehby <sup>3</sup>; F. Deleyiannis <sup>4</sup>; J. Hecht <sup>5</sup>; R. Weyant <sup>1</sup>; R. Crout <sup>2</sup>; E. Feingold <sup>1</sup>; M. Marazita <sup>1</sup>; J. Shaffer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) University of Pittsburgh, Pittsburgh, PA; 2) West Virginia University, Morgantown, WV; 3) University of Iowa, Iowa City, Iowa; 4) University of Colorado, Denver, Colorado; 5) University of Texas Health Science Center at Houston, Houston, Texas

## Introduction

Enamel secretion and maturation is a tightly regulated process that impacts the quality of enamel and its ability to resist carious attacks. The Endoplasmic Reticulum (ER) stress signaling pathway IRE1/XBP1 has been indicated in enamel secretion in an IRE1 conditional knockout mouse model and an *amelogenesis imperfecta* mouse model. Here, we hypothesize that coding variants in the *ERN1* (which encodes IRE1) and *XBP1* genes contribute to caries susceptibility in humans. We investigated the association between common and low-frequency coding variants in *ERN1* and *XBP1*, with caries in the primary dentition.

## Methods

Decayed and filled primary teeth (dft) indices were collected from Caucasian, non-Hispanic children ages 1 to 12.9 years from two cohorts: the Center for Oral Health Research in Appalachia (COHRA, N=619), and The Pittsburgh Orofacial Clefting (POFC, N=559). Exome variants were genotyped on the Illumina Exome chip for COHRA and on Exome add-on for POFC. Variance components modeling was used to individually test the association between common missense SNPs (*ERN1*:rs186305118 and *XBP1*:rs5762809) and dft. Low-frequency coding variants in *ERN1* and *XBP1* were tested in aggregate by Combined and Multivariate Collapsing (CMC) and Sequence Kernel Association Test (SKAT) using an MAF cutoff of 1% in a gene-based framework. Analyses were adjusted for ancestry and pertinent covariates.

## Results

We discovered that the common missense variant in *ERN1* rs186305118 was associated with dft (beta=1.57 (SE = 0.09), p=0.009) in the COHRA cohort, but not in the POFC cohort. This variant was predicted to be damaging by Combined Annotation-Dependent Depletion (CADD). Gene-based association tests revealed a significant association between low-frequency coding variants in the *XBP1* locus and dft in the POFC cohort ( $p^{\text{CMC}} = 0.0006$ ,  $p^{\text{SKAT}} = 0.005$ ). The association was mainly driven by a variant that causes a frameshift in exon 1 of *XBP1* (c.14\_15insGCCG), and is predicted to result in the production of a truncated nonfunctional protein via PROVEAN (Protein Variation Effect Analyzer).

## Conclusion

Protein-coding variants predicted to be damaging at the *ERN1* and *XBP1* loci are associated with dental caries in the primary dentition. These data, coupled with experimental evidence, indicate a possible regulatory role for *ERN1* and *XBP1* genes in enamel development and caries etiology in children.

## PgmNr 2243: Assessment of oxidative stress among agriculture workers exposed to pesticides in Punjab.

**Authors:**

M. Ahluwalia; A. Kaur

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, Guru Nanak Dev University, Kapurthala, Punjab, India

---

Pesticides are exogenous products that are metabolised to active intermediates leading to generation of free radicals, creating oxidative stress in the biological systems and obligate changes in human defence system. Farmers are occupationally exposed to pesticides and thus are more prone to oxidative stress. The present study was aimed towards analysis of oxidative stress among agricultural workers. A total of 538 subjects were enrolled in the study which included agricultural workers (n=260) and non-exposed subjects (n=278) from Punjab. Calorimetric techniques were used to analyse the oxidative stress by estimation of malondialdehyde (MDA), reduced glutathione (GSH) levels and superoxide dismutase (SOD) activity. A significant increase in lipid peroxidation was observed among exposed group indicated by significantly increased levels of MDA as compared to non-exposed subjects. Alternatively, a decreased SOD activity was seen in exposed subjects. The GSH levels were slightly higher in agriculture workers but, the difference was not significant. The exposed and non-exposed subjects were stratified w.r.t. demographic factors, perturbation in levels of oxidative stress parameters dependent upon alcohol intake, previous medical history of co-morbid conditions, duration of exposure, use of personal protective equipment, recent exposure to pesticides, occurrence of acute health problems were seen among exposed subjects. However, in the absence of occupational exposure, all demographic factors affected the one or more oxidative stress parameters among non-exposed subjects. Thus, a positive association of pesticide exposure with oxidative stress was observed. Since oxidative stress affects reproduction, progression of cancer, and other disorders, strict regulations for use are recommended.

## PgmNr 2244: Assaying lung-specific accessible chromatin to predict the causal variants in COPD.

**Authors:**

C.J. Benway; F. Du; J. Liu; M.H. Cho; E.K. Silverman; X. Zhou

[View Session](#) [Add to Schedule](#)

**Affiliation:** Channing Division of Network Medicine, Brigham & Women's Hospital, Boston, Massachusetts.

---

Genome-wide association studies have identified dozens of loci associated with risk of chronic obstructive pulmonary disease (COPD). However, identifying the causal variants and their functional role in the appropriate cell type remains a major challenge; the latter in particular because of the dearth of regulatory profiling in human lung cells.

We generated ATAC-seq data for four primary lung cell types implicated in disease pathology: ATII, bronchial epithelial (NHBE), small airway epithelial (SAEC), and lung fibroblast (LF) cells, as well as the 16HBE human bronchial epithelial cell line. We used these profiles as well as ENCODE data to generate cell type-specific regulatory sequence vocabularies using deltaSVM.

We predicted the effect of 1,235 COPD risk variants corresponding to 22 COPD loci in five lung cell types. We observed that in 14 of the 22 loci (63.6%) the index variant deltaSVM scored in the highest decile for positive or negative effect for at least one cell type, suggesting a functional role. Four of the index variants (near CHRNA5, RIN3, EEFSEC, and DSP) scored large deltaSVM values across the four primary lung cell types. However, at most loci the largest deltaSVM values were at non-index variants. Even in the EEFSEC locus, two nearby variants, rs2955084 (127 bp downstream) and rs2999082 (5 kb upstream), scored comparably large deltaSVM values. The top scoring variants in the COPD credible set suggest novel functional roles for several single nucleotide polymorphisms (SNPs) for nine additional loci.

We further compared the functional consequences of COPD variants in a diversity of cell types and tissues from ENCODE and observed that many variants with the largest predicted effect sizes were not cell type specific, which could be consistent with functional variants in COPD affecting multiple tissues. Correlation and clustering of the samples revealed cell-specific patterns of variant effects, suggesting distinct regulatory effects of COPD genetic factors in lung epithelium, endothelial, and immune cells. In total, this analysis identified >100 COPD risk variants with novel putative functional effects in specific lung cell types. Further experimental validation will be required to evaluate function of these candidate causal variants in vitro and in vivo.



## PgmNr 2245: Dissecting the genetic basis of variation in cocaine and methamphetamine consumption in *Drosophila melanogaster*.

### Authors:

B.M. Baker; R.R.H. Anholt; T.F.C. Mackay

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clemson University, Greenwood, SC

---

Studies on *Drosophila melanogaster* can identify genetic and transcriptional networks that underlie variation in voluntary consumption of cocaine and methamphetamine to serve as a blueprint for subsequent studies on humans. Exposure to these psychostimulants in flies results in behavioral and physiological effects that resemble those observed in humans. We derived an outbred advanced intercross population (AIP) from 37 of the sequenced inbred wild-derived lines of the *Drosophila melanogaster* Genetic Reference Panel (DGRP). These lines are maximally genetically divergent, have minimal residual heterozygosity, are not segregating for common inversions, and are not infected with *Wolbachia pipientis*. We assessed voluntary consumption of sucrose, methamphetamine-supplemented sucrose and cocaine-supplemented sucrose and found significant phenotypic variation in the AIP, in both sexes, for consumption of both drugs. We performed whole genome sequencing and extreme QTL mapping on the top 10% of consumers for each replicate, sex and condition, and an equal number of randomly selected flies. We evaluated changes in allele frequencies genome-wide among high consumers and the control flies and identified 3,033 variants associated with increased consumption that reside in 1,963 genes, enriched for genes associated with nervous system and mesoderm development. We assessed the effects of ubiquitous RNA interference (RNAi) on consumption for 22 candidate genes, of which 14 showed a significant increase or decrease in consumption. Extensive recombination in the AIP generates increased statistical power compared to genome-wide association analysis of the DGRP and illustrates the polygenic genetic architecture that underlies variation in cocaine and methamphetamine consumption. Supported by NIH grant U01DA041613.

## PgmNr 2246: Distinguishing genetic and tissue regulation of gene expression in human multi-traits by network-based stratification.

### Authors:

Y. Dai <sup>1</sup>; G. Pei <sup>1</sup>; Z. Zhao <sup>1,2,3</sup>; P. Jia <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 2) Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA

---

Context-specific genetic regulation of gene expression plays significant roles in disease etiology. Transcriptome-wide association studies (TWAS) using tissue expression profile opened a new avenue to understand complex diseases in a tissue-specific way. In this study, we collected summary statistics of genome-wide association studies (GWAS) for 57 complex traits from 4 major groups: immune related, psychiatric, cardiovascular, and anthropometric traits. For each trait, we determined trait-related tissues by conducting a tissue-specific enrichment analysis using our recent method, deTS, and conducted TWAS in these tissues. As a result, we constructed a *gene*×*TWAS* matrix containing ~500 TWAS in 32 tissues for 57 traits, with each column representing a tissue TWAS result for a trait. The matrix was then overloaded to a human protein-protein interaction network, followed by network propagation and stratification. Using the uniform manifold approximation and projection (UMAP) method, we found that a majority of traits were clustered by the original trait groups (51/57) while others by tissues (6/57). With a stratified non-negative matrix factorization method, we identified 17 principal patterns (PPs) that represent the major variation in the data. Among them, 10/17 PPs were trait-driven, i.e., the significant trait-tissue pairs were enriched with a trait group, and 5/17 PPs were tissue-driven, i.e., the significant pairs were enriched with a few tissues. For example, we found the 8<sup>th</sup> PP (trait-driven) was enriched with education level and college attendance and the 13<sup>th</sup> PP (trait-driven) was enriched with height TWAS from various tissues. We also found genes contributing differently to each PP. For example, we found the gene *C4A*, which was recently identified as a causal gene in schizophrenia, was most contributive to a brain-driven PP and two trait-driven PPs (immune and blood related traits). Another gene, *SORT1*, a causal gene to lipoprotein cholesterol, was highlighted in a liver-driven PP for total cholesterol, high-density lipoproteins and low-density lipoproteins. Interestingly, the gene *STXBP5*, which was originally reported in plasma and attention-deficit/hyperactivity disorder GWAS, was found in the 8<sup>th</sup> PP (education and college attendance). In sum, our findings distinguished and quantified the contribution of genetic and tissue-specific regulation underlying complex traits, providing insights into the pleiotropic effects of genes and disease etiology.

## PgmNr 2247: Association of lung function polygenic risk scores with chronic obstructive pulmonary disease, quantitative computed tomography imaging features, and lung growth patterns.

### Authors:

B.D. Hobbs<sup>1,2</sup>; M. Moll<sup>1,2</sup>; P. Sakornsakolpat<sup>1,3</sup>; N. Shrine<sup>4</sup>; S. Gharib<sup>5</sup>; M. Obeidat<sup>6</sup>; L. Lahousse<sup>7,8</sup>; D. Meyers<sup>9</sup>; M.J. McGeachie<sup>1</sup>; R. Tal-Singer<sup>10</sup>; A. Manichaikul<sup>11,12</sup>; S. Won<sup>13,14,15</sup>; S.T. Weiss<sup>1</sup>; E.K. Silverman<sup>1,2</sup>; M.D. Tobin<sup>4,16</sup>; M.H. Cho<sup>1,2</sup>; International COPD Genetic Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; 2) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA; 3) Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; 4) Genetic Epidemiology Group, Department of Health Sciences, University of Leicester, Leicester, UK; 5) Computational Medicine Core, Center for Lung Biology, UW Medicine Sleep Center, Department of Medicine, University of Washington, Seattle, WA, USA; 6) University of British Columbia Center for Heart Lung Innovation, St. Paul's Hospital, Vancouver, British Columbia, Canada; 7) Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 8) Department of Bioanalysis, Ghent University, Ghent, Belgium; 9) Department of Medicine, University of Arizona, Tucson, AZ, USA; 10) GlaxoSmithKline R&D, Collegeville, PA, USA; 11) Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA; 12) Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA; 13) Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea; 14) Interdisciplinary Program of Bioinformatics, College of National Sciences, Seoul National University, Seoul, Republic of Korea; 15) Institute of Health and Environment, Seoul National University, Seoul, Republic of Korea; 16) National Institute for Health Research Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, UK

---

Chronic obstructive pulmonary disease (COPD) is one of the leading causes of morbidity and mortality worldwide. Cigarette smoking is a major risk factor, though only a minority of smokers develops COPD; the estimated heritability is approximately 40%. We hypothesized that a polygenic risk score (PRS) developed from genome-wide association studies (GWAS) of lung function predicts moderate-to-severe COPD, quantitative computed tomography (CT) imaging phenotypes, and patterns of reduced lung growth in children and young adults.

We constructed a PRS using penalized regression accounting for linkage disequilibrium (lassosum) based on summary statistics from GWASs of two measures of lung function in UK Biobank (n=321,047) and SpiroMeta (n=79,055). We tuned model parameters using participants from the GenKOLS case-control study and calculated PRSs in 16 cohorts: COPDGene (non-Hispanic white [NHW] and African American [AA] subjects), ECLIPSE, NETT/NAS, MESA (4 ancestry groups), CHS (African-American and European), Lung Health Study, SPIROMICS NHW, Rotterdam (3 NHW cohorts), and KWU (Korean). In the 16 cohorts, we tested for association between a combined lung function PRS and: moderate-to-severe COPD and five distinct quantitative CT imaging phenotypes. Results were meta-

analyzed across cohorts of either European or non-European ancestry. We also tested the combined PRS association with patterns of reduced lung growth in the Childhood Asthma Management Program (CAMP) cohort.

The combined lung function PRS was associated with COPD in European (OR per PRS SD: 1.7-1.8,  $p=1.6e-156$ ) and non-European ancestry (OR per PRS SD: 1.3-1.5,  $p=4.7e-24$ ) cohorts. Compared to individuals in the first decile of the combined PRS, the tenth decile had an increased odds of COPD in European (OR 6.1 [95% CI: 5.1-7.4,  $p=2.4e-76$ ]) and non-European ancestry (OR 4.8 [95% CI: 3.5-6.8],  $p = 3.5e-16$ ) cohorts. The PRS was associated with four of the five tested CT imaging phenotypes, including percent emphysema ( $p=3.0e-17$ ), wall area percent ( $p=4.7e-38$ ), gas trapping ( $p=0.015$ ), and the 15<sup>th</sup> percentile of the lung density histogram ( $p=2.5e-58$ ). The combined PRS was also associated with a reduced lung growth pattern ( $p = 1.1e-5$ ) in CAMP.

A combined lung function PRS predicted COPD in European and non-European ancestry cohorts and was associated with COPD-related CT imaging features and patterns of reduced lung growth.

## PgmNr 2248: Use of whole transcriptome sequencing to detect molecular signatures from stress reduction interventions.

### Authors:

P.S. Lai <sup>1</sup>; R. Sethi <sup>1</sup>; R.P. Ebstein <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Paediatrics, National University of Singapore, Singapore 119228; 2) Southwestern University of Finance and Economics, Chengdu, P.R. China 611130; 3) Psychology, The Hebrew University of Jerusalem, Israel 91905

---

Mindfulness Based Stress Reduction (MBSR) and Music Therapy Stress Reduction (MTSR) programs have been used as clinical interventions to reduce stress and improve chronic disease outcomes in complex conditions such as depression, chronic pain, dementia, etc. However, molecular understanding of the biological underpinnings of such interventions remains limited. In this study, we investigate the feasibility of using whole transcriptome sequencing on RNA from peripheral blood to evaluate gene expression changes following either MBSR or MTSR programs in a cohort of 154 healthy individuals at three time points: baseline level, 2 and 3 months after intervention. All subjects were randomly age-and sex-matched into the two groups of either MBSR or MTSR. Each stress reduction course consisted of structured 2-hour sessions over an 8 week period. The courses were led by certified MBSR and music therapy facilitators. RNA-sequencing was carried out from peripheral blood cells of subjects and reads were aligned using HISAT2 and gene expression changes were estimated at baseline and compared between subjects after 2 and 3 months of MBSR and MTSR using Ballgown R package. MBSR and MTSR groups were also compared with each other at Baseline, 2 months and 3 months. The genes significantly upregulated or downregulated were estimated using a p-value < 0.01 and fold change > 2 (upregulated) and fold change < ½ (downregulated). The results show differential expression of genes involved in immune, neurological, signaling and stress-related pathways such as NF-kappaB Signaling, Mitogen-activated protein kinase (MAPK) pathway and IL2 signaling pathway. Both mindfulness and music interventions could elicit downregulation of pro-inflammatory genes as much as 4 folds while adaptive and innate immune genes were upregulated by 2 -14 folds from baseline. The data points to a complex network of genes and transcripts that regulate stress response and nervous system and provide evidence of molecular response to environmental stimuli through mindfulness meditation and music therapy.

## PgmNr 2249: Whole genome sequence association study of sleep duration in the TOPMed program.

### Authors:

J. Lee<sup>1</sup>; T. Sofer<sup>1,2,3</sup>; S. Redline<sup>1,2,4</sup>; D. Gottlieb<sup>1,2,5</sup>; TOPMed Sleep Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Sleep and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 2) Division of Sleep Medicine, Harvard Medical School, Boston, MA, USA.; 3) Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA, USA.; 4) Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA.; 5) VA Boston Healthcare System, Boston, MA, USA.

---

**Background:** Sleep duration is heritable, and short and long sleep durations are each associated with numerous adverse health outcomes. Prior genetic studies did not evaluate rare variants. We performed a whole genome sequence (WGS) association study using multi-ethnic data from the Trans-Omics for Precision Medicine (TOPMed) program in Freeze 5b.

**Methods:** We harmonized self-reported usual sleep duration across eight cohorts. We performed associated analyses with sleep duration, short vs normal sleep, and long vs normal sleep, where short sleep was defined as usual sleep  $\leq 5$ hrs, normal sleep as 6-8hrs, and long sleep as  $\geq 9$ hrs. For binary traits, we removed cohort-ethnicity subgroups with fewer than two cases. Sample sizes were 22304, 20021, and 19486, for the continuous, short, and long sleep duration analyses, respectively. We applied linear (sleep duration) and logistic (short and long sleep) mixed-models, with random effects modeled via a genetic relatedness matrix to account for population structure. All models were adjusted for age, sex, and cohort-ethnicity subgroup. For sleep duration, rank-normalized residuals were used in the association analyses. We tested single variants with minor allele count using the Score test for sleep duration, and SAIGE/SPA for short and long sleep. We also tested rare variants in aggregate units of genes and functional categories, and non-overlapping 3000bp sliding windows of variants with a FATHMM-XF score  $> 0.5$  and maximum alternate allele frequency of 0.01. For aggregate tests we used both the SKAT test with weights of Beta(1, 25), and unweighted burden test. In each set, we required at least 3 variants, and at least 20 counts of alternate alleles for the burden test.

**Results:** Rs141617869 (MAF=0.08) located 27kb downstream of *BMP15* in the Xp11.22 region, was associated with sleep duration ( $p=2.7e-08$ ); and two rare variants (MAF=0.002) in the 2p22.3 region were associated with short sleep ( $p=2.6e-08$ ). The top gene associations were *TNFRSF10D* for short sleep (high confidence loss of function; SKAT  $p=3.4e-09$ ), *IL31RA* for long sleep (high confidence loss of function; SKAT  $p=1.0e-08$ ), and *XRCC1* for sleep duration (missense; burden  $p=9.9e-07$ ). There were additional significant sliding window associations ( $p<1e-8$ ).

**Conclusion:** Novel variants and genes were associated with sleep duration, short sleep, and long sleep, including an X-linked variant and genes associated with inflammation and DNA damage.

## PgmNr 2250: A GWAS identifies novel susceptibility loci for irritable bowel syndrome.

### Authors:

J. Li <sup>1,2</sup>; T. Wang <sup>2</sup>; F. Mentch <sup>1</sup>; X. Chang <sup>1</sup>; H. Qu <sup>1</sup>; J. Glessner <sup>1</sup>; H. Hakonarson <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.; 2) Department of Cell Biology, 2011 Collaborative Innovation Center of Tianjin for Medical Epigenetics, Tianjin Key Laboratory of Medical Epigenetics, Tianjin Medical University, Tianjin, China.; 3) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Irritable bowel syndrome (IBS) is the most frequently diagnosed functional gastrointestinal disorder, which is characterized with recurrent abdominal pain and discomfort. The worldwide prevalence of IBS ranges from 10% to 20% and over 60% of IBS patients are female. The pathogenesis of IBS is complex and there is no specific laboratory test or effective cure treatment. It is essential to characterize the genetic components of IBS in order to further understand its pathogenesis. We conducted a genome-wide association study including 2164 IBS cases and 2097 controls of African American ancestry. After quality control filtering and genome-wide imputation, we found a significant locus at 8q21.13 (rs113344155, OR=1.98, P-value=2.52E-08). SNP rs113344155 is located in gene *CHMP4C* which encodes a protein belonging to the chromatin-modifying protein/charged multivesicular body protein (CHMP) family. Several loci were approaching genome-wide significance, such as SNP rs116608329 at 4p13.1 (OR=0.48, P-value=2.42E-07). GWAS in 2348 IBS cases and 4481 controls of European ancestry did not yield any genome-wide significant locus. SNP rs10152423 showed suggestive level of significance (OR= 0.69, P-value=1.7E-06). Further replication and functional characterization of these loci are underway.



## PgmNr 2251: Genome-wide association study identifies genetic polymorphisms and genes possibly associated with chronic pain and postherpetic neuralgia.

### Authors:

D. Nishizawa <sup>1</sup>; H. Arita <sup>2</sup>; K. Hanaoka <sup>2</sup>; C. Yajima <sup>2</sup>; M. Iseki <sup>3</sup>; J. Kato <sup>4</sup>; S. Ogawa <sup>5</sup>; A. Hiranuma <sup>1,6</sup>; S. Kasai <sup>1</sup>; J. Hasegawa <sup>1</sup>; Y. Ebata <sup>1</sup>; K. Nakayama <sup>1</sup>; M. Hayashida <sup>1,3,7</sup>; K. Ikeda <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Psychiatry and Behavioral Science (Addictive Substance Project), Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Anesthesiology and Pain Relief Center, JR Tokyo General Hospital, Tokyo, Japan; 3) Department of Anesthesiology & Pain Medicine, Juntendo University School of Medicine, Tokyo, Japan; 4) Department of Anesthesiology, Nihon University School of Medicine, Tokyo, Japan; 5) Nihon University University Research Center, Tokyo, Japan; 6) Department of Surgery, Toho University Sakura Medical Center, Sakura, Japan; 7) Department of Anesthesiology, Saitama Medical University International Medical Center, Hidaka, Japan

---

**Purpose:** Human twin studies and other studies have indicated that chronic pain has heritability ranging from 30-70%. Various kinds of drugs were administered to patients with chronic pain for the treatment of pain, and some of these drugs may be effective for certain patients, whereas others may not. We conducted genome-wide association study (GWAS) in patients who suffer from chronic pain to identify potential genetic variants that may significantly contribute to susceptibility to chronic pain and effects of administered drugs. **Methods:** Subjects enrolled in our analyses were 191 patients with chronic pain or pain-related disorders including neuropathic pain such as postherpetic neuralgia (PHN) who visited several hospitals in Tokyo in Japan for the treatment of chronic pain. We conducted GWAS by using whole-genome genotyping arrays with more than 700,000 markers in 191 patients and 282 healthy control subjects in trend, dominant, and additive genetic models. Genome-wide associations between genetic polymorphisms and susceptibility to PHN were also investigated by comparing only patients with PHN and control subjects. Further, we applied GWAS to the effects of drugs for the treatment of pain, which included opioids, antidepressants, anticonvulsants, non-steroidal anti-inflammatory drugs (NSAIDs), and gamma-aminobutyric acid (GABA) receptor agonists, in the patient subjects. **Results:** Although none of the single-nucleotide polymorphism (SNP) markers were found to be genome-wide significantly associated with chronic pain in a GWAS in all patients ( $P \geq 1.858 \times 10^{-7}$ ), another GWAS including only PHN patients revealed that rs4773840 SNP within the *ABCC4* gene region, which mapped to 13q32.1, was identified to be significantly associated with PHN in trend model (nominal  $P = 1.638 \times 10^{-7}$ ). In additional gene-based analyses, only one gene, *PRKCQ*, was significantly associated with chronic pain in trend model (adjusted  $P = 0.03722$ ). However, none of the genes were significantly associated with PHN in any genetic models. Moreover, further GWAS on the effects of drugs for the treatment of pain in the patient subjects observed no SNPs significantly associated with the phenotypes in any genetic models. **Conclusions:** Our results suggested that the *PRKCQ* gene and the rs4773840 SNP within the *ABCC4* gene region could be related to susceptibility to chronic pain conditions and PHN, respectively.



## PgmNr 2252: Correlation between gut microbiota and 6 facets of neuroticism in Korean adults.

### Authors:

E. Park <sup>1</sup>; H.N. Kim <sup>2</sup>; Y. Chang <sup>3</sup>; S. Ryu <sup>3</sup>; H. Shin <sup>4</sup>; S.C. Jung <sup>1</sup>; H.L. Kim <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biochemistry, School of Medicine, Ewha Womans University, Seoul, Korea, Republic of; 2) Medical Research Institute, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea; 3) Center for Cohort Studies, Total Healthcare Center, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea; 4) Department of Family Medicine and Health Screening Center, Kangbuk Samsung Hospital School of Medicine, Sungkyunkwan University, Seoul, Republic of Korea

---

Personality affects fundamental behavior patterns and has been related with health outcomes and mental illness. Neuroticism is a heritable personality trait that is comprised of distinct sub-facets. We reported that Neuroticism showed significantly correlated with diversity and composition of gut microbiota (Han-Na Kim, BBI, 2018). In the present study, diversity of gut microbiota association studies for six facets of neuroticism which includes N1-Anxiety, N2-Hostility, N3-Depression, N4-Self-consciousness, N5-Impulsiveness and N6-Vulnerability was performed. This research was performed using the Korean version of the Revised NEO Personality Inventory and the sequencing data of the fecal 16S rRNA gene in 513 men and 311 women. Each subject provides a stool sample and bacterial DNA is extracted. The V3 and V4 hypervariable region of bacterial 16S rRNA gene are sequenced using Illumina MiSeq platform. DADA2 pipeline of R was used for identifying ASV feature table and 16S rRNA data was analyzed using QIIME2 software.

The diversity and the composition of the human gut microbiota exhibited significant difference when stratified by Neuroticism facet quartiles: For male, in groups with high N1, N3 and N6 scores, the observed diversity was increased while evenness shows significantly decreased. On the other hands, female showed less significant difference in microbiota diversity between low and high neuroticism facets.

Our findings implicated association between the gut microbiota and personality, and providing useful visions toward developing and testing of personality- and microbiota-based interventions for improving health condition. This research was supported by National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018 R1D1A1B07-050067).

## PgmNr 2253: Beyond GWAS: Fine-mapping, functional, and phenotypic follow-up of 44 genome-wide significant endometriosis associations in 61K cases and 711K controls.

### Authors:

N. Rahmioglu <sup>1,2</sup>; S. Mortlock <sup>3</sup>; P. Loof <sup>4,5</sup>; M. Ghiasi <sup>6</sup>; L. Stefansdottir <sup>7</sup>; R. Magi <sup>8</sup>; A. Giri <sup>9</sup>; G. Galarneau <sup>10</sup>; D. Velez-Edwards <sup>9</sup>; P. Yurttas-Beim <sup>10</sup>; V. Steinthorsdottir <sup>7</sup>; M. Nyegaard <sup>4,5</sup>; G. Montgomery <sup>3</sup>; S. Missmer <sup>6,11,12</sup>; A.P. Morris <sup>13</sup>; K.T. Zondervan <sup>1,2</sup>; International Endometriosis Genomics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK.; 2) Endometriosis CaRe Centre Oxford, Nuffield Department of Women's and Reproductive Health, University of Oxford, Oxford, UK.; 3) Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, Australia.; 4) Department of Biomedicine, Aarhus University, Aarhus, Denmark.; 5) The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark.; 6) Department of Obstetrics, Gynecology and reproductive biology, Michigan State University, Michigan, USA.; 7) deCODE genetics/Amgen, Reykjavik, Iceland.; 8) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia.; 9) Department of Obstetrics and Gynecology, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA.; 10) Celmatix Inc., New York, NY, USA.; 11) Department of Epidemiology, Harvard University, Boston, USA.; 12) Boston Center for Endometriosis, Boston, MA, USA.; 13) School of Biological Sciences, University of Manchester, Manchester, UK.

---

Endometriosis is an oestrogen-dependent condition seen in 5-10% of women of reproductive age, associated with infertility and chronic pelvic pain. The heritability has been estimated to be ~50% with 26% due to common genetic variants. We have undertaken the largest genome-wide association study (GWAS) meta-analysis of endometriosis to date, including 61,077 cases and 711,319 controls of European and Japanese ancestry. We identified 44 genome-wide significant (GWS,  $P < 5 \times 10^{-8}$ ) loci associated with endometriosis, 32 of which are novel with 22 showing larger effect sizes for severe (American Society of Reproductive Medicine (ASRM) stage III/IV) disease vs. mild (ASRM stage I/II). Conditional analyses revealed multiple distinct association signals for 11 GWS loci, including 7 mapping near the Estrogen Receptor 1 (*ESR1*) gene. A functional GWAS (fGWAS) analysis revealed enrichment of endometriosis association signals mapping to binding sites for 5 transcription factors (EGR1, GATA3, RELA, SUZ12, TFAP2C) and enhancers in the ovary (OVRY-EnhA). EGR1 mediates oestrogen activity to establish uterine receptivity for embryo implantation. GATA3 is induced by oestrogen in a dose/response manner promoting cytokine secretion in endometrial cells. TFAP2C is involved in epigenetic regulation of *ESR1* expression in breast cancer. RELA and SUZ12 are involved in immune and inflammatory response pathways. Fine-mapping of endometriosis loci, informed by enriched annotations from the fGWAS analysis, uncovered 15 high-confidence variants (>50% probability of being causal for the association signal): rs9789525 in *GREB1*, rs6456259 near *ID4*, rs851983 in *ESR1* and rs71575922 in *SYNE1*, rs979165 near *miR148a* and rs6970537 in *HOXA10*, rs10757279 near *CDKN2-BAS1*, rs507666 in *ABO*, rs10828247 near *MLLT10*, rs4071558 near *FSHB*, rs3803042 near *HOXC10*, rs6538618 near *VEZT*, rs3742716 in *RIN3*, rs7183386 in *SRP14-AS1*, and

rs73241342 near *PLAC1*. Furthermore, we have conducted analyses of RNAseq data from endometrium tissue collected from endometriosis cases (N=58) and controls (N=70), revealing 12 significantly differentially expressed genes mapping to GWS loci. Phenotypic stratified analyses of the GWS loci in deeply phenotyped datasets are ongoing. Taken together the results will provide better understanding of the underlying functional mechanisms of genetic risk variants on endometriosis and its subtypes.

## PgmNr 2254: Structural variants associated with GWAS SNPs provide mechanistic explanation of phenotypic associations.

### Authors:

S.I. Berger <sup>1,2,3</sup>; H. Barseghyan <sup>1,3</sup>; S. Bhattacharya <sup>1</sup>; E. Vilain <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Medicine Research, Children's National Medical Center, Washington, DC; 2) Rare Disease Institute/Division of Genetics and Metabolism, Children's National Medical Center, Washington DC; 3) Department of Genomics and Precision Medicine, School of Medicine and Health Science, George Washington University, Washington DC

---

**Background:** Genome wide association studies (GWAS) use common single nucleotide polymorphisms (SNPs) to identify genomic regions associated with phenotypic traits. The SNPs identified in GWAS are rarely the causal variants for the evaluated phenotype. Typically, these are genetically linked on a common haplotype carrying functional variants. Structural variants (SVs) have been reported to co-occur with SNPs identified by GWAS (Liu et al., 2018) and insertion variants occur disproportionately at GWAS loci (Payer et al., 2017). Optical genomic mapping (OGM) allows for improved detection of SVs not found using short-read DNA sequencing. SVs are among the most common sources of genomic variation and occur at polymorphic frequencies similar to SNPs (Levy-Sakin et al., 2019).

**Objective:** Using public datasets of SVs in well genotyped individuals, we characterized common SVs co-occurring with GWAS SNPs.

**Methods:** SVs identified using OGM on 154 individuals (Levy-Sakin et al., 2019) were merged with phased haplotypes from the 1000genomes project (The 1000 Genomes Project Consortium, 2015). SNPs associated with phenotypes were obtained from the NHGRI-EBI GWAS Catalog (Buniello et al., 2019). Genotype counts were calculated for 1,315,891 pairs of SNPs and SVs with allele frequencies greater than 5% located within 10 megabases of each other. Phased haplotypes were counted for all genotypes except the doubly heterozygous cases where phase is ambiguous. P-values for linkage were calculated using Fisher's exact test under the null hypothesis of independent assortment of the two loci and Bonferroni correction applied.

**Results:** There were 1,146 pairs of SVs and GWAS-reported SNPs with significant associations including loci associated with diverse traits including biometric measurements and disease susceptibilities. These pairs involved 1,098 distinct SNPs and 361 distinct SVs with candidate gene involvement consisting of 197 deletions, 157 insertions, and 7 inversions. The SV more frequently associated with the SNP minor allele in 612 pairs and the major allele in 534 pairs.

**Conclusions:** Common SVs provide plausible mechanistic explanations of many GWAS associated SNPs. Due to larger size, SVs likely have larger effects on gene expression, regulation, and splicing than SNPs. As the marker SNP is observed on haplotypes both with and without the SV, direct evaluation of polymorphic SVs may demonstrate stronger associations than SNP based GWAS.

## PgmNr 2255: Upper airway microbiota diversity in infancy is associated with nasal mucosa DNA methylation patterns and allergic rhinitis at 6 years of age.

### Authors:

A. Morin<sup>1</sup>; C. McKennan<sup>1,2</sup>; J. Thorsen<sup>3</sup>; J. Stokholm<sup>3</sup>; B. Chawes<sup>3</sup>; A.M. Schoos<sup>3</sup>; K. Naughton<sup>1</sup>; C.E. Pedersen<sup>3</sup>; D. Vercelli<sup>4</sup>; D. Nicolae<sup>1,2</sup>; K. Bønnelykke<sup>3</sup>; H. Bisgaard<sup>3</sup>; C. Ober<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, University of Chicago, Chicago, IL, USA.; 2) Department of Statistics, University of Chicago, Chicago, IL, USA.; 3) COPSAC, Copenhagen Prospective Studies on Asthma in Childhood, Herlev and Gentofte Hospital, University of Copenhagen, Gentofte, Denmark.; 4) Department of Cellular and Molecular Medicine, Arizona Center for the Biology of Complex Diseases, and Asthma and Airway Disease Research Center, University of Arizona, Tucson, AZ, USA.

---

Allergic rhinitis (AR) is a common inflammatory disease of the nasal mucosa, characterized by the presence of specific IgE (sensitization) to inhaled allergens and symptoms such as sneezing, rhinorrhea and nasal congestion. Nasal mucosa forms a physical and immune barrier to inhaled allergens and to microbes, both of which can alter immune responses and subsequent risk for disease, such as AR. We hypothesized that early life exposure to microbes in the upper airway alters the epigenetic landscape in nasal mucosa, leading to allergen sensitization (AS) and AR later in childhood. To address this hypothesis, we examined airway (hypopharyngeal) microbiota assessed by 16S rRNA gene sequencing at 1 week, 1 month and 3 months of life and DNA methylation (Illumina 850k EPIC array) profiles in cells obtained by nasal brushings at age 6 from 454 children in the Copenhagen Prospective Studies on Asthma in Childhood (COPSAC)<sub>2010</sub> birth cohort; 24 of these children were diagnosed with AR and 331 had neither AS or AR at age 6. Our study revealed that children with AR at age 6 had a significantly less diverse nasal microbiota at 1 week (richness: median 17 and 21, Wilcoxon rank sum test  $P=2.9 \times 10^{-3}$ ) and 1 month (Shannon diversity index: median 0.87 and 1.12, Wilcoxon rank sum test  $P=0.039$ ), but not at 3 months of age, compared to children with neither AS or AR. Using linear models (limma) adjusted for cell composition, we identified 720 differentially methylated probes (DMPs; FDR 5%) in nasal brush cells at age 6 (80% were less methylated in children with AR). Genes near DMPs were enriched for relevant gene ontology pathways, such as immune system process and response. Richness measures at 1 week were correlated with methylation levels at 31 DMPs (Spearman  $\rho > |0.15|$ ;  $P < 0.01$ ), which was never observed in 1000 subsamples of 720 probes that were not differentially methylated (median and maximum number of correlations with  $\rho > |0.15| = 7$  and 18; permutation  $P < 1 \times 10^{-3}$ ). These data indicate that a less diverse upper airway microbiota in infancy is associated both with subsequent development of AR and altered DNA methylation patterns in airway mucosal cells that differ between children with and without AR at age 6. Taken together our results suggest that trajectories for development of AR is modulated by microbial exposures in infancy, at least in part through epigenetic modification of nasal mucosal cells. Supported by NIH R01 HL129735.

## PgmNr 2256: Exome sequencing identify a damaging mutation in *ABI3BP* in a family with hereditary gingival fibromatosis from Nigeria.

### Authors:

W. Awotoye <sup>1,2</sup>; P. Ayanbadejo <sup>3</sup>; K. Umezudike <sup>3</sup>; O. Olagundoye <sup>4</sup>; T. Busch <sup>2</sup>; F. Ajayi <sup>5</sup>; A. Akinshipo <sup>5</sup>; A. Ogundana <sup>6</sup>; M. Li <sup>2</sup>; M. Musa <sup>2</sup>; M. Hassan <sup>2</sup>; W.L. Adeyemo <sup>7</sup>; A. Butali <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Iowa Institute of Oral Health Research, College of Dentistry, University of Iowa, Iowa, IA.; 2) Department of Oral Pathology, Radiology, and Medicine, College of Dentistry, University of Iowa, IA; 3) Department of Preventive Dentistry, Faculty of Dental Sciences, University of Lagos, Lagos, Nigeria; 4) Department of Preventive Dentistry, Faculty of Dentistry, Lagos State University College of Medicine Ikeja, Lagos, Nigeria; 5) Department of Oral and Maxillofacial Pathology/Biology, College of Medicine, University of Lagos, Lagos, Nigeria; 6) Department of Preventive Dentistry, Lagos University Teaching Hospital, Idi-Araba, Lagos, Nigeria; 7) Department of Oral and Maxillofacial Surgery, Faculty of Dental Sciences, University of Lagos

---

Hereditary gingival fibromatosis (HGF) is a slow and progressive enlargement of the attached gingiva covering the teeth and affecting speech, mastication and aesthetics. This is a rare oral disease which may be localized or generalized, may be isolated or part of a syndrome. Several genetic factors have been associated with the syndromic type however little is known about the genetic factors associated with the isolated type. In this study, we conducted exome sequencing using DNA from a Nigerian family. The father presented with a localized form of HGF while the 4 year old child has a 2 year history of generalized form of HGF. However, the mother does not have any history of this disease. Out of the 255,692 variants called 207,026 passed the quality check. VarSeq® software was used in the annotation and analyses of the variants called. The called variants were filtered for shared dominant variants (present only in the father and son but absent in the mother). These variants are expected to affect the structure and function of the protein hence results in the phenotype based on bioinformatics predictions using Silico tools like Polyphen, SIFT, HOPE and CADD scores. These variants were also filtered for very low minor allele frequency by comparing to the publicly available control database which contains over 5000 individuals from African populations. Out of the 207,026 variants that passed the quality check, 10 variants met the filtering criteria and were validated by Sanger sequencing. Among the 10 variants, the *ABI3BP* variant (c.53G>T p.Cys18Tyr) has the highest Combined Annotation Dependent Depletion (CADD) score of 24 meaning it is among the top 1% deleterious mutation in the human genome. Sorting Intolerance From Tolerance (SIFT) and Polymorphism Phenotyping (PolyPhen) predicted this variant to be deleterious and probably damaging, respectively. *ABI3BP* variant is predicted to be pathogenic by ClinVar and has an increased size and reduced hydrophobicity according to HOPE (Help Your Protein Explained). This gene has 2 fibronectin domains and it is involved in collagen binding and organization of the extracellular matrix. Whole exome sequencing of a triad identified a gene involved in the organization of extracellular matrix, which is associated with HGF. Identification of variants in *ABI3BP* in more families with HGF and functional study in human gingival mesenchymal stem cell lines to validate these variants is ongoing.

## PgmNr 2257: Novel mutations screening by family based whole exome sequencing in premature ovarian insufficiency patients.

### Authors:

J.Y. Park <sup>1</sup>; M.Y. Go <sup>1</sup>; J.W. Kim <sup>2</sup>; S.H. Shim <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Science, College of Life Science, CHA University, Seongnam 13488, Korea.; 2) Fertility Center of CHA Gangnam Medical Center, CHA University, Seoul 06125, Korea.

---

Premature Ovarian Insufficiency (POI) is defined as amenorrhea over 4 months before the age of 40-year and the incidence of POI is about 1~2 % of women worldwide. Autoimmune dysfunctions, metabolic disorders and genetic abnormalities such as chromosomal abnormalities or gene mutations are well-known etiologies of POI, however, most cases of POI are still idiopathic. In this study, whole exome sequencing (WES) analysis was performed in three pairs of POI patient and her mother. All patients had a normal 46, XX karyotype, and *FMR1* premutation was ruled out. To investigate causative genetic mutations, rare variants were selected with minor allele frequencies (MAF) < 0.1% based on the public population genomic data bases. After several filtering steps, one or two convincing variants were elucidated in each proband. These variants were confirmed by Sanger sequencing in the patients and 115 control women.

A heterozygous non-synonymous variant of *AKT1* (c.225C>G, p.Ile75Met) was found in proband P1, but not in her unaffected mother and control subjects. SIFT and PolyPhen-2 predicted the variant as a damaging mutation. AKT1, RAC- $\alpha$  Ser/Thr protein kinase, is involved in various of regulating processes including cell proliferation and growth via phosphorylation activity to downstream targets. In another family, a heterozygous premature stop variant of *BLM* (c.2410G>T, p.Gly804\*) was discovered in the proband P2. This novel premature stop variant was not found in her mother and controls. BLM, a bloom syndrome protein is an ATP-dependent DNA helicase that required for DNA replication and repair. A heterozygous missense variant in *NPPC* (c.131A>G, p.Gln44Arg) was identified in proband P3 and her mother with early menopause. A natriuretic peptide C (NPPC) and natriuretic peptide receptor 2 (NPR2) maintain cGMP and cAMP-mediated meiotic arrest in oocytes for ovarian reserve. The physiological roles of these newly identified genetic variants will be explored by *in-vitro* functional analysis in further studies. This work is supported by National Research Foundation of Korea (NRF) Grants from the Korean Government, No. 2019R1A6A1A03032888 and NRF-2017R1C1B5077023.



## PgmNr 2258: Polygenic effects and outlier individuals on gene expression to assess disease susceptibility.

### Authors:

A. Sartori <sup>1</sup>; A. Brown <sup>2</sup>; E.T. Dermitzakis <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) University of Dundee, School of Medicine, Dundee, UK; 3) iGE3 Institute of Genetics and Genomics of Geneva, Switzerland; 4) SIB, Swiss Institute of Bioinformatics, Geneva, Switzerland

---

It has been shown that variants that are rare in the population can have extreme effects on the expression of genes and also extreme effects on risk of disease. It has also been shown that genes can be regulated by multiple independent common variants. Similarly to how polygenic risk scores are used to predict individuals whose combinations of genotypes place them at the extreme end of the disease risk spectrum, we hypothesize that particular combinations of these common regulatory variants, representing combinations (cis haplotypes) that are rare in the general population, could also have extreme effects on expression and downstream effects on disease risk. These rare haplotypes of common variants may also show evidence of selective constraint.. Linking particular haplotypes both to extreme effects on expression and risk of disease would provide stronger evidence for mechanistic effects on disease. Our analysis focuses on individuals with extremely high or extremely low expression of a particular gene -because of the combined effect of multiple eQTLs on that gene- compared with the population. We aim to discover rare combinations of common variants with extreme effects on gene expression. We replicate the analysis in different datasets including the 49 tissues in GTEx for which eQTL data is available and UK Biobank, to link these combinations with diseases and other common traits in large GWAS studies. We further investigate if there is any evidence of selection acting on those haplotypes by implementing extended haplotype homozygosity (EHH) methodologies. By using lists of independent eQTLs from the GTEx project, we aim to find individuals whose predicted expression from genetic code is an extreme outlier. After defining haplotypes with extreme effects, we are planning to assess, using the UK Biobank dataset, if these haplotypes are associated with disease or other quantitative traits. Our preliminary results show that these effects while rare can have a profound impact on identifying individuals with increased disease risk due to cumulative effects in a few genes.



## PgmNr 2259: Polygenic risk score accuracy is dependent on local ancestry.

### Authors:

T.B. Cavazos <sup>1</sup>; J.S. Witte <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Biological and Medical Informatics, University of California, San Francisco, San Francisco, CA; 2) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 3) Department of Urology, University of California, San Francisco, San Francisco, CA

---

**Background:** Polygenic risk scores (PRS) have been developed and optimized in European populations, resulting in potentially biased and less predictive scores among non-Europeans. It is unknown whether these limitations are due to improper SNP selection or incorrect weighting, although previous studies have focused on the latter. In addition, the amount of bias may be individual-specific and dependent on the local ancestry at a given disease locus. To better understand the lack of generalizability of European PRSs across non-European populations, we will assess the relationship between bias and local ancestry at PRS variants. **Methods:** Genotypes for individuals of European (CEU) and African (YRI) ancestry were simulated. A subset of individuals were used as founders to create a new admixed population (n=4000) with RFMIX. True PRSs were computed for admixed and non-founder European individuals as described in *Martin et. al. 2017 (AJHG)* assuming 1000 causal variants and heritability of 0.67. To reflect European bias, a GWAS was conducted using 100,000 CEU individuals with 10,000 cases to identify independent GWAS significant variants and derive weights for an empirical PRS. **Results:** In our simulations we found that the Pearson's correlation between the true (m=1000) and empirical (m=820) PRSs diminished with increasing proportion of African ancestry at PRS variants. Individuals of European-only ancestry had high correlation (r=0.645) and those with low European ancestry (CEU<20%) had low correlation (r=0.432). Admixed individuals with high European ancestry at PRS variants (CEU>80%) had lower correlation (r=0.606) than European-only; however, this trend reflects that increasing noise is added into the empirical PRS with higher divergence from European ancestry at PRS variants. Utilizing this simulation framework, we will test different approaches for creating less-biased models, including effect size adjustment and SNP selection specific to local-ancestry population allele frequency and LD. **Conclusion:** Through simulations we show that the bias introduced by European GWAS when applied to non-European individuals is dependent on local ancestry and seen even when assuming identical causal SNPs across populations. This is likely a result of GWAS identifying the best European tag-SNP, which may be in lower LD with the causal variant in the non-European population. Local ancestry will be crucial to generalizing polygenic risk scores across admixed populations.

## PgmNr 2260: The genetic background of intrahepatic cholestasis of pregnancy: Findings from the FINNGEN study.

### Authors:

A.S. Havulinna<sup>1,2</sup>; T. Kiiskinen<sup>1</sup>; S. Lemmelä<sup>1,2</sup>; S. Pikkarainen<sup>3</sup>; J. Karjalainen<sup>4</sup>; M. Kurki<sup>4</sup>; E. Widen<sup>1</sup>; A. Palotie<sup>1,4</sup>; M. Daly<sup>1,4</sup>; C. Lindgren<sup>5</sup>; H. Laivuori<sup>1,6,7</sup>; FINNGEN

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland - FIMM, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Helsinki University Central Hospital, Helsinki, Finland; 4) Broad Institute of MIT and Harvard, Cambridge, MA; 5) Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; 6) Medical and Clinical Genetics, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; 7) Department of Obstetrics and Gynecology, Tampere University Hospital, and Tampere University, Faculty of Medicine and Health Technology, Tampere, Finland

---

Intrahepatic cholestasis of pregnancy (ICP, 0.4-1.5% of pregnancies) is a reversible liver disease of the late second or third trimester of pregnancy. It is characterized with pruritus, elevated serum bile acids, and elevated liver enzymes. It is linked to increased fetal risk: stillbirth, meconium passage, respiratory distress syndrome and asphyxia as well as maternal risk for gestational diabetes and pre-eclampsia. ICP has a complex aetiology with a genetic component. A handful of candidate genes, mainly involved in metabolism and transport of bile acids, have been implicated. No genome wide association studies have been published yet. Given this we set out to investigate ICP comorbidity associations, as dissecting the genetics underpinnings of disease.

The FinnGen study aims to create a dataset of ~500,000 genotyped Finns (~10% of the population) connected to electronic health records (EHRs) by unique personal identification number. Using the EHR's we constructed disease endpoints for ICP and a range of comorbidities. The current FINNGEN data freeze contains ca. 140,000 study participants. There were 581 ICP cases, among 26,343 women (2.2%) who had delivered at least once.

Both prevalent autoimmune disorders (Fisher test  $p=0.00015$ ), and diabetic pregnancy ( $p=5.9e-6$ ) were associated with ICP, while we saw no evidence of association to malignancies, cardiovascular diseases or pre-eclampsia ( $p>0.05$ ).

Next, we explored whether ICP predisposed to later morbidity (age adjusted Cox regression). The set of morbidities was in line with those in the J Hepat article by Wikstöm Shemer et al (2015), including malignancies, autoimmune diseases, cardiovascular diseases and pregnancy-related outcomes. Among ICP patients the following outcomes were more frequent: Crohn's disease ( $p=0.012$ ), polyarthropathies ( $p=8.0e-4$ ), adverse pregnancy outcomes ( $p=0.011$ ), including preeclampsia ( $p=6.5e-10$ ).

Finally, we performed a PheWAS scan for the FINNGEN participants. ICP was associated with regions in several liver transporter genes. Some of these regions were also associated with cholelithiasis.

In conclusion, ICP has a strong genetic background, and it is associated to autoimmune disorders and adverse pregnancy outcomes. We are currently expanding our analysis with >2,000 cases from an additional cohort. The understanding of the genetic underpinnings of this disease will lend itself to a greater understanding of the mechanisms responsible for the disease susceptibility.

## **PgmNr 2261: A genome-wide association study on landmark based phenotypes identifies novel loci for facial variations in East Asian populations.**

### **Authors:**

S. Du; S. Wu; M. Zhang

[View Session](#) [Add to Schedule](#)

**Affiliation:** CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai, China

---

Numerous studies have shown evidence of a genetic basis for facial morphology. Previous studies based on 3D facial surface images have identified a number of genetic loci associated with facial quantitative traits. However, most studies on facial morphology were performed on Caucasian and African populations, while few studies were based on East Asian populations. In this study based on 19 anatomical landmarks automatically placed on 3D facial surface meshes of 6,132 individuals of East-Asian ancestry, we calculated 192 linear distances between each two landmarks as traits. We then carried out a genome-wide association study of linear distances and observed genome-wide significant associations ( $p < 5 \times 10^{-8}$ ) for traits related to nose and eyes with six genetic loci including previously reported signals at 1p36.32, 2q36.1, 4q31.3 and 17q24.3, as well as novel signals at 3p25.2, 6p12.1 and 15q23. By visualizing the morphological effects of the associated loci at the local and global levels, we could observe that the loci associated with linear distances in nose and eyes also influenced the global facial morphology. These results provided a more comprehensive understanding of the relation between genetic loci and human facial morphology.

## **PgmNr 2262: SNPs in *PDE3A*, *OLFM4*, and *CYP1A2* were associated with caffeine consumption in a genome-wide association study of the Coriell Personalized Medicine Collaborative.**

### **Authors:**

D. Kusic; L. Scheinfeldt; S. Zajic

[View Session](#) [Add to Schedule](#)

**Affiliation:** Coriell Institute for Medical Research, Camden, NJ.

---

More than 90% of Americans consume caffeine daily, but the amount consumed varies widely between individuals. Previous studies have shown that genes related to caffeine metabolism and receptor binding may influence caffeine consumption. Better understanding the relationship between genotype and caffeine consumption has the potential to inform potential mechanisms of the cardioprotective effects of caffeine, as well as its role in headache including migraine.

We performed a genome-wide association study (GWAS) of 2,783 participants in the Coriell Personalized Medicine Collaborative for the phenotype of self-reported caffeine consumption in drinks per week using a generalized linear model using age, cohort, gender, and principal components to account for population structure. Participants reported the number of caffeinated drinks they consumed in a typical week at baseline.

Our analysis identified three genetic loci associated with caffeine consumption near the Bonferroni-corrected alpha significance threshold. One locus, rs12580173 ( $p\text{-value}=2.8\times10^{-7}$ ), is located in *PDE3A*, the gene for phosphodiesterase (PDE) 3A. PDE3A is one of several PDEs inhibited by caffeine that catalyze the elimination of cyclic adenosine monophosphate (cAMP) and has been linked with cardiovascular disease and migraine. The second locus, rs11620007 ( $p\text{-value}=8.2\times10^{-7}$ ), is near *OLFM4*, which is reported to play roles in metabolic function, pancreatic cell proliferation, and cell adhesion. The third locus, rs2472297 ( $p\text{-value}=2.2\times10^{-6}$ ), upstream of the gene for CYP1A2 (the enzyme primarily responsible for caffeine metabolism), has been associated with caffeine consumption in prior studies. Though these loci did not meet the threshold for genome-wide significance ( $p\text{-value}=5\times10^{-8}$ ) in this relatively small GWAS, the biological relevance of CYP1A2 and PDE3A to caffeine metabolism and pharmacodynamics lend plausibility to the associations.

These findings add to a growing body of literature characterizing the influence of genetics on caffeine consumption that has the potential to inform lifestyle choices (i.e. consumption behavior), as well as associations with health outcomes. These results suggest that further research into the role of PDEs such as PDE3A in mediating the effects of caffeine on cardiovascular health and headache may be warranted. Such work has the potential to eventually inform personalized interventions aimed at improving cardiovascular, metabolic, and quality of life outcomes.

## PgmNr 2263: Genetically mediated variation in one carbon metabolism is associated with the development of edematous severe acute malnutrition.

### Authors:

N.C. Lie <sup>1,2</sup>; K.V. Schulze <sup>2,3</sup>; S. Swaminathan <sup>2,3</sup>; S. Howell <sup>4</sup>; O. Brown <sup>4</sup>; R. Sadat <sup>2</sup>; N. Hall <sup>2</sup>; K. Marshall <sup>4</sup>; A. Choudhury <sup>5</sup>; D. Sengupta <sup>5</sup>; M.E. Reid <sup>4</sup>; C. Taylor-Bryan <sup>4</sup>; X. Wang <sup>2,3</sup>; J.W. Belmont <sup>3</sup>; Y. Guan <sup>6</sup>; M.J. Manary <sup>7,8</sup>; I. Trehan <sup>7,8</sup>; C.A. McKenzie <sup>9</sup>; N.A. Hanchard <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Integrative Program in Molecular and Biomedical Sciences, Baylor College of Medicine, Houston, TX, USA; 2) USDA/ARS/Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 4) Tropical Metabolism Research Unit, Caribbean Institute for Health Research, University of the West Indies, Mona, Jamaica; 5) Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, 2193, South Africa; 6) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA; 7) 6Departments of Paediatrics and Child Health and Community Health, University of Malawi, Blantyre, Malawi; 8) Department of Pediatrics, Washington University in St. Louis, St. Louis, MO, USA; 9) Posthumous

Severe acute malnutrition (SAM), which affects 16.9 million children globally and contributes to the death of 3 million annually, occurs as either edematous SAM (ESAM), characterized by swelling and extensive organ failure and seen most commonly in Central Africa and the Caribbean, or non-edematous SAM (NESAM), characterized by extensive wasting and seen most commonly elsewhere. The diet and environment of SAM-affected children is similar regardless of sub-type, so the reasons why some children develop ESAM while others develop NESAM remain unclear. However, we have previously shown that ESAM is associated with significant hypomethylation of DNA at metabolic and nutritional gene loci. This is consistent with a growing body of work that suggests abnormal one-carbon metabolism (OCM) is a key factor in the development of ESAM. As genetic variation is known to significantly impact OCM, we sought to determine whether variants in genes regulating OCM contribute to the development of ESAM versus NESAM.

Using samples from 711 Jamaican and Malawian children, evenly split between those with ESAM or NESAM, we assayed 1.8 million SNPs using the H3Africa SNP microarray. Logistic regression coupled with hypergeometric sampling uncovered evidence of a cumulative significance of OCM-related SNPs associated with ESAM ( $z = 3.058$ ). This was partially driven by SNPs in *GABBR2*, a gene associated with OCM homocysteine levels, that seem to confer protection from ESAM ( $P_{\text{locus}} = 2.18 \times 10^{-5}$ , odds ratio = 0.5659). Logistic regression in Jamaica and Malawi samples separately also uncovered strong evidence of association between SNPs in *PRICKLE2*, a modulator of OCM folate concentration, and ESAM in Malawi that was attenuated in Jamaica ( $P_{\text{locus}} < 5 \times 10^{-4}$ ). Admixture analyses identified differing proportions of ancestral Bantu-speaking components in Jamaica and Malawi, suggesting that differences in local ancestry may contribute to differing strengths of disease association and geographical patterns of disease prevalence.

Our findings provide the first molecular evidence for genetic variation as a mediator of clinical outcomes in severe nutritional stress, and emphasize the importance of integrating local ancestry and

population genetics with disease studies to elucidate associations and understand geographical variation in disease. Work to leverage local ancestry and haplotype analyses to further clarify these associations and to integrate previously identified methylation QTLs is ongoing.

## PgmNr 2264: First genome-wide analysis of developmental stuttering identifies genome-wide significant signal near *EFNA5*.

### Authors:

H. Polikowsky <sup>1</sup>; L.E. Petty <sup>1</sup>; D.M. Shaw <sup>1</sup>; D.G. Pruett <sup>2</sup>; S.J. Kraft <sup>3</sup>; J.E. Below <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 2) Hearing and Speech Sciences, Vanderbilt University, Nashville, TN; 3) Behavioral Speech & Genetics Lab, Department of Communication Sciences and Disorders, Wayne State University, Detroit, MI

---

Developmental stuttering is an often debilitating speech disorder characterized by prolongation of sounds and interruptions in speech with childhood onset. It can negatively affect job performance and employability in adults and frequently results in decreased classroom participation and increased bullying in children. Persistent stuttering is also very common, with a prevalence of 1% of the general adult population. Moreover, despite large heritability estimates and strong familial trends, the genetic architecture and etiology of developmental stuttering remain elusive, in part due to the challenges associated with amassing a large enough cohort for well-powered population-based analyses. To date, no genome-wide association study of developmental stuttering has been published.

Here, through collaborative efforts and an innovative social media campaign, we genotyped 809 developmental stuttering cases on the Illumina multi-ethnic genotyping array (MEGA) measuring 1,842,793 variants. Approximately 3000 ancestry matched randomly selected population-based controls were drawn from Vanderbilt's identically genotyped and called DNA biobank, BioVU. Using logistic regression in an unrelated, Caucasian subset of cases (n=589) and controlling for the first three principal components we analyzed 790,192 common variants measured in both case and control data sets, and identified one variant reaching genome-wide significance, rs17159693 (p-value  $4.302e^{-08}$ , OR 1.692) and 22 variants with suggestive significance (p-values less than  $5e^{-06}$ ). For variant rs17159693, developmental stuttering cases had a minor allele frequency (MAF) of about 16%; whereas, controls had a MAF of about 10%, aligning with reported MAF in European ancestry populations (per dbSNP). Validation analyses include cross-ancestry analysis in an additional 800 developmental stuttering cases, fine mapping, and rare variant analysis in 1500 whole exomes.

Interestingly, variant rs17159693, is physically adjacent to *EFNA5*, a member of the ephrin gene family implicated in mediating developmental events, particularly in the nervous system. These intriguing preliminary results are the first to implicate genes that regulate nervous system development in the etiology of developmental stuttering.



# PgmNr 2265: Functional haplotypes in the *ADIPOQ* gene are associated with underweight, immunosuppression and viral suppression in Kenyan HIV-1 infected injection substance users.

## Authors:

N. Shaviya <sup>1</sup>; V. Budambula <sup>2</sup>; T. Were <sup>3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Medical Laboratory Sciences, Masinde Muliro University of Science and Technology, Kakamega, Kakamega, Kenya; 2) Environmental Health, Technical University of Mombasa, Kenya; 3) Center for Global Health and Research, Kenya Medical Research Institute, Kisumu, Kenya

---

## Background

Human immunodeficiency virus (HIV) and injection substance use have both direct and epigenetic effects on genes. These effects could be beneficial or detrimental in defining disease outcomes. *ADIPOQ* gene is key in modulating metabolic and immunoregulatory functions. Understanding the effects of HIV and injection substance use on the gene in the context of antiretroviral therapy is important for predicting disease outcomes.

## Methods

This cross-sectional genetic study determined polymorphisms in the promoter region of the *ADIPOQ* gene. Two loci were analyzed rs2241766 and rs266729. The polymorphisms were associated with clinical markers of disease outcome; underweight, immunosuppression and viral suppression. The selected variants were amplified via PCR then genotyped via random fragment length polymorphism.

## Results

GC haplotype associated with higher odds of having underweight (OR, 2.21; 95% CI, 1.83-4.60;  $P=0.008$  vs. OR, 2.30; 95% CI, 1.89-4.71;  $P=0.006$ ) in ART-naïve and experienced ISUs and immunosuppression (OR, 1.90; 95% CI 1.67-3.98,  $P=0.041$ ) in ART-naïve while TC haplotype associated with lower odds of having underweight (OR, 0.49; 95% CI, 0.31-0.74;  $P=0.019$  vs. OR, 0.43; 95% CI, 0.19-0.89;  $P=0.028$ ) in both ART-naïve and -experienced ISUs and immunosuppression (OR, 0.42; 95% CI 0.16-0.81,  $P=0.023$ ) in ART-naïve ISUs. Circulating total adiponectin levels were higher in ART-naïve (median, 19.5; IQR, 7.9 µg/ml) than -experienced (median, 12.0; IQR, 4.4 µg/ml) ISUs ( $P<0.0001$ ). Underweight and immunosuppressed ISUs presented with lower serum adiponectin levels while virally suppressed participants had higher serum adiponectin levels. GC carriers presented with low while TC carriers presented with higher adiponectin levels in both ART-naïve and -experienced ISUs.

## Conclusion

The study revealed haplotypes of the *ADIPOQ* gene at loci rs2241766 and rs266729 that could determine disease outcomes in HIV-1 ART naïve and experienced ISUs.

## PgmNr 2266: Elucidation of the causative gene of non-obstructive azoospermia by whole-exome sequencing.

### Authors:

Y. Sato <sup>1</sup>; C. Fukunaga <sup>1</sup>; K. Kojo <sup>2,3</sup>; M. Uchida <sup>2,4</sup>; H. Tsuchiya <sup>2</sup>; K. Yamasaki <sup>2,4</sup>; T. Iwamoto <sup>2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pharmaceutical Information Sciences, Biomedical Sciences Tokushima Graduate School of Biomedical Sciences, Tokushima University University Graduate School, Tokushima, Tokushima, Japan; 2) Center for Infertility and IVF, International University of Health and Welfare Hospital; 3) Department of Urology, University of Tsukuba Hospital; 4) Urology department, Tsukuba Gakuen Hospital; 5) Department of Male Infertility, Reproduction Center, Sanno Hospital

---

**BACKGROUND:** About 10% of male infertility patients do not have sperms in their semen and are diagnosed with azoospermia. The condition without obstruction of the post testicular genital tracts and no sperm production is called non-obstructive azoospermia (NOA). The causes of NOA, excluding Klinefelter syndrome, includes deletion of the Y chromosome azoospermia factor (AZF) region. A diagnostic test for detection of the AZF region is useful to predict sperm recovery by microdissection testicular sperm extraction (MD-TESE). About 10% of NOA patients have a deletion of the AZF region; however, other causes remain to be elucidated. This study aimed to clarify the pathogenesis and condition of individual NOA patients by genetic analysis using next-generation sequence analysis. **METHODS:** Five men diagnosed with NOA at the International University of Health and Welfare Hospital and who had undergone MD-TESE were recruited in this study. After extracting the DNA, exome sequencing was performed, and genes that affected protein function and with an allele frequency of 1% or less were narrowed down. Subsequently, we searched relevant literature for relationship between genes and spermatogenesis.

**RESULTS and DISCUSSIONS:** The BRWD1, HFM1 and NLRP14 genes, which were narrowed down in three patients in whom sperm could not be recovered with MD-TESE, were reported to be involved in spermatogenesis. Therefore, we speculate that variants in these genes may have caused spermatogenesis failure in these three patients. On the contrary, no gene reported to be associated with spermatogenesis was found in the two subjects whose sperms could be recovered by MD-TESE; this may be due to unexplained genetic or other environmental factors. These results indicate that whole-exome sequencing would enable us to understand the causes of NOA. Furthermore, because genes that are associated with spermatogenesis were found only in patients whose sperm could not be recovered by MD-TESE, whole-exome sequencing for NOA patients undergoing MD-TESE may lead to the prediction of sperm recovery.

## PgmNr 2267: Genome wide association analysis suggests three novel genetic associations with specific language impairment in a large DNA biobank.

### Authors:

D. Shaw <sup>1</sup>; L. Petty <sup>1</sup>; H. Chen <sup>1</sup>; H. Polikowsky <sup>1</sup>; R. Nitin <sup>3</sup>; C. Walters <sup>3</sup>; R. Gordon <sup>2,3</sup>; J. Below <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN.; 2) Department of Otolaryngology, Vanderbilt University Medical Center; 3) Vanderbilt Brain Institute, Vanderbilt University

Specific language impairment (SLI) is a developmental language disorder in which children will often have reduced or absent language capabilities during early childhood. In addition to late onset of speaking skills, children with SLI often have difficulty mastering basic grammar patterns and utilizing proper tenses. Children with SLI are phenotypically distinguishable from patients whose speaking capabilities are directly affected by alternative correlative phenotypes including hearing loss, autism, and other developmental disorders. This diagnosis is relatively common for children in their kindergarten years, having a prevalence of approximately 7% within school aged children in the US. SLI profiles within patients can vary, and patients are often stratified into those with expressive language delays, receptive language delays, or both.

Early twin studies demonstrated moderate heritability for SLI, with estimates ranging from 0.5 to 0.75 for school aged children. Despite this strong genetic component, the genetic architecture of SLI is still largely unknown. Previous linkage analyses have demonstrated an association between abnormalities in the FOXP2 gene and SLI, providing a potential causative pathway. However, most patients with SLI have no abnormalities in FOXP2, suggesting undiscovered genetic components of SLI remain.

We utilized BioVU, a large DNA biobank at Vanderbilt University Medical Center linked to electronic health records (EHR) and deployed a novel EHR-based SLI phenotyping algorithm to identify likely cases of SLI. We matched probable cases to controls based on genetic ancestry principal components, including in our preliminary analysis 247 cases and a total sample size of 1,235. We ran an association analysis using logistic regression, controlling for three principal components, in order to identify variants associated with specific language impairment.

Our top hits following the regression include variants rs7642482 (OR=1.821;  $\beta$  = 4.694; p=2.679E-6) and rs1692421 (OR=4.646;  $\beta$  = 1.821; p=3.391E-6) which flank *ROBO2* (19kb 3', chr5) and *MAP1B* (83kb 5', chr5) respectively. Each of these genes are associated with neuronal extension/guidance. The analysis additionally suggests an association with the variant rs35938541 (OR=5.084;  $\beta$  = 2.141; p=3.368E-7), found within the first intron of the KIF3A gene. This analysis will be extended to an additional 1,322 SLI cases in BioVU.

## PgmNr 2268: A genome-wide association study of the gut microbiome in Africans.

### Authors:

A.A. Adeyemo<sup>1</sup>; A.P. Doumatey<sup>1</sup>; A.R. Bentley<sup>1</sup>; J. Zhou<sup>1</sup>; S. Adebamowo<sup>2</sup>; O. Olusola-Taiwo<sup>3</sup>; C. Adebamowo<sup>4</sup>; C.N. Rotimi<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) University of Maryland School of Medicine, Baltimore, MD; 3) Ring Road State Hospital, Ibadan, Nigeria; 4) Department of Epidemiology and Public Health; Institute of Human Virology; Greenebaum Comprehensive Cancer Center, School of Medicine, University of Maryland, Baltimore, MD

---

Human genetic variation is one of several factors that influence the gut microbiome. However, genetic association studies of the gut microbiome have not yet been done in African populations. In the present study, we conduct a genome-wide association study (GWAS) of gut microbiome composition in 291 adult continental Africans (224 women, 67 men; 56.1 (SD 12.6) years; mean BMI 31.0 (SD 6.1) kg/m<sup>2</sup>) enrolled from Nigeria. Gut microbiome profiling was done using fecal 16S rRNA V4 sequencing followed by identification, filtering and classification of operational taxonomy units (OTU). Genotypes were obtained using the Illumina Multi-Ethnic Genotyping Array (MEGA) followed by imputation into the African Genome Resources Reference Panel. Gut microbiome characteristics analyzed included measures of alpha diversity (Shannon Diversity and OTU richness), beta diversity (Bray-Curtis Dissimilarity) and relative abundance (RA) of selected OTU. All analyses were adjusted for age, sex, diabetes status, antibiotic use, metformin use and principal components of the genotypes. The most abundant OTU at the phylum level were Firmicutes, Actinobacteria, and Bacteroidetes. There was a single genome-wide significant (GWS) locus for beta-diversity (lead SNP rs12341440, MAF=0.093,  $p=1.66 \times 10^{-8}$ ), near *MED27*. None of the two alpha diversity indices had a GWS variant (best SNP for OTU richness rs35347638  $p=1.30 \times 10^{-7}$ ; Shannon Diversity chr3:138998067,  $p=2.46 \times 10^{-7}$ ). Each OTU displayed a unique pattern of genetic association. Proteobacteria had two GWS associations (rs77476041 near *MTUS2*,  $p=3.28 \times 10^{-8}$ ; rs112957979 in *DNAH7*,  $p=4.24 \times 10^{-8}$ ) while Akkermansia and Bacteroidetes each had a single GWAS locus at rs59849632 (*RP11-544L8\_B.4*,  $p=1.69 \times 10^{-8}$ ) and rs116703580 (chromosome 12;  $p=2.26 \times 10^{-7}$ ), respectively. Notably, Bifidobacterium showed GWS in *AMPD3* (lead SNP rs6484218, MAF=0.409,  $p=4.66 \times 10^{-8}$ ); in previous studies, this SNP was significantly associated with depression, bipolar disorder and schizophrenia. The phyla Firmicutes and Blautia showed only suggestive association ( $5 \times 10^{-7} > p > 5 \times 10^{-8}$ ). Most of the GWS SNPs in the study are eQTL and/or have regulatory annotation (alter motifs or are in promoter or enhancer histone marks). This first gut microbiome GWAS in Africans has identified distinct patterns of association with overall composition and specified taxa. Our understanding of microbiome composition would be advanced by studies that model both genetic and non-genetic effects.

## PgmNr 2269: Genetics of gut microbiome composition: Results of MiBioGen consortium.

### Authors:

A. Zhernakova; on behalf of MiBioGen consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics, University Medical Center Groningen, The Netherlands

---

Gut microbiome composition is influenced by environment and host factors, including host genetics. Several population-based genome-wide association studies (GWAS) have identified dozens of SNPs influencing microbiome composition and function, however, the overlap in findings from different cohorts is limited, suggesting that current studies are underpowered and/or subject to population-specific effects. To address both problems, we established the MiBioGen consortium to perform multi-cohort microbiome GWAS. The consortium includes >18,000 samples from 23 cohorts from the USA, Canada, Israel, South Korea and seven European countries, with both microbiome (16S rRNA gene sequencing) and genome-wide SNP data imputed using the Haplotype Reference Consortium reference panel. Processing of genetic and microbiome data were performed following standardized procedures developed within the consortium, and then meta-analysed using individuals summary statistics data.

We performed the genetic analysis in relation to three microbial traits: alpha-diversity of gut microbiome, abundance of bacteria (quantitative analysis) and presence of bacteria (binary analysis). At genome-wide significance level, we identified 40 QTLs that affect presence or abundance of specific microbial taxa (mbQTLs), 38 of them novel. The two established mbQTLs include the lactose intolerance locus (*LCT*) affecting the abundance of *Bifidobacterium* and the *FUT2* locus that affects abundance of two genera from the Lachnospiraceae family).

We explored the potential relevance of microbiome-affecting loci to host health by testing for causality between host genetics, microbiome composition, diseases and disease-related phenotypes using PheWAS and Mendelian Randomization (MR) approaches. We focused on the genetic overlap of mbQTLs with non-communicable diseases, metabolomics, nutritional preferences and immunological markers. PheWAS analysis indicated an overlap of mbQTLs with genetic variants that influence metabolic, nutritional and immune phenotypes. Using MR we were able to causally link serum metabolites (68 associations passed 5% FDR), nutrition phenotypes (18 signals passed 5% FDR) and systemic non-infectious diseases to microbial taxa abundance. In turn, microbial taxa act as mediators of genetic predisposition to ulcerative colitis.

Overall, our results provide broader insight into the effect of host genetics on the microbiome, and consequently on human health and disease.

## PgmNr 2270: eMERGE phenome-wide association study of biogeographic ancestries predicts ocular, immune system, renal, cardiometabolic, gynecological, and vector-borne disease risk.

### Authors:

J.M. Keaton<sup>1,2,3,4</sup>; J.N. Hellwege<sup>2,5</sup>; E.S. Torstenson<sup>1,2</sup>; K. Actkins<sup>2,6</sup>; R. Knevel<sup>7,8,9</sup>; S. Asgari<sup>10</sup>; G. Jarvik<sup>11</sup>; N. Shang<sup>12</sup>; S. Raychaudhuri<sup>13</sup>; L. Davis<sup>2,5</sup>; J.C. Denny<sup>2,4,14</sup>; D.M. Roden<sup>2,4,14,15</sup>; T.L. Edwards<sup>1,2,3</sup>; D.R. Velez Edwards<sup>2,3,4,16</sup>; on behalf of the eMERGE Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 2) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA; 3) Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, TN, USA; 4) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, USA; 5) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 6) Department of Microbiology, Immunology, and Physiology, Meharry Medical College, Nashville, TN, USA; 7) Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands; 8) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 9) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 10) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 11) Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA, USA; 12) Department of Biomedical Informatics, Columbia University, New York, NY, USA; 13) Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA; 14) Department of Medicine, Vanderbilt University, Nashville, TN, USA; 15) Department of Pharmacology, Vanderbilt University, Nashville, TN, USA; 16) Division of Quantitative Sciences, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA

---

Racial disparities may arise in part due to phenomena that increase the frequency of trait-increasing alleles in one geographic parental subpopulation relative to another. In admixed offspring populations, those differences manifest as association between proportions of genetically inferred ancestry and traits. We therefore estimated six geographic genetic ancestry proportions based on 1000 Genomes reference populations in 60,267 non-Hispanic white (NHW) and 10,168 non-Hispanic black (NHB) participants from the electronic MEdical Records and GENomics (eMERGE) Network and performed a phenome-wide association study (PheWAS) searching for ancestry-dependent phenotypes. Associations of ancestry proportion, ancestry-body mass index (BMI) interactions, and ancestry-ancestry interactions with PheWAS outcomes were adjusted for age, sex, and BMI. We identified 509 phenotypes significantly associated ( $p < 2.75 \times 10^{-5}$ ) with at least one ancestry. Among NHW participants, Northern European (NEUR) and Southern European (SEUR) ancestries were associated with eye disease (e.g. presbyopia;  $p_{\text{NEUR}} = 9.67 \times 10^{-276}$ ,  $p_{\text{SEUR}} = 2.20 \times 10^{-207}$ ) and immune system disorders ( $p_{\text{NEUR}} = 2.80 \times 10^{-172}$ ,  $p_{\text{SEUR}} = 3.29 \times 10^{-153}$ ). Among NHB participants, East African (EAfr) and West African (WAfr) ancestries were associated with rosacea ( $p_{\text{EAfr}} = 6.73 \times 10^{-6}$ ), stage III chronic kidney

disease (CKD;  $p_{\text{WAFR}}=2.51 \times 10^{-6}$ ), diabetic kidney disease ( $p_{\text{WAFR}}=1.27 \times 10^{-5}$ ), and hypertensive kidney disease ( $p_{\text{WAFR}}=4.29 \times 10^{-6}$ ). Protective effects for WAFR ancestry were observed for pregnancy complications among NHB women (e.g. malposition and malpresentation of fetus or obstruction;  $p_{\text{WAFR}}=9.37 \times 10^{-8}$ ), which is consistent with a model of increasing admixture promoting reproductive success. Significant ancestry-BMI interactions were observed for psychiatric (e.g. suicide;  $p_{\text{NEUR\_BMI}}=5.03 \times 10^{-11}$ ;  $p_{\text{SEUR\_BMI}}=1.08 \times 10^{-8}$ ), respiratory (e.g. pulmonary congestion;  $p_{\text{NEUR\_BMI}}=1.07 \times 10^{-6}$ ;  $p_{\text{SEUR\_BMI}}=4.67 \times 10^{-6}$ ), and infectious (e.g. tuberculosis;  $p_{\text{WAFR\_BMI}}=7.82 \times 10^{-6}$ ) diseases. Significant ancestry-ancestry interactions were observed for infectious (e.g. arthropod-borne disease;  $p_{\text{WAFR\_NEUR}}=3.08 \times 10^{-7}$ ), metabolic (e.g. mixed hyperlipidemia;  $p_{\text{WAFR\_NEUR}}=1.16 \times 10^{-6}$ ;  $p_{\text{EAFR\_NEUR}}=2.97 \times 10^{-5}$ ), hematopoietic (e.g. hereditary hemolytic anemias;  $p_{\text{WAFR\_NEUR}}=1.64 \times 10^{-5}$ ), and renal diseases (e.g. stage IV CKD;  $p_{\text{WAFR\_SEUR}}=1.24 \times 10^{-5}$ ). These results demonstrate the ability of geographic genetic origin to predict many types of disease risk.



# PgmNr 2271: Using phenome-wide association study results to predict Mendelian phenotypes.

**Authors:**

L. Tang; Q. Chen; J. Denny; L. Bastarache

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt University Medical Center, Nashville, TN.

---

**Background:**

Mendelian and complex diseases have traditionally occupied separate spaces. However, common variants with minor allele frequencies (MAFs)  $>1\%$  in and around Mendelian genes, or Mendelian-matched variants (MMVs), have been linked to complex diseases that can share similar clinical phenotypes to their linked Mendelian disease. We tested the efficacy of using the phenome-wide association studies (PheWAS) to elucidate the phenotype patterns of Mendelian disease. We hypothesized that MMVs were more likely to be associated with corresponding Mendelian disease features than non-Mendelian-matched variants (nMMVs).

**Methods:**

We obtained clinical features and their linked genes of Mendelian diseases from Online Mendelian Inheritance in Man (OMIM). We performed PheWAS on 72,000 individuals of European ancestry, adjusted for age, sex, and principal components for MMV and nMMVs. We analyzed associations between 978,030 MMV-phecode pairs (86,271 SNPs from 3,409 genes and 853 phecodes). We randomly sampled nMMVs from 10,000 randomly-selected non-Mendelian genes to obtain a 1:1 MMV-to-nMMV ratio for each Mendelian phecode. We used logistic regression to measure the enrichment of significant PheWAS results (i.e. PheWAS results with  $p < 1e-5$ , the same cutoff for significance that is used in the GWAS Catalog) for MMVs vs. nMMVs. Among MMVs, we also evaluated the significance of disease inheritance pattern (autosomal dominant vs. recessive), SNP function (e.g. intronic, exonic), and affected biological systems as predictors for significant PheWAS results. We adjusted for the number of SNPs tested in each gene.

**Results:**

Globally, the odds of a significant SNP-Mendelian phecode result were 56 times greater for MMVs vs. nMMVs ( $p = 2.97e-46$ ). Among MMVs, inheritance pattern was not a significant predictor ( $p = 0.49$ ). The odds of a significant result were 15 times greater for intronic vs. intergenic SNPs ( $p = 8.47e-3$ ). Compared to connective tissue phecodes, the odds of a significant result were 5 times greater for the endocrine system ( $p = 9.95e-5$ ).

**Conclusion:**

We explored the utility of using PheWAS data on common variants to predict the effect of rare, potentially deleterious variants. Such data can be useful for interpreting genetic testing results that can reveal deleterious variants in genes not linked to Mendelian diseases. Our findings support the use of association data from common variants to predict the Mendelian phenotype.



## PgmNr 2272: The status of vitamin D and its receptor *Fok I* gene polymorphism among adult TB patients of Mulago Hospital, Kampala, Uganda.

### Authors:

E. Acen <sup>1</sup>; W. Worodria <sup>3,4</sup>; P. Mulamba <sup>5</sup>; I. Andia <sup>6</sup>; D. Kateete <sup>7</sup>; M. Joloba <sup>7</sup>; J. Erume <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MICROBIOLOGY, MAKERERE UNIVERSITY, KAMPALA, P.O. Box 7072 Kampala, Uganda; 2) 2Makerere university College of Veterinary Medicine, Animal Resources and Bio-Security, Department of Biomolecular Resources and Biolab Sciences, P.O Box 7062, Kampala Uganda.; 3) Pulmonary Unit, Mulago National Referral and Teaching Hospital, P.O.Box 7051. Kampala Uganda; 4) Makerere University College of Health Sciences, Department of Medicine, P.O. Box 7072 Kampala Uganda; 5) 5Makerere University College of Agricultural and Environmental Science, Department of Agricultural and Biosystem Engineering, P.O.Box 7062 Kampala Uganda; 6) 6 Makerere College of health sciences, Infectious diseases division department of Internal medicine, P.O. Box 7072 Kampala Uganda; 7) Makerere College of health sciences, Immunology and Molecular Biology department. P.O. Box 7072 Kampala Uganda

---

*Tuberculosis* (TB) is an infectious disease causing severe problems throughout the world. However this disease is prevalent in Africa. The antimicrobial activity of vitamin D may plays a significant role in prevention of TB. The vitamin D receptor gene is associated with susceptibility. The functional Single Nucleotide Polymorphism rs2228570, *Fok I* gene has been found to be inconsistent among the TB patients. Additionally, patients with active TB have significantly lower vitamin D serum concentration than their contacts from the same ethnicity. This study investigated the relationship between vitamin D status and Vitamin D receptor *Fok I* gene polymorphism, in TB patients. A pilot case-control study was conducted in 41 newly diagnosed Tuberculosis patients and 41 non-TB healthy workers enrolled between April and June, 2013 at Mulago National Referral Hospital. Levels of Vitamin D and PTH were analyzed by Electrochemiluminescence using Cobas 6000. Genotyping of *Fok I* gene was done by Polymerase Chain Reaction and direct sequencing method using the ABI Sequencer. The Prevalence of hypovitaminosis in the TB patients was 24.4% and 26.8% in the healthy subjects. Vitamin D deficiency in TB patients was 9.7% and 4.9 in healthy subjects respectively. Severe vitamin D deficiency was only found in the TB patients. Subjects with vitamin D deficiency were only found in the FF genotype. There was no significant difference between vitamin D levels among the *Fok I* gene variants ( $p = 0.78$ ). Although there was no significant difference between vitamin D status in the different genotypes of Vitamin D receptor *Fok I* gene polymorphism among TB patients and controls in this Ugandan cohort, the hypovitaminosis D noted dominantly in the FF genotype should not be under estimated. On the other hand optimal levels of vitamin D were predominantly found in both groups.

## PgmNr 2273: Using genome-wide SNPs to estimate components of heritability for 24 quantitative and dichotomous psychological characteristics in 14,675 Han Chinese.

### Authors:

L. Wang<sup>1</sup>; Y. Chen<sup>2</sup>; X. Yao<sup>1</sup>; Q. Zhang<sup>1</sup>; D. Jiang<sup>2</sup>; S. Tang<sup>1</sup>; T. Li<sup>1</sup>; H. Weng<sup>1</sup>; Y. Li<sup>1</sup>; X. Lin<sup>1</sup>; X. Wu<sup>1</sup>; Y. Gan<sup>2</sup>; G. Chen<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) WeGene, Shenzhen, China; 2) Department of Psychology and Cognitive Science, Peking University, Beijing, China

---

Perceived stress characteristics, resilience, future orientation and healthy behaviors have been major research topics over the last decades. However, genetic and environmental factors and their interrelationships remain unclear and have not been well conducted among Han Chinese populations. In the present study, we made use of psychological phenotype data fulfilled by 14,675 Han Chinese participants, who had been genotyped by ~700,000 SNPs on microarrays as part of the WeGene Personal Genome Services. Ten psychological scales and demographic questions were used to define 24 quantitative and dichotomous psychological traits. We firstly validated 71 SNPs reported being associated with perceived stress, future orientation and healthy behaviors discovered by candidate-gene approaches, only 4 SNPs can be verified ( $P$ -value  $< 0.05$ ). We further conducted genome-wide association studies (GWAS) using these traits and only 21 loci can reach genome-wide statistical significance ( $P$ -value  $< 5e-8$ ) in 7 phenotypes defined by Childhood Trauma Questionnaire, Future-oriented Coping Inventory, Connor-Davidson Resilience Scale, Work Stress Scale and Maslach Burnout Inventory. We also obtained heritability estimates using array and imputed data in unrelated individuals. The estimated heritability are around 5~12% and 12~18% of the variance in quantitative and dichotomous traits respectively. These scales of heritability and sample size should obtain a sufficient GWAS power to detect genotype-phenotype associations. We propose that the majority of heritability is attributable to genome-wide variants of small effect and variance explained by individual effects are too small to pass stringent significance tests. We verified our hypothesis by further partitioning additive variance onto chromosomes. We found that the variance explained by each chromosome is proportional to its length (17 of 24 traits with  $P$ -value  $< 0.05$ , Pearson's product-moment correlation; the most significant phenotypes is defined by Maslach Burnout Inventory-achievement). As far as we are aware of, this is the first study estimating the shared and unique heritability of psychological characteristics in Han Chinese, which provide insights for future experimental designs to map additional trait loci and to understand the association between gene and environments. According to our findings, larger GWASs will be needed to further detect casual SNPs for these psychological traits.

## PgmNr 2274: Exome-wide low-frequency genetic variants contribute to human craniofacial morphology.

### Authors:

D. Liu <sup>1</sup>; N. Alhazmi <sup>2</sup>; M.K. Lee <sup>3</sup>; J. Li <sup>4</sup>; J.T. Hecht <sup>5</sup>; G.L. Wehby <sup>6</sup>; L.M. Moreno <sup>7</sup>; C.L. Heike <sup>8</sup>; J. Roosenboom <sup>3</sup>; E. Feingold <sup>1</sup>; M.L. Marazita <sup>3</sup>; C. Claes <sup>4</sup>; E.C. Liao <sup>9</sup>; S.M. Weinberg <sup>3</sup>; J.R. Shaffer <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Oral Biology, Harvard School of Dental Medicine, Boston, MA; 3) Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA; 4) Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium; 5) Department of Pediatrics, University of Texas McGovern Medical Center, Houston, TX; 6) Department of Health Management and Policy, University of Iowa, Iowa City, IA; 7) Department of Orthodontics, University of Iowa, Iowa City, IA; 8) Department of Pediatrics, Seattle Children's Craniofacial Center, University of Washington, Seattle, WA; 9) Department of Surgery, Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA

---

The genetic basis of normal-range variation in human craniofacial traits is still poorly understood. Most studies to date have approached facial morphology via univariate measurements with genetic hypotheses centered on common genetic variants. In recent years, studies of other complex traits have identified fruitful associations with rare and low-frequency variants involved. To better understand the genomic architecture of normal-range facial traits, we studied the influence of low-frequency coding variants on multi-dimensional facial shape phenotypes. We genotyped a cohort of 2329 healthy individuals of European ancestry for approximately 245,000 coding variants on the Illumina Exome v1.2 array. Using three-dimensional facial images, we partitioned the full face into 31 hierarchically arranged segments, which we call modules, to model global-to-local features, and generated multi-dimensional phenotypes representing the shape variation within each module. We used multivariate kernel regression (implemented in the MultiSKAT R package) to test the association between the multivariate facial phenotypes and exome-wide variants with frequencies <1% in a gene-based manner. After accounting for multiple tests, eight genes (*AR*, *CARS2*, *FTSJ1*, *HFE*, *LOC108783645*, *LTB4R*, *TELO2*, *NECTIN1*) showed significant association with morphology of the cheek, chin, nose and philtrum. These genes displayed a wide range of phenotypic effects, with some impacting the full face and others affecting only localized regions. Certain cellular processes and functions were enriched in the group of nominated genes, including metal ion transport, steroid-hormone signaling, tRNA metabolism, DNA repair and cell-cell adhesion. Notably, *NECTIN1* is a well-established craniofacial gene that underlies the etiology of a syndromic form of cleft lip and palate (CLPED1; MIM#225060). We identified the significant missense variant rs142863092 in *NECTIN1*, predicted bioinformatically to be deleterious to the encoded protein. Little is known about the roles of the other implicated genes in craniofacial development, and future investigation holds the promise of yielding new insights. These results have expanded our understanding of the genetic basis of normal-range human facial morphology by implicating rare and low frequency coding variants in novel candidate genes.

## **PgmNr 2275: Creating a reporting framework for polygenic risk scores to improve transparency and standardization: Results from a ClinGen complex disease working group literature review.**

### **Authors:**

H. Wand <sup>1,2</sup>; C. Tamburro <sup>3</sup>; C. Sillari <sup>3</sup>; K. Vlessis <sup>1</sup>; K. Kinnear <sup>1</sup>; R. Rowley <sup>3</sup>; G. Wojcik <sup>1</sup>; ClinGen Complex Disease Working Group

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Stanford University, Stanford, California; 2) Stanford Hospital and Clinics, Stanford, California; 3) National Human Genome Research Institute (NHGRI), National Institutes of Health, Bethesda, Maryland

---

Over the past few years, the polygenic risk score (PRS) has become a more common tool for bridging the gap between the hypothesis-generating genome-wide association study (GWAS) and clinical application for disease risk estimation. In contrast with other clinical risk tools, there are no accepted standards for the development, reporting, and application of PRS. To address this gap, the Complex Disease Working Group (CDWG) of the Clinical Genome Resource (ClinGen, [clinicalgenome.org](http://clinicalgenome.org)) is working to specify and standardize reporting criteria for PRS with a two-fold approach: (1) assess the comprehensivity and clarity of a pilot reporting framework and (2) conduct a literature review of current PRS reporting practices including major gaps and trends. ClinGen's CDWG is uniquely positioned with membership of 30+ experts in the fields of epidemiology, statistics, disease-domain expertise, implementation, actionability, and ELSI with interests in the application of PRS.

To create a preliminary PRS reporting framework, we iteratively built upon previously established expert consensus guidelines, starting with the GRIPS statement (2011). This statement recommends a checklist of 25 items when reporting out genetic risk estimation studies. The ClinGen CDWG expanded on the original checklist to include 44 unique items, of which 33 should be reported for both training and validation sets, by incorporating expert feedback on anticipated reporting gaps and inconsistencies. Here we present the pilot results evaluating the utility of this expanded reporting framework through a literature review of 36 articles selected by NHGRI's Genomic Medicine XII (Genomics and Risk Prediction) planning committee as a comprehensive sample of PRS across a variety of disease domains. Each article was assessed by two independent reviewers according to our expanded reporting framework for completeness and clarity. Here we will present trends captured in this review with respect to our reporting framework, highlighting areas that lack reporting such as model fit and calibration, as well as inconsistent interpretations of common terms. We will also discuss the CDWG's revised reporting framework, which will enable the standardized dissemination of PRS while weighing barriers to participation in these reporting standards and ultimately facilitating translation into clinical care by allowing for the establishment of management guidelines based on PRS risk estimation.

## PgmNr 2276: Whole-blood RNA-Seq analysis detects differences in exon expression and alternatively spliced genes in smokers with emphysema.

### Authors:

A. El Boueiz <sup>1,2</sup>; Z. Xu <sup>1</sup>; J.H. Yun <sup>1,2</sup>; A. Saferali <sup>1</sup>; M. Parker <sup>3</sup>; R. Chase <sup>1</sup>; S. Lee <sup>1</sup>; E.K. Silverman <sup>1,2</sup>; C.P. Hersh <sup>1,2</sup>; P.J. Castaldi <sup>1,4</sup>; the COPDGene investigators

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Alnylam Pharmaceuticals, Cambridge, MA; 4) General Medicine and Primary Care, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

---

**Introduction:** Emphysema, irreversible destruction of lung parenchyma, is commonly observed in smokers. GWAS have revealed multiple genetic loci associated with emphysema. An integrative analysis of GWAS with blood and lung tissue gene expression has identified associations of *LILRA3* and *DCBLD1* genes with emphysema. However, no investigation to-date has evaluated the possible associations of emphysema with complex regulatory phenomena such as alternative splicing and differential isoform usage.

**Methods:** Short reads were generated from whole-blood RNA-Seq data from 1,984 subjects in the COPDGene study. Read alignment was performed using the STAR aligner. We quantified exon level counts using Rsubread and inferred isoform counts using Salmon. We tested gene expression, exon usage, and isoform usage associations with emphysema percentage (CT densitometry low-attenuation area less than -950 Hounsfield units (%LAA-950)) using the voom/limma method and adjusting for age, sex, race, current smoking, pack-years of smoking, FEV1, cell counts, and library batch. We also conducted gene set enrichment analyses using Gene Ontology (GO) categories and findings from prior COPD studies to provide biological and disease relevance.

**Results:** At FDR 10%, 1113 genes were associated with emphysema with 647 up-regulated and 466 down-regulated genes. Exon level usage analysis identified 26 exons from 25 genes with significant %LAA-950-related differential usage. 23 of the 26 differentially used exons are either in the first or last exon of the isoforms (OR=3.70,  $P=0.01$ ). At the isoform level, 586 isoforms in 524 genes were differentially used with 299 up-used and 287 down-used isoforms. These genes were enriched in GO pathways related to cell morphogenesis, positive regulation of T-helper 17 cell differentiation, peptidyl-tyrosine auto-phosphorylation, negative regulation of I-kappaB kinase/NF-kappaB signaling, ubiquitin-dependent protein degradation, negative regulation of protein kinase B signaling, sulfur oxidation, microtubule sliding, and TORC2 signaling (*adjusted*  $P<0.005$ ). Only 9 genes with differential exon usage and 75 genes with differential isoform usage also showed gene level differential expression.

**Conclusion:** Complex transcriptomic changes at the exon and isoform levels from whole blood

highlighted candidate genes and pathways that complement the findings from conventional gene level analyses and may provide further insight into the pathophysiology of emphysema.

## PgmNr 2277: Rare missense variants in intracerebral hemorrhage.

### Authors:

J. Chung<sup>1</sup>; M. Kim<sup>2</sup>; J. Henry<sup>3,5,6</sup>; B. Montgomery<sup>3,5</sup>; C.D. Langefeld<sup>7</sup>; S. Demel<sup>8</sup>; D. Woo<sup>8</sup>; J. Rosand<sup>3,4,5,6</sup>; C.D. Anderson<sup>3,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical Genetics, Department of Medicine, Boston University, Boston, MA; 2) Department of Bioinformatics, Korea University, Sejong, South Korea; 3) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 4) Program in Medical and Population Genetics, Broad Institute, Boston, MA, USA; 5) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 6) McCance Center for Brain Health, Massachusetts General Hospital, Boston, MA, USA; 7) Center for Public Health Genomics and Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA; 8) Department of Neurology and Rehabilitation Medicine, University of Cincinnati College of Medicine, Cincinnati, OH, USA

### Background

Common genetic variations in *COL4A1* and *COL4A2* at 13q34 have been previously associated with cerebral small vessel disease, including intracerebral hemorrhage (ICH) and small vessel ischemic stroke, as well as coronary artery disease. We hypothesized that rare coding variants in *COL4A1* and *COL4A2* may be responsible for ICH-associated common variant signals at 13q34.

### Methods

We performed targeted sequencing across 559Kbp at 13q34 including *COL4A1* and *COL4A2* among 2,133 individuals (1,055 ICH cases and 1,078 controls) and selected rare coding variants with predicted functional impact and appearance in at least two ICH cases and no controls. Sequence annotation for functional impact was conducted using gnomAD and the Ensembl Variant Effect Predictor (VEP). Phenomic impact in carriers of identified variants was examined using the Geno2MP database. We extended our findings to UK Biobank (UKB) exome sequence data (339 strokes and 42,610 controls) using BinomiRare. We further assessed structural effects of selected variants on protein structures of *COL4A1* and *COL4A2* using molecular dynamic simulation (MDS).

### Results

We identified two highly conserved missense variants at 13q34 appearing in six ICH cases but not in controls: 1) rs138269346 (p.I110T, *COL4A1*) in four cases; 2) rs201716258 (p.H203L, *COL4A2*) in two cases. Both variants were classified as pathogenic in gnomAD and VEP based on low frequency (MAF < 0.00035 in EUR) and predicted functional impact (*probably* or *possibly damaging*). Carriers of these two variants in Geno2MP expressed phenotypic abnormalities consistent with rare collagen mutation syndromes. In UKB, neither variant was associated with stroke (rs201716258 P=0.25; rs138269346 P=0.09) by ICD10 diagnosis but we confirmed the rarity of the variants (MAF < 0.001 for each) in the UKB EUR population. Using MDS, both variants caused a substantial decrease in physical length of the two protein structures compared with wildtype (> 8% shortening).

### Conclusion

We have identified two rare coding variants in *COL4A1* and *COL4A2* that appear only in ICH cases but not in controls. Our annotation and simulation studies imply that these variants are highly functional, disrupting the structure complexed by *COL4A1* and *COL4A2*. Associations with stroke in UKB were

limited by power. Further experiments will examine the contribution of these variants *in vitro* and test for associations with additional refined vascular phenotypes in UKB.



## PgmNr 2278: Whole blood gene expression analysis reveals active tuberculosis transcriptional signature in HIV co-infected children in Uganda and Botswana.

### Authors:

E. Katagirya <sup>1</sup>; S. Mwesigwa <sup>1</sup>; G. Mboowa <sup>1</sup>; G. Retshabile <sup>2</sup>; L. William <sup>2</sup>; B. Mlotshwa <sup>2</sup>; S. Kyobe <sup>1</sup>; M. Wayengera <sup>1</sup>; A. Mirembe <sup>3</sup>; W.S. Mpoloka <sup>2</sup>; I. Kasvosve <sup>2</sup>; K. Maplanka <sup>2</sup>; G.P. Kisitu <sup>3</sup>; A. Kekitiinwa <sup>3</sup>; M.L. Joloba <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) College of Health Sciences, Makerere University, Kampala Uganda; 2) University of Botswana, Gaborone, Botswana; 3) Baylor College of Medicine Children's Foundation Uganda (Baylor Uganda), Kampala Uganda; 4) Botswana Baylor Children's Clinical Centre of Excellence, Gaborone Botswana; 5) Department of Molecular and Human Genetics Baylor College of Medicine, Houston, TX; 6) University of Tennessee Health Science Center, LeBonheur Children's Hospital and St. Jude Children's Research Hospital, Memphis, TN

---

**SIGNIFICANCE:** Diagnosis of tuberculosis (TB) is problematic in individuals co-infected with HIV because of the atypical presentation of the disease, low infective dose of the organisms among other reasons. The difficulty is further compounded in children who have an added difficulty in the collection of sputum of optimum quantity and quality. There is, therefore, need for alternative diagnostic approaches and markers of disease. In this study, we use RNA sequence data from HIV infected children with different clinical TB phenotypes i.e. possible, probable and confirmed TB as well as non-TB infected children to determine the transcriptional signature for Active TB and/or different clinical phenotypes.

**METHODS:** IRB approval was obtained from the two study sites in Botswana and Uganda and RNA collected as a case-control study. RNA samples from a discovery data set of 24 Active TB cases and 24 non-TB infected age and gender-matched controls were sequenced on the Illumina HiSeq platform. FastQC was used for Quality Control of the sequences. The sequences were then aligned against the Human Reference Genome using Bowtie2/Cufflinks2. Transcriptome assembly was done using Cufflinks2 while differential gene expression was determined using CuffDiff2. CuffDiff2 determined the fold change in gene expression (normalized to log base 2). Geneset enrichment analysis was done using the GSEA desktop platform (Broad Institute) on the over 20,000 genes identified and ranked by fold change in expression. Multidimensional scaling (MDS) was done using the Orange Data Mining software.

**RESULTS:** Over 20,000 genes were identified in the cases and controls with 1500 and 300 genes upregulated and downregulated respectively in the cases. The gene set enrichment analysis showed enrichment of genes in gene sets involved in innate immunity and cytokine pathways.

**CONCLUSION:** Active TB disease upregulates gene expression of genes involved in innate immunity. We also show that whole blood gene expression profiles may be able to delineate clinical TB phenotypes.

## PgmNr 2279: Genome-wide linkage study of childhood gender nonconformity in males.

### Authors:

A. Sanders<sup>1,2</sup>; G. Beecham<sup>3,4</sup>; S. Guo<sup>3</sup>; K. Dawood<sup>5</sup>; G. Rieger<sup>6</sup>; R. Krishnappa<sup>7</sup>; A. Kolundzija<sup>8</sup>; J. Bailey<sup>9</sup>; E. Martin<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Psychiatry and Behavioral Sciences, NorthShore Univ HealthSystem, Evanston, IL; 2) Dept of Psychiatry and Behavioral Neuroscience, Univ of Chicago, Chicago, IL; 3) John P. Hussman Institute for Human Genomics, Univ of Miami Miller School of Medicine, Miami, FL; 4) Dr. John T. Macdonald Foundation Dept of Human Genetics, Univ of Miami Miller School of Medicine, Miami, FL; 5) Dept of Psychology, Pennsylvania State Univ, University Park, PA; 6) Dept of Psychology, Univ of Essex, Colchester, England, UK; 7) Dept of Psychiatry, Icahn School of Medicine at Mount Sinai, Elmhurst, NY; 8) Collective Impact, Washington, DC; 9) Dept of Psychology, Northwestern Univ, Evanston, IL

---

Male sexual orientation is a scientifically and socially important trait shown by family and twin studies to be influenced by environmental and complex genetic factors. Childhood gender nonconformity (CGN) is one of the strongest correlates of homosexuality, especially in males, and also has substantial familiarity. We have utilized a collection of families each with two or more homosexual brothers (409 concordant sibling pairs in 384 families), with a 23-item CGN questionnaire administered to all brothers and then converted to a quantitative scale. To map loci for CGN, we have conducted a genome-wide linkage scan (GWLS) using genotypes from the Affymetrix 5.0 SNP array analyzed with MERLIN after rigorous quality control steps. The strongest linkage peaks, each with multipoint LOD scores over 2.0, were on chromosomes 5q (maximum LOD = 2.11) and 8q (maximum LOD = 2.25), with the latter not overlapping with previously reported strongest linkage region for male sexual orientation on pericentromeric chromosome 8. Family-based association analyses were used to identify associated variants in the linkage regions. The current GWLS results may help prioritize findings from future large-scale sexuality GWAS on samples with CGN data available now being collected. Further increasing genetic knowledge about CGN and its relationships to male sexual orientation, especially via large GWAS, should help advance our understanding of the biology of these important traits. This work was supported NIH grants R01HD041563 and R21HD080410.

## PgmNr 2280: Pedigree based analysis of dizygotic twinning from whole genome sequencing.

### Authors:

J.J. Beck <sup>1</sup>; E.A. Ehli <sup>1</sup>; J.J. Hottenga <sup>2</sup>; H. Mbarek <sup>2</sup>; E.J.C.N. de Geus <sup>2</sup>; D.I. Boomsma <sup>1,2</sup>; G.E. Davies. <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Avera Institute for Human Genetics, Sioux Falls, SD, USA.; 2) Department of Biological Psychology, Behavioral and Movement Sciences, Vrije Universiteit, Amsterdam, The Netherlands.

---

Spontaneous dizygotic (DZ) twinning results from the release and fertilization of two or more oocytes during a single pregnancy. Considered a common complex trait, DZ twinning events affect approximately 1-4% of women worldwide with large regional differences known to exist. Genome-wide association studies of DZ twinning have implicated common genetic variants in *FSHB* and *SMAD3*. Some rare genetic variants, including a loss of function mutation in *GDF9*, have been thought to influence DZ twinning, but with conflicting results. To optimize rare genetic variant detection, we analyzed whole genome sequences of four distantly related mothers of DZ twins from a large Dutch pedigree with a rich history of DZ twinning. The pedigree spans seven generations and includes 18 mothers of DZ twins and 21 sets of DZ twins. Samples were sequenced on an Illumina Hi-Seq 2500 instrument (101bp paired-end reads). Sequence data were aligned to human reference genome build 37 with Burrows-Wheeler Aligner. Variant discovery was performed with GATKv3.8. To define haplotypes that segregate with being a mother of DZ twins, we identified regions in common between the four sequenced mothers. Elucidation of high-priority regions was done by comparing genotypic data, obtained from an Illumina GSA array, of an additional nine affected mothers from the same pedigree using a minor allele frequency filter (MAF 0.001) to remove monomorphic variants. Overlapping regions were determined to be located on chromosomes 11, 14, and X. Based on these overlapping regions between all affected mothers, we will construct haplotype blocks for further investigation and test for deleterious mutations. In a next step, we will make use of the Genome of the Netherlands (GoNL) resource, in which 250 trios were sequenced on an Illumina Hi-Seq 2000 (see Genome of the Netherlands, 2014). Of the trios, 110 were derived from the Netherlands Twin Register and included 46 mothers of dizygotic twins. We plan to evaluate if either the identified mutation is only present in the pedigree or whether the haplotype is shared between DZ twinning affected mothers.

## PgmNr 2281: Sequencing-based, multi-population fine mapping of the 10q24.32/AS3MT arsenic metabolism efficiency locus.

### Authors:

M. Chernoff<sup>1</sup>; D. Delgado<sup>2</sup>; L. Tong<sup>2</sup>; M. Stephens<sup>3,4</sup>; F. Jasmine<sup>2</sup>; L. Huang<sup>5</sup>; S.A. Cole<sup>6</sup>; K. Haack<sup>6</sup>; J. Umans<sup>7</sup>; L.G. Best<sup>8</sup>; A. Navas-Acien<sup>9</sup>; M.R. Kargas<sup>10</sup>; H. Ahsan<sup>2,3,12</sup>; M.G. Kibriya<sup>2</sup>; B. Pierce<sup>2,3,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Interdisciplinary Scientist Training Program: Public Health Sciences, University of Chicago, Chicago, Illinois.; 2) Department of Public Health Sciences, University of Chicago, Chicago, IL; 3) Department of Human Genetics, University of Chicago, Chicago, IL; 4) Department of Statistics, University of Chicago, Chicago, IL; 5) Center for Research Informatics, The University of Chicago, Chicago, IL; 6) Texas Biomedical Research Institute, San Antonio, TX; 7) Georgetown-Howard Universities Center for Clinical and Translational Science, Washington, DC; 8) Missouri Breaks Industries Research Inc., Timber Lake, South Dakota; 9) Mailman School of Public Health, Columbia University, New York City, NY; 10) Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH; 11) Comprehensive Cancer Center, University of Chicago, Chicago, IL; 12) Department of Medicine, University of Chicago, Chicago, IL

---

Arsenic is a carcinogen affecting ~56 million people in Bangladesh and 13 million in the U.S. via drinking water. Once consumed, inorganic arsenic is converted to mono- and di-methylated (DMA) forms. DMA, expressed as a percentage of urinary arsenic, is a measure of arsenic metabolism efficiency (AME). Inherited genetic variation in the 10q24.32 region, containing the *AS3MT* (arsenic methyltransferase) gene, impacts AME, with 2 independent association signals identified in our prior studies. The causal variants and mechanisms underlying these associations are unknown. To comprehensively study variation in the 10q24.32 region, we performed targeted sequencing in 3 populations. This includes 2436 Bangladeshi individuals from the Health Effects of Arsenic Longitudinal Study (HEALS), 874 American Indians from the Strong Heart Study (SHS), and 706 European Americans from the New Hampshire (NH) Skin Cancer Study of squamous cell carcinoma. Conditional, forward stepwise regressions in each population identified variants independently associated with DMA%. We identified potential causal sets in each population and the estimated likelihood of a shared causal variant, using a Bayesian method implemented in the Sum of Single Effects Regression (SuSiE) R package. We used tissue-specific eQTL results from GTEx and Bayesian co-localization analysis to determine if SNPs affecting DMA% also impact expression of *AS3MT* or nearby genes. We identified 3 independent association signals in the 10q24.32 region in HEALS (p-values  $10^{-17}$  to  $10^{-11}$ ), 4 in SHS (p-values  $10^{-20}$  to  $10^{-5}$ ), and 1 in NH (p-value  $7 \times 10^{-5}$ ). HEALS and SHS share a lead SNP, rs4919687, and moderate LD was observed between this SNP and the lead SNP in NH. All identified SNPs are in a 360kb region containing *AS3MT*, *BORCS7*, *CNNM2*, and *NT5C2*. There were cis-eQTLs for *AS3MT* in 13 GTEx tissue types for which the lead SNP was in LD ( $r^2 > 0.7$ ) with one of the 3 lead SNPs in HEALS. Our primary association signal co-localized with cis-eQTLs for *AS3MT* in 11 tissue types (posterior probability >80%). We also found evidence of co-localization between 5 additional genes in the 10q24.32 region and cis-eQTLs in 39 tissue types. Our results suggest 1) the presence of shared associations, allowing us to leverage population LD differences to identify

candidate causal variants and 2) at least one variant likely acts via regulation of local gene expression. This increases our understanding of the 10q24.32 region and its role in AME.

## PgmNr 2282: Dissecting out the bones of height.

### Authors:

B. Harrington<sup>1</sup>; B. Hill<sup>1</sup>; J.F. Wilson<sup>1,2</sup>; P. Navarro<sup>1</sup>; C. Haley<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MRC Human Genetics Unit, University of Edinburgh, Edinburgh, United Kingdom; 2) Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom; 3) Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

---

Despite years of study, and hundreds of identified associated genetic variants, not all of the genetic variation in morphological traits, such as height, is currently explained. Dividing height into its constituent parts, such as the long bones in the leg, may help to identify novel height-associated variants, and to better understand bone morphology, as well as the components of growth. The major long bones of leg are the largest contributors to height in humans, and, while their length correlates with height, significant variation exists independent of variation in height. Genetic variants acting on the length aspect of morphology may have larger relative effect sizes on long bone length than they do on height, so focusing on these more immediate phenotypes should result in greater power to uncover single nucleotide polymorphism (SNP)-trait associations. Bone lengths have previously seen limited study in humans due to the difficulty of measuring them accurately for living participants, and the lack of cohorts with both long bone lengths and genetic data available.

The major long bone lengths in both the arms (humerus, radius, ulna) and legs (femur, tibia, fibula) were obtained from dual x-ray absorptiometry (DXA) whole-body scans in the ORCADES cohort ( $N=1,150$ ). Genotyping was performed using the Illumina<sup>®</sup> HumanHap300 and OmniExpress arrays (278,618 and 599,638 markers), and data were imputed to the Haplotype Reference Consortium panel (12.3 million SNPs after imputation and QC). Bone length SNP heritability estimates for these traits range from 34% for ulna length, to 59% for fibula length, after accounting for height, age, and sex. We carried out a genome-wide association study (GWAS) using a mixed-linear model framework to look for genetic variants associated with the length of the major long bones in both the arms and legs. GWAS uncovered genome-wide significant associations with SNPs not identified in height GWAS, or known to be in linkage disequilibrium with height-associated variants. Many of the identified variants are found in intergenic regions near genes whose function provides plausible biological explanations for the associations, however replication is still required. Based on these preliminary findings, further work is underway using the UK Biobank DXA whole-body scans ( $N=15,000$ ) and image-analysis software that has been developed in-house to automatically extract long bone lengths from the scans.

## PgmNr 2283: Cross-population fine-mapping of 50 complex traits and diseases in ~675,000 individuals across three global biobanks.

### Authors:

M. Kanai<sup>1,2,3,4,5,6</sup>; J.C. Ulirsch<sup>2,7</sup>; Z.R. McCaw<sup>8</sup>; C. Albers<sup>2</sup>; Z. Fan<sup>9</sup>; K. Ishigaki<sup>10,11</sup>; J. Karjalainen<sup>1,2,3,4</sup>; M. Kurki<sup>1,2,3,4</sup>; Y. Kamatani<sup>10,12</sup>; Y. Okada<sup>6,13</sup>; M.J. Daly<sup>1,2,3,4</sup>; H.K. Finucane<sup>2</sup>; FinnGen

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 5) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA; 6) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan; 7) Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; 8) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 9) Department of Statistics and Data Science, Yale University, New Haven, CT; 10) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 11) Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 12) Kyoto-McGill International Collaborative School in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 13) Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan

Identifying causal variants for complex traits is one of the major challenges in human genetics. The causal variants in most GWAS associated loci remain unknown due to insufficient power and to high linkage disequilibrium (LD) among variants in a locus. Cross-population fine-mapping studies may improve fine-mapping resolution by leveraging sample size and differences in LD and minor allele frequency (MAF) across populations. However, lack of availability of non-European GWAS and proper statistical methods have limited the success of previous studies.

Here, we fine-mapped 50 complex traits and diseases from three large-scale biobanks, UK Biobank (UKB;  $n = 361,194$ ), FinnGen ( $n = 135,638$ ) and Biobank Japan (BBJ;  $n = 179,066$ ). We first conducted single-population fine-mapping using two state-of-the-art methods, FINEMAP (Benner *et al*, 2016) and SuSiE (Wang *et al*, 2018). In total, 51,302 putative causal variants were identified (posterior inclusion probability [PIP] > 0.1), with a median number of 21 variants in 95% credible sets. Our results suggested that single-population fine-mapped variants are rarely shared across populations (915 variants; 1.8%), whereas their posterior effect sizes in allelic scale are largely consistent if shared (Pearson's  $r = 0.83$ , 95% CI: 0.81–0.84). This inconsistency of PIP is partly due to difference in MAF, imputation quality, and various quality control criteria, which all resulted in missing putative causal variants in one population or incorrect imputation of haplotypes. For example, among 42,345 non-shared fine-mapped variants in UKB, we observed lower MAF in BBJ than in UKB (mean MAF = 0.19 and 0.24, respectively). In addition, 16.7% of them are monomorphic (MAF = 0 or non-existent in

imputation reference) and 15.6% are missing (available in reference but removed after QC) in BBJ. Our results highlight the importance of these factors for robust fine-mapping, even in a single population.

To tackle this issue and maximize our power, we investigate a novel cross-population fine-mapping method that leverages single-population fine-mapping results from a harmonized pipeline and applies an empirical Bayes approach that leverages genome-wide data to learn patterns of shared causality across populations (see McCaw *et al*, ASHG19). We apply our method to a set of biobank-scale simulated datasets and the 50 complex traits and diseases. Our preliminary results suggest that our method successfully improves fine-mapping power and resolution.



## PgmNr 2284: 203 SNPs with combinatorial effects contribute to the genetic architecture of facial morphology.

### Authors:

J.D. White<sup>1</sup>; K. Indencleef<sup>2,3,4</sup>; R.J. Eller<sup>5</sup>; S. Naqvi<sup>6,7</sup>; J. Roosenboom<sup>8</sup>; M.K. Lee<sup>8</sup>; J. Li<sup>2,3</sup>; J. Mohammed<sup>6,9</sup>; S. Richmond<sup>10</sup>; E. Feingold<sup>11</sup>; M.L. Marazita<sup>8,11</sup>; H. Peeters<sup>12</sup>; G. Hens<sup>4</sup>; J.R. Shaffer<sup>8,11</sup>; J. Wysocka<sup>6,9</sup>; S. Walsh<sup>5</sup>; S.M. Weinberg<sup>8,11,13</sup>; M.D. Shriver<sup>1</sup>; P. Claes<sup>2,3,12,14,15</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Anthropology, Pennsylvania State University, State College, PA.; 2) Department of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium; 3) Medical Imaging Research Center, UZ Leuven, Leuven, Belgium; 4) Department of Otorhinolaryngology, KU Leuven, Leuven, Belgium; 5) Department of Biology, Indiana University Purdue University Indianapolis, Indianapolis, IN; 6) Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA; 7) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 8) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA; 9) Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA; 10) Applied Clinical Research and Public Health, School of Dentistry, Cardiff University, Heath Park, Cardiff, United Kingdom; 11) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 12) Department of Human Genetics, KU Leuven, Leuven Belgium; 13) Department of Anthropology, University of Pittsburgh, Pittsburgh, PA; 14) Murdoch Children's Research Institute, Melbourne, Victoria, Australia; 15) Department of Biomedical Engineering, University of Oxford, Oxford, United Kingdom

In this study, we highlight groups of SNPs with potential large to fine scale combinatorial interactions affecting the human face. The initial pool of SNPs comes from a meta-analysis GWAS of 3D facial shape in two large cohorts ( $N_{US} = 4,680$ ;  $N_{UK} = 3,566$ ) in which each face was densely registered with 7,160 quasi-landmarks, which were then hierarchically grouped into segments representing global-to-local facial variation. For each SNP, we first identify the linear combination of principal components that are maximally correlated with the SNP in the identification cohort. We then project the verification cohort onto this learned phenotype, creating univariate facial shape variables to be tested for association. This process is repeated, switching the dataset used for identification and verification, and the p-values are meta-analyzed. Genome-wide significant SNPs were grouped by genomic location and similarity of effect in the two cohorts, refining our follow-up analyses to 203 lead SNPs at 138 genomic positions that, in total, explain ~5% of the phenotypic variation in our cohorts and are enriched for H3K27ac activity in cranial neural crest cells and craniofacial tissues. We additionally used Structural Equation Modeling (SEM) to better understand the large-scale groupings of SNPs that together best explain the variance observed in each facial segment. For the 25 segments passing all recommended model fit parameters, the correlation between H3K27ac activity across 100 different cell-types and tissues of the SEM-refined SNPs was high, indicating that the refined SNPs potentially have coordinated actions in the same cell-types or tissues. By performing a linear regression between the facial traits identified alongside each lead SNP and the genotypes at all other lead SNPs, we additionally identified 110 traits as having associations with multiple SNPs, giving rise to a smaller set of SNPs that might work in concert to create these facial traits. Lastly, the latent univariate phenotype produced for each SEM model was used to assess whether interactions between genotypes mask or

boost the genotypic effect on a phenotype, resulting in four SNP combinations with evidence of interacting effects, three of which have been previously reported in the literature. In sum, our analyses provide new insights into the genetic architecture of human facial shape, both in terms of SNPs with single effects as well as SNPs showing combinatorial interactions.

## **PgmNr 2285: Developing a data resource for pediatric research: The Gabriella Miller Kids First Pediatric Research Program (Kids First).**

### **Authors:**

D. Winchester <sup>1</sup>; V. Cotton <sup>2</sup>; The Gabriella Miller Kids First Pediatric Research Program Working Group

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Office of the Director, Office of Strategic Coordination - The Common Fund, National Institutes Health, Rockville MD; 2) Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health Rockville,MD

---

Epidemiological studies have reported that children with congenital anomalies are at a higher risk of developing certain childhood cancers. Little is known about the genetic cause-and -effect relationship between these disorders. Furthermore, a large dataset is needed to examine the role of genetics in childhood cancer and birth defects. The National Institutes of Health Common Fund's Gabriella Miller Kids First Pediatric Research Program (Kids First) is developing a large-scale data resource of clinical and genetic data from patients with childhood cancers and structural birth defects and their families. These data are now available through the Gabriella Miller Kids First Data Resource Portal which was launched by the program's Data Resource Center in 2018. To date, Kids First has selected 27 patient cohorts for whole genome sequencing through a peer-review process. Clinical and genetic sequence data from over 5,000 patient samples are accessible through the portal, including data from patients with Ewing Sarcoma and Congenital Diaphragmatic Hernia. Data from more than 30,000 DNA and RNA samples are expected to be added to the Kids First Data Resource Portal over the next few years. The Kids First program is focused on data sharing to develop tools and resources to foster collaborative analyses. The program aims to help researchers uncover new insights into the biology of childhood cancer and structural birth defects and identify shared genetic pathways among these various pediatric conditions. A complete list of conditions represented in the Kids First data resource and additional information about the Kids First program will be described in this presentation.

## PgmNr 2286: LabWAS: Investigating relationships between biomarkers and genetic risk for complex diseases.

### Authors:

J. Sealock <sup>1</sup>; J. Dennis <sup>1,2</sup>; L.K. Davis <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Department of Medical Genetics, University of British Columbia, Vancouver, Canada

---

Clinical laboratory testing provides physicians with targeted biochemical measurements (i.e., biomarkers) to aid in diagnosing and treating patients for a variety of diseases. Combining genetics of complex traits to biomarker and diagnostic data from electronic health records could point towards mechanisms of disease, provide information on the relationship between biomarkers and diagnosis, and facilitate the development of diagnostic panels. We describe a method that pairs polygenic scores with lab data derived from EHRs to screen for associations. Results from 3,800 labs were extracted for 70,184 individuals in Vanderbilt University Medical Center's biobank, BioVU. Usable labs were required to have a minimum of 1,000 observations over at least 100 individuals, and 70% of measurements to be recorded in one set of units. The remaining 475 labs were filtered to exclude observations outside 4 standard deviations from the mean. The median value of each lab for each individual was calculated, inverse normalized, and adjusted for cubic splines of age at measurement. Labs exhibiting no calculable heritability using GCTA were excluded leaving 272 labs for analysis. To screen for associations between PRS and lab measurements, we created a lab-wide association study (LabWAS) pipeline. This pipeline fits a linear regression model to each lab and determines the association between the PRS after further adjusting for sex, top 10 principal components of ancestry, and genotyping batch. We compared polygenic scores calculated using PRS-CS, PRSice, and LDpred, and selected the PRS with the highest predictive performance with CAD, PRS-CS. In a proof-of-principle analysis, PRS for coronary artery disease reproduced known associations with canonical heart disease risk factors. The strongest association was nucleated red blood cells (OR = 0.97, p-value = 1.43e-8), followed by HDL cholesterol (OR = 0.96, p-value = 9.51e-7), and blood glucose levels (OR = 1.03, p-value = 1.12e-4). Results from LabWAS provide a starting point for further sensitivity analyses to determine shared genetic architecture, relevant points in life, and effects of medication. We are currently working on ways to integrate effects of age and medication in our analysis. We will share the development of the pipeline, available software, and the results of ongoing replication efforts.

## PgmNr 2287: Estimation of heritability mediated by assayed gene expression levels elucidates gene architecture of complex traits.

### Authors:

D. Yao<sup>1,2</sup>; L. O'Connor<sup>1,2,4</sup>; A. Price<sup>3,4</sup>; A. Gusev<sup>3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA; 2) Harvard Graduate School of Arts and Sciences, Cambridge, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 5) Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA

---

Disease variants identified by genome-wide association studies (GWAS) tend to overlap with expression quantitative trait loci (eQTLs), suggesting that genetic effects on disease could be mediated by gene expression levels. However, existing methods that aim to quantify overlap between GWAS loci and eQTLs cannot distinguish mediation (where genetic variants causally influence disease through expression levels) from pleiotropy (where genetic variants independently influence disease and expression levels), even though the latter is not informative of disease etiology. Here we introduce a new method, Mediated Expression SCore regression (MESc), to estimate heritability mediated by the cis-genetic component of assayed gene expression levels using summary association statistics from GWAS and eQTL studies. MESc distinguishes mediated from pleiotropic effects in a set of genes by relying on the idea that the magnitude of effects of a SNP on the expression of genes in that gene set should be proportional to the magnitude of effects of the SNP on the disease if any mediation is occurring, whereas these quantities will not be proportional if effects of the SNP on the trait are not mediated by expression. We show that MESc produces robust estimates of expression-mediated heritability across a wide range of simulations.

We applied MESc to GWAS summary statistics for 42 complex traits (average N = 323K) and cis-eQTL data across 49 tissues from the GTEx consortium. **We determined that a statistically significant but modest proportion of disease heritability (mean estimate 0.11 with S.E. 0.02) is mediated by the cis-genetic component of assayed gene expression levels across traits, though estimates varied substantially among individual traits (point estimates from 0 to 0.38).** We further estimated the heritability mediated by expression of different subsets of genes. **Across all tested genes, we observed a strong inverse relationship between cis-heritability of expression and proportion of heritability mediated by expression, suggesting that genes with weaker eQTLs have larger causal effects on disease.** Moreover, across a large number of functional gene sets, we observed broad patterns of mediated heritability enrichment that recapitulate known biology and implicate specific gene sets in disease. Our results provide concrete evidence for the hypothesis that SNPs impact disease via regulation of gene expression levels, and give insight into the *gene* architecture of disease.

## PgmNr 2288: Implicating effector genes at GWAS loci associated with age at menarche through genome-wide promoter-focused Capture C in human ESC-derived hypothalamic neurons.

### Authors:

C.A. Doege<sup>1</sup>; K.M. Hodge<sup>2</sup>; C. Su<sup>2</sup>; E. Manduchi<sup>2,3</sup>; S. Lu<sup>2</sup>; M.E. Leonard<sup>2</sup>; J.A. Pippin<sup>2</sup>; M.E. Johnson<sup>2</sup>; R. Rausch<sup>1</sup>; B.F. Voight<sup>3</sup>; A. Chesi<sup>2,3</sup>; A.D. Wells<sup>2,3</sup>; R.L. Leibel<sup>1</sup>; S.F.A. Grant<sup>2,3</sup>; D.L. Cousminer<sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Columbia University, New York, NY; 2) Children's Hospital of Philadelphia, Philadelphia, PA; 3) University of Pennsylvania, Philadelphia, PA

---

Puberty marks the onset of reproductive maturity and impacts adult health outcomes including diabetes and cancer. At pubertal onset, genetic, nutritional and environmental signals are integrated to reawaken the pulsatile release of gonadotropin-releasing hormone (GnRH) from the hypothalamus to the pituitary. Despite the discovery of hundreds of GWAS signals for age at menarche (AAM) in girls and age at voice break in boys, the molecular mechanisms underlying pubertal initiation are not fully understood. Determining the effector genes and causal variants at these established GWAS loci should provide insight into molecular mechanisms for functional maturation of the hypothalamic pubertal trigger. We performed molecular fine-mapping in human embryonic stem cell (hESC)-derived hypothalamic arcuate neurons using high-resolution promoter-focused capture C coupled with ATAC-seq. We identified 32,614 proxy SNPs in LD ( $r^2 > 0.5$ ) with 379 AAM-associated sentinel SNPs, which were then further reduced using ATAC-seq to 714 proxies residing in open chromatin. Some of these open proxies were located in gene promoters, such as for *CENPW*. We used the genome interactome map established by Capture C to investigate direct contacts between open non-promoter SNPs and promoters of their putative target genes. We detected such contacts for 152 target genes at 62 loci. At the '*CADPS2*' locus, we identified an open proxy-to-promoter interaction for *FEZF1*, a gene ~150kb upstream whose product disrupts migration of GnRH neurons, resulting in hypogonadotropic hypogonadism. At the '*RNLS*' locus, we observed a contact with the promoter region ~600kb upstream of *PTEN*, a tumor suppressor gene in which mutations have been implicated in cases of precocious puberty. RNAseq further confirmed that these three target genes were highly expressed in hypothalamic neurons. Ingenuity pathway analysis across all target genes revealed significant enrichment for processes related to the central nervous system and brain development, but also to 'growth of genital organ.' Our results help clarify the possible hypothalamic effector genes among known plausible candidates, and also implicate novel target genes. Functional verification with CRISPR-Cas9 manipulation of promoter-interactors followed by target gene expression assays and assessment of key readouts, such as GnRH, should aid in confirming novel target effector genes associated with the timing of puberty.

## PgmNr 2289: Cross-population fine-mapping to identify shared and population-specific causal effects.

### Authors:

Z. McCaw<sup>1,2</sup>; M. Kanai<sup>1,3</sup>; C. Albers<sup>1</sup>; J. Ulirsch<sup>1,3</sup>; H. Finucane<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA.; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.; 3) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA.

Fine-mapping methods aim to isolate the causal genetic variants within a genomic region previously associated with a complex trait or disease. Such loci are of interest for understanding biological mechanisms and as potential therapeutic targets. Jointly analyzing multiple populations can improve power for detecting shared causal variants. We propose a cross-population fine-mapping model that allows for shared and population-specific causal effects. Our model accommodates multiple causal variants within a single locus, and allows for cross-population correlations in effect sizes. Working within the Bayesian paradigm, we implement a variational algorithm to estimate the posterior distribution of target parameters. From this, we can construct credible sets that contain the causal variant(s) with high probability.

We present realistic, bio-bank scale simulation studies validating our estimation and inference procedures. Individual-level genotypes for  $N = 500K$  subjects from each of African, European, and East Asian ancestry were simulated using *msprime* (Kelleher, et al., 2016). The coalescent simulation incorporated a demographic model (Gravel, et al., 2011), estimated from 1000 Genomes data, to ensure the simulated genotypes reflect modern human population histories. The simulated genomes were partitioned into loci, and within each locus, causal variants were assigned to populations based on two parameters: the probability of a locus harboring at least one causal variant, and the probability of a causal variant being shared across populations. These parameters have been estimated empirically using data from multiple biobanks (see Kanai, et al., ASHG19). Effect sizes at the causal variants were drawn from a multivariate normal distribution, with the covariance structure depending on the SNP heritability, MAF, and the cross-population genetic correlation. Using these data, we characterize the sensitivity and specificity of the proposed fine-mapping model for identifying shared and population-specific causal effects.

## PgmNr 2290: Investigating genetic links between normal-range facial variation and orofacial clefting.

### Authors:

M. Lee <sup>1</sup>; J. Roosenboom <sup>1</sup>; K. Indencleef <sup>2</sup>; J.D. White <sup>3</sup>; J.C. Carlson <sup>4</sup>; E. Feingold <sup>5</sup>; T.H. Beaty <sup>6</sup>; E.J. Leslie <sup>7</sup>; J. Wysocka <sup>8</sup>; M.L. Marazita <sup>1</sup>; S. Walsh <sup>9</sup>; M.D. Shriver <sup>3</sup>; P. Claes <sup>2</sup>; J.R. Shaffer <sup>5</sup>; S.M. Weinberg <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, USA; 2) Department of Electrical Engineering, KU Leuven, Belgium; 3) Department of Anthropology, Penn State University, USA; 4) Department of Biostatistics, University of Pittsburgh, USA; 5) Department of Human genetics, University of Pittsburgh, USA; 6) Department of Epidemiology, Johns Hopkins University, USA; 7) Department of Human genetics, Emory University, USA; 8) Department of Chemical and Systems Biology, Stanford University School of Medicine, USA; 9) Department of Biology, Indiana University Purdue University Indianapolis, USA

---

Using multivariate phenotypes derived from 3D facial images, we recently identified 203 genetic signals influencing human facial shape in a meta-GWAS of 8242 European individuals. These facial GWAS signals showed significant enrichment for genes implicated in dysmorphic facial traits, most notably orofacial clefts (OFCs), suggesting some overlap between genes responsible for normal and abnormal facial variation. While a number of genetic loci have been recognized as risk factors for OFCs, a large portion of the heritable variation remains unexplained. Loci that contribute to normal-range facial traits may help to identify new candidate genes and pathways involved in the etiology of OFC and prioritize possible biologically meaningful associations that fail to achieve standard thresholds for genome-wide significance. To evaluate this hypothesis, we selected a subset of 155 of the 203 GWAS signals implicated in normal-range facial morphology. We then tested these signals in two independent nonsyndromic OFC cohorts (n=7053 and n=5417) for which GWAS results were available. We observed a significant ( $p < 0.0001$ ) association between OFC and 55 loci influencing facial morphology. Fourteen loci pointed to known candidate genes for OFC and showed strong evidence of association in one or both of our OFC cohorts. The remaining 41 loci were novel for OFC. Of these 41 novel loci, seven showed independent evidence of association across the two OFC cohorts, including *THRB/RARB*, *CACNA2D3*, *PRDM5/TNIP3*, *4q28*, *FREM1*, *ALX4*, and *DAAM/DACT1*. Variants at these loci may modify risk for OFC, possibly by altering embryonic facial shape in ways that could alter lip and palate fusion. Future work will evaluate whether these variants have an effect on phenotypic severity through interactions with other well-established OFC risk variants. This research is supported by NIH grants U01-DE020078, R01-DE016148, and R01-DE027023.



## PgmNr 2291: Using wearable devices and genetics to estimate and validate mechanisms of sleep.

### Authors:

A.R. Wood <sup>1</sup>; S.E. Jones <sup>1</sup>; V.T. van Hees <sup>2</sup>; D.R. Mazzotti <sup>3</sup>; P. Marques-Vidal <sup>4</sup>; S. Sabia <sup>5,6</sup>; A. van der Spek <sup>7</sup>; H.S. Dashti <sup>8,9</sup>; J. Engmann <sup>5</sup>; D. Kocovska <sup>7</sup>; M. Hillsdon <sup>10</sup>; A.I. Luik <sup>7</sup>; N. Amin <sup>7</sup>; J.M. Lane <sup>8,9</sup>; R. Saxena <sup>8,11</sup>; M.K. Rutter <sup>12,13</sup>; H. Tiemeier <sup>5,14</sup>; Z. Kutalik <sup>4,15</sup>; M. Kumari <sup>16</sup>; T.M. Frayling <sup>1</sup>; M.N. Weedon <sup>1</sup>; P. Gehrman <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics of Complex Traits, University of Exeter Medical School, Exeter, Devon, United Kingdom; 2) Netherlands eScience Center, Amsterdam, The Netherlands; 3) University of Pennsylvania, Philadelphia, PA, USA; 4) Lausanne University Hospital, Lausanne, Switzerland; 5) University College London, UK; 6) Université de Paris, Paris, France; 7) Erasmus MC University Medical Center, Rotterdam, Netherlands; 8) Massachusetts General Hospital, Boston, MA, USA; 9) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 10) Sport and Health Sciences, College of Life and Environmental Sciences, University of Exeter, Exeter, EX1 2LU, UK; 11) Harvard Medical School, Boston, MA, USA; 12) University of Manchester, Manchester, UK; 13) Manchester University NHS Foundation Trust, Oxford Road, Manchester, UK; 14) Harvard TH Chan School of Public Health, Boston, MA, USA; 15) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 16) University of Essex, Colchester, UK

---

The widespread availability of wearable devices, such as wrist worn accelerometers, offers the potential to objectively measure sleep, and thus better understand the role of sleep in disease development. While polysomnography (PSG) is regarded as the gold standard method of measuring sleep, it is impractical to perform in large cohorts. To better understand the mechanisms of sleep, we aimed to derive accelerometer estimates of sleep duration, timing and quality and perform subsequent genetic analyses on these phenotypes to identify novel genetic associations.

Using accelerometer data in up to 85,670 participants from the UK Biobank study, we derived measures of sleep duration (duration and variability), quality (sleep efficiency and number of nocturnal sleep episodes), and timing (including nocturnal sleep and the least active 5-hours) using an algorithm previously validated against PSG data. Genetic data available from the UK Biobank was used for genome-wide association analyses and validation of phenotypes.

The phenotypic correlation between accelerometer-estimates of sleep timing and measures of sleep duration and quality were low ( $-0.10 \leq R \leq 0.12$ ), consistent with data from self-reported chronotype ("morningness") and sleep duration in the UK Biobank ( $R = -0.01$ ). We observed a stronger correlation between sleep duration and sleep efficiency ( $R=0.57$ ). Heritability estimates ranged from 2.8% (95% CI 2.0%, 3.6%) for sleep duration variability, to 22.3% (95% CI 21.5%, 23.1%) for the number of nocturnal sleep episodes. We identified 47 genetic associations at  $P < 5 \times 10^{-8}$ . These include 26 novel associations with measures of sleep quality and 10 with nocturnal sleep duration. We replicate a previously reported gene locus (*PAX8*) associated with sleep duration based on self-report data. We observe variants previously associated with restless legs syndrome associate with multiple sleep traits. As a group, sleep quality loci are enriched for serotonin processing genes.

Using accelerometer data in the UK Biobank, we have derived eight measures of sleep characteristics. Genetic associations between these measures and known restless legs syndrome associations provides validation of our methods.

## PgmNr 2292: A large-scale genome-wide association scan in East Asians revealed genetic mechanisms underlying variations of human facial morphology.

### Authors:

S. Wu <sup>1,2,3</sup>; M. Zhang <sup>1,2,3</sup>; S. Wang <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200438, China;; 2) CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China; 3) Human Phenome Institute, Fudan University, 825 Zhangheng Road, Shanghai 201203, China

---

The facial surface is visible and recognizable based on the facial shape variation which represents a set of highly polygenic and correlated complex traits. Understanding the genetic basis underlying facial shape traits has important implications in population genetics, developmental biology, and forensic science. However, relatively little is known about which and how genes influence human facial variation in East Asians. Here, we conduct a genome-wide association scan for facial features in ~9,500 East Asians from high resolution 3dMD facial images. We identified 54 loci that significantly associated with facial shape at multiple levels of organization, showing distinctive patterns of global-to-local genetic effects on facial shape. Among the 54 loci, 33 were supported in the literature, while 21 were newly identified. Functional enrichment analysis showed that most of genes were linked to regionalization, appendage morphogenesis and embryonic organ development, suggesting that the facial morphology variation be shaped during the pattern specification process. Finally, we found that the associated loci were enriched for signals of positive selection, indicating that facial morphology had undergone strong local adaptation during the recent evolutionary history. These results substantially advanced our understanding of the genetic mechanisms underlying human facial variations.

## PgmNr 2293: Phenotypic and genetic associations of growth differentiation factor-15 plasma levels in a Finnish population cohort.

### Authors:

S. Lemmelä<sup>1</sup>; E. Wigmore<sup>2</sup>; C. Benner<sup>1</sup>; A. Havulinna<sup>1,5</sup>; T. Kempf<sup>3</sup>; K.C. Wollert<sup>3</sup>; S. Blankenberg<sup>4</sup>; T. Zeller<sup>4</sup>; V. Salomaa<sup>5</sup>; A. Platt<sup>2</sup>; R. March<sup>2</sup>; A. Palotie<sup>1,6,7</sup>; M. Daly<sup>1,6,7</sup>; M. Kinnunen<sup>1</sup>; A. Matakidou<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine, Helsinki, Finland; 2) Centre for Genomics Research, Discovery Sciences, Biopharmaceutical R&D, AstraZeneca, Cambridge, UK.; 3) Department of Cardiology and Angiology, Hannover Medical School, Germany.; 4) Clinic for General and Interventional Cardiology, University Heart Centre, Hamburg, Germany.; 5) National Institute for Health and Welfare, Helsinki, Finland.; 6) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA.; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

*GDF15* is a stress response cytokine known to be upregulated in cardiovascular diseases and diabetes. Animal model studies have reported knock-out of *GDF15* to result in increased body weight, whereas overexpression leads to reduced body weight and food intake and improved metabolic profiles. *GDF15* has been identified as a potential target for treatment of obesity and related diseases. However, data supporting its causal role in human disease is sparse (O’Rahilly 2017). Here, we undertake an unbiased assessment of phenotypic and genetic associations of *GDF15* by utilizing a well characterized, Finnish population-based cohort with baseline plasma *GDF15* concentrations (n=6610).

Associations between *GDF15* levels and 677 disease endpoints and 96 quantitative biomarkers were examined. *GDF15* was significantly associated with 37 diseases and 33 biomarkers ( $p < 6.5 \times 10^{-5}$  multiple test correction threshold). The most significant association was found for all-cause mortality (OR=1.8,  $p=9.8 \times 10^{-27}$ ) and mortality due cardiac causes (OR=1.8,  $p=7.7 \times 10^{-14}$ ). A strong association was found for type 2 diabetes (OR=1.5,  $p=2.2 \times 10^{-9}$ ) with both incident and prevalent cases showing increased *GDF15* levels. Elevated *GDF15* levels were detected in cardiovascular diseases (OR=1.2,  $p=5.6 \times 10^{-6}$ ; major coronary heart disease and stroke OR=1.3,  $p=1.6 \times 10^{-9}$ ; atherosclerosis OR=1.7,  $p=1.3 \times 10^{-10}$ ; hypertension OR=1.2,  $p=3.1 \times 10^{-5}$ ) and key components of metabolic syndrome (triglycerides  $b=0.1$ ,  $p=2.4 \times 10^{-22}$ ; fasting insulin  $b=0.1$ ,  $p=1.2 \times 10^{-13}$ ; waist-to-hip ratio  $b=0.1$ ,  $p=7.0 \times 10^{-10}$ ). Associations were also observed for chronic kidney disease (OR=2.5,  $p=6.8 \times 10^{-8}$ ), pneumonia (OR=1.3,  $p=2.7 \times 10^{-8}$ ), schizophrenia (OR=1.8,  $p=9.8 \times 10^{-6}$ ) as well as biomarkers of sepsis and inflammation (Mid regional pro-adrenomedullin  $b=0.2$ ,  $p=1.2 \times 10^{-91}$ ; C-reactive protein  $b=0.2$ ,  $p=4.6 \times 10^{-64}$ ; hepatocyte growth factor  $b=0.2$ ,  $p=4.2 \times 10^{-43}$ ) supporting role for *GDF15* in stress adaptive responses. Contrary to the proposed role of *GDF15* in cancer we found no associations with cancer phenotypes.

To understand the genetic factors regulating plasma *GDF15* levels, we performed GWAS of *GDF15* (n=5817). We identified 100kb locus on chromosome 19, around the *GDF15* gene, strongly associated with *GDF15* levels (top  $p=4.6 \times 10^{-83}$ ) and fine-mapping revealed 4 putative causal variants (rs16982345, rs1054221, rs1059369, rs189593084). Our results suggest that *GDF15* has a role in stress adaptive

responses in several diseases and genetic factors play a significant role in regulation of *GDF15* levels in blood.

## PgmNr 2294: Accurate estimation of SNP-heritability from biobank-scale data irrespective of genetic architecture.

### Authors:

K. Burch<sup>1</sup>; K. Hou<sup>2,3</sup>; A. Majumdar<sup>3</sup>; H. Shi<sup>1,4</sup>; N. Mancuso<sup>3</sup>; Y. Wu<sup>5</sup>; S. Sankararaman<sup>1,5,6,7</sup>; B. Pasaniuc<sup>1,3,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA; 2) College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, China; 3) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA; 4) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 5) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA; 6) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA; 7) Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA

The proportion of phenotypic variance attributable to the additive effects of a given set of SNPs (i.e. SNP-heritability) is a fundamental quantity in the study of complex traits. Recent works have shown that existing methods to estimate genome-wide SNP-heritability often yield biases when their assumptions are violated. While various approaches have been proposed to account for frequency- and LD-dependent genetic architectures, methods based on REML are highly computationally intensive when applied to datasets such as the UK Biobank ( $N > 400K$ ), and it is currently unknown whether methods based on summary statistics produce accurate estimates of total SNP-heritability. Thus, it remains unclear which estimates of SNP-heritability computed from biobank-scale GWAS are reliable.

Here, we describe an approach to estimate genome-wide SNP-heritability accurately from biobank-scale data irrespective of the underlying genetic architecture of the trait, without specifying a heritability model or stratifying SNPs by minor allele frequency (MAF) and/or LD. We use theoretical justifications coupled with extensive simulations starting from real genotypes from the UK Biobank ( $N = 337K$  unrelated white British individuals) to show that, unlike existing methods, the accuracy of our closed-form estimator is invariant to the strength of coupling between effect size and genomic features such as MAF/LD. Across 126 simulated genetic architectures with different parametric forms of the per-SNP effect sizes, the maximum bias we observe with our estimator is 2% of the simulated SNP-heritability whereas existing methods yield biases between -64% and 28%.

We apply our approach to estimate SNP-heritability for 22 complex traits and diseases in the UK Biobank and show that, consistent with our results in simulations, existing biobank-scale methods yield estimates up to 30% different from our theoretically-justified approach. Across 18 of the traits with SNP-heritability estimates greater than 0.05, estimates from stratified LD score regression (Finucane et al. 2015, Gazal et al. 2017) and SumHer (Speed & Balding 2019) differ from our estimates by a median of -9% and 11%, respectively. Overall, our results show that while existing methods can yield biases, for the purpose of estimating total SNP-heritability, most methods are relatively robust to different genetic architectures.

## **PgmNr 2295: Overfitting in polygenic risk score analyses: Exploring the impact of sample overlap and first degree relatives.**

**Authors:**

S. Medland

[View Session](#) [Add to Schedule](#)**Affiliation:** Psychiatric Genetics, QIMR Berghofer, Brisbane, Queensland, Australia

---

Independence of the discovery and target samples is essential to avoid estimation bias due to overfitting within polygenic risk score analyses. Although this is a well-known methodological issue, in recent years perceptions have arisen that this may not matter with biobank type discovery samples when the overlap is very small and few authors have considered the impact of relatedness across the discovery and target samples. To empirically examine these issues a discovery sample of ~340,000 individuals were extracted from the UK Biobank (app. 25331) and GWAS were conducted for a continuous (height) and a binary trait (day-time sleepiness) and for random continuous and binary traits. Polygenic risk scores were calculated and polygenic risk score analyses were conducted using target samples comprising randomly selected baseline samples of 2,000, 5,000 and 10,000 individuals who were not in the discovery sample. To examine overfitting due to non-independence, the target samples were spiked with 5, 10, 50, 100 and 200 individuals who had been part of the discovery sample or their first degree relatives. Sample size was maintained between baseline and spiked samples and 1,000 replicates were performed. As expected degree of overfitting increased with degree of overlap, decreased with total target size and also arose due to non-independence due to relatives. Overfitting was present with as few as 5 overlapping individuals (0.001% of the initial discovery sample) confirming the importance of independence for accurate estimates of variance explained. The impact of non-independent controls (in the absence of overlapping cases) will also be presented.

## PgmNr 2296: PheWAS with a comprehensive health check-up database across 10,349 Korean population.

### Authors:

D. Kim <sup>1,2</sup>; M. Shivakumar <sup>1</sup>; A. Verma <sup>3</sup>; S. Verma <sup>3</sup>; E.K. Choe <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, Epidemiology & Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA; 3) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Surgery, Seoul National University Hospital Healthcare System Gangnam Center, Seoul, 06236, South Korea

---

The rapid growth of resource of genomics linked to the electronic health records (EHRs) database has increasingly leveraged by phenome-wide association study (PheWAS) to elucidate not only the phenotype – gene associations but also the genetic pleiotropy. Most PheWASs introduce International Classification of Diseases, Ninth Revision (ICD-9) billing codes, to define phenotypes, which was not primarily set up for research purpose. Thus, there are limitations by specificity and accuracy of billing information-based phenotypes. In this study, we analyzed the results from comprehensive medical check-ups and screening in a single healthcare center in the Korean population, as a confirmative phenotype to perform PheWAS. Phenotypic data on 10,239 samples who went through check-ups including physical examination, laboratory test, imaging test (ultrasonography, CT, fundoscopy or MRI), and functional studies (electrocardiography, pulmonary function test or endoscopy) were obtained. The genotyping was done using Affymetrix Axiom KORV1.1-96 Array, customized for the Korean population. After standard quality controls, 9,742 samples were used for PheWAS. In our dataset, total 166 EHR-derived phenotypes (94 binary, 72 continuous) were extracted. We calculated the associations between these health check-up phenotypes and 548,762 common frequency SNPs with minor allele frequency > 0.01. We found many SNPs that replicate the previous reports from GWAS as well as novel SNPs associated with depression, tumor markers, skeletal muscle mass, or vitamin D. Among the significant associations, there were novel PheWAS associations potentially driven by pleiotropy. One of the most meaningful findings as potential pleiotropy was the cross-phenotype associations with a variant near SLC19A1 gene for depression, vitamin D3, obesity, liver function, diabetes, and hematologic trait. SNPs associated with tumor markers such as carcinoembryonic antigen had a cross-phenotype associations with other clinical laboratory measures. Using the phenotypes defined by the results during the comprehensive check-up, we were able to identify significant links between several diseases. This is the first large-scale single-cohort PheWAS study using the comprehensive health check-up database in the Asian population. By using this database, it may enable us to overcome some of the limitations of Phecode and to discover novel disease-disease associations.



## PgmNr 2297: Genetic architecture of phenome-wide latent factors in the UK Biobank.

### Authors:

C.E. Carey<sup>1,2</sup>; R. Shafee<sup>2,3</sup>; D.S. Palmer<sup>1,2</sup>; L.E. Abbott<sup>1</sup>; A.L. Elliott<sup>1,2,4,5,6</sup>; D. King<sup>1</sup>; A. Ganna<sup>2,6</sup>; D.P. Howrigan<sup>1,2</sup>; C. Churchhouse<sup>1,2</sup>; R.K. Walters<sup>1,2</sup>; B.M. Neale<sup>1,2</sup>; E.B. Robinson<sup>1,2,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA.; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Department of Medicine, Harvard Medical School, Boston, MA; 5) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 6) Institute for Molecular Medicine Finland, Helsinki, Finland; 7) Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA

---

Large-scale biobanks and electronic health records can offer unprecedented insight into the genetic architecture of a wide range of traits in the general population. However, the high dimensionality of phenotypic data, with thousands of traits measured per individual, can make it difficult to interpret genetic results. Here we derive and genetically characterize latent phenotypic factors representative of the spectrum of traits measured in UK Biobank (UKB).

Beginning with 4203 traits, we identified individuals and traits with low missingness (i.e., ~10%) in order to limit association with survey structure. We performed exploratory factor analysis in this core dataset of 33,860 individuals and 730 items and confirmed the fit of the initial model in an additional low-missingness group (N=8,465). The final model includes 36 stable, interpretable factors accounting for ~30% of total phenotypic variance and spanning a wide range of physical (e.g., body size, general pain), behavioral (e.g., neuroticism, smoking), and lifestyle (e.g., education, urbanicity) dimensions.

To investigate the genetic architecture of these latent factors, we performed a genome-wide association study (GWAS) of each factor, weighted by factor-specific missingness per individual, in the full European ancestry subset (N=361,144). Using LD score regression, we observe a roughly twofold increase in median common variant heritability for the derived factors compared to that of the individual items. We also evaluate whether each latent factor is associated with an increased burden of rare coding variants, especially in loss-of-function intolerant genes, in exome sequencing of UKB.

We contextualize these factors by estimating genetic correlation to psychiatric and somatic disorders studied outside of UKB. In preliminary results, genetic correlations confirm expected associations (e.g.,  $r_g=0.58$  for coronary artery disease and the heart attack factor,  $r_g=-0.61$  for ADHD and the education factor) and also reveal a surprisingly broad, diffuse pattern of genetic correlation between common psychiatric disorders (i.e., ADHD and major depressive disorder) and phenotypic factors.

Overall, these results suggest that phenotypic factor analysis enhances interpretability, boosts power for heritability analyses, and can yield meaningful reduction in dimensionality to drive the next generation of genetic studies.

## PgmNr 2298: Population-specific non-synonymous variants in *PRSS53* influence morphology of human scalp hair.

### Authors:

Q. Ayub<sup>1,2</sup>; K. Rajandran<sup>1</sup>; F. Wahyud<sup>1</sup>; M. Szpak<sup>3</sup>; J.K. Dhaliwal<sup>4</sup>; Y. Xue<sup>3</sup>; T.Y. Ying<sup>5</sup>; S. Rahman<sup>1,2</sup>; C. Tyler-Smith<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Science, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia.; 2) Monash University Malaysia Genomics Facility, Tropical Medicine and Biology Multidisciplinary Platform, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia.; 3) Wellcome Sanger Institute, Hinxton, CB10 1SA, United Kingdom.; 4) School of Information Technology, Monash University Malaysia, 47500 Bandar Sunway, Selangor Darul Ehsan, Malaysia.; 5) Saw Swee Hock School of Public Health, National University of Singapore, Singapore.

Fine-Mapping of Adaptive Variation (*FineMAV*) of 1000 Genomes Project continental populations has previously identified two population-specific high-frequency non-synonymous transitional mutations, rs201075024 (g.31087679C>T; p.Gly34Ser) and rs11150606 (g.31087690T>C; p.Gln30Arg), in the South and East Asian continental populations, respectively. These mutations are 10 bp apart and lie in the third exon of serine protease 53 (*PRSS53*) on human chromosome 16. The gene has been associated with curly hair in knock-out animal models and with hair shape in genome-wide association analysis of admixed Latin Americans. We tabulated *FineMAV* scores for these variants in published high-coverage whole-genome sequenced population samples from Singapore and from the Han Chinese and observed high frequencies for rs201075024 in the Indian and rs11150606 in the Malay populations from Singapore. We also examined the genotype and allele frequencies of these variants in 98 additional individuals from Malaysia using PCR amplification and Sanger sequencing. Subsequently, we used scanning electron microscopy to examine phenotypic differences in the shape, thickness and cross section of scalp hair in a subset of the Malaysian samples (n= 25). In Malaysians, the derived allele frequency was 0.75 for rs201075024 and 0.15 for rs11150606. Several Malaysian individuals (n = 15) were heterozygous for both variants and this has not been reported previously. A Multivariate ANOVA (MANOVA) model was used to compare the two *PRSS53* variants and ethnicity to the measures obtained with electron microscopy. The results show that the number of derived alleles for both SNPs significantly affected the shorter cross-sectional diameter of scalp hair (p = 0.005) when individuals with 0, 1 or 2 copies of the derived alleles were compared. However, the effects of the two SNPs were in opposite directions. The most extreme variation in the average shorter diameter was observed between the homozygous derived rs201075024 (South Asian) genotype (Mean  $\pm$  SD = 39.42  $\pm$  5.49  $\mu$ m) and homozygous derived rs11150606 (East Asian) genotype (81.01  $\pm$  4.58  $\mu$ m). These results suggest an opposing role for these two variants in shaping the cross-sectional diameter of scalp hair and further studies will be required to elucidate the mechanism of action of these derived alleles and their interaction with other loci such as the ectodysplasin A receptor, *EDAR*, that has also been associated with hair thickness in East Asians.

## PgmNr 2299: Genome-wide association studies for metabolic traits in the Ryukyu populations.

### Authors:

M. Matsunami <sup>1</sup>; M. Imamura <sup>1,2</sup>; K. Koganebuchi <sup>3</sup>; R. Kimura <sup>4</sup>; M. Horikoshi <sup>5</sup>; C. Terao <sup>6</sup>; Y. Kamatani <sup>6</sup>; H. Ishida <sup>4</sup>; S. Maeda <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Nishihara-cho, Okinawa, Japan; 2) Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Nishihara-cho 903-0215, Japan; 3) Advanced Medical Research Center, Faculty of Medicine, University of the Ryukyus, Nishihara-cho 903-0215, Japan; 4) Department of Human Biology and Anatomy, Graduate School of Medicine, University of the Ryukyus, Nishihara-cho 903-0215, Japan; 5) Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan; 6) Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan

---

Genome-wide association studies (GWAS) have identified more than several hundreds of variants for individual common diseases or quantitative traits in various ethnic groups, including Japanese. Previous researches have shown that Japanese populations are genetically divided into two clusters, one is the Hondo cluster, which is a major Japanese cluster. The other is the Ryukyu cluster including individuals originally from the Ryukyu Archipelago, which is a chain of Japanese islands that stretch southwest from Kyushu to Taiwan, and is composed of dozens of islands. Because previous GWAS have focused on people belong to the Hondo cluster, there are few genetic information for individuals belong to the Ryukyu cluster. To obtain the genetic information for people living in the Ryukyu Archipelago, we have been attempting to establish a bio-resource, Okinawa Bio-information Bank, consisting of genomic DNA and clinical information from ~13,000 participants living in the Ryukyu Archipelago. To understand the regional population structure, we first obtained genome-wide single nucleotide variations (SNVs) data using Asian Screening Array (ASA, Illumina, CA, U.S.A.) for ~4,000 individuals, living in Ryukyu islands, including Miyako, Kumejima and Okinawa islands. Interestingly, results of principal component and admixture analyses indicated distinct clusters of each island group in the Ryukyu Archipelago. Subsequently, GWAS for metabolic traits, including serum uric acid (sUA), low density lipoprotein cholesterol (LDL-C), triacylglyceride (TG) and high density lipoprotein cholesterol (HDL-C), were conducted, and we identified several loci significantly associated with these metabolic traits ( $p < 5 \times 10^{-8}$ ); sUA ( $n = 2,945$ ); rs57633992 (Chromosome (Ch) 11, *NRXN2-SLC22A12*,  $p = 3.3 \times 10^{-18}$ ), rs6822046 (Ch 4, *SPP1*,  $p = 7.5 \times 10^{-9}$ ), TG ( $n = 2,863$ ); rs662799 (Ch 11, *ZPR1-APOA5*,  $p = 6.7 \times 10^{-23}$ ), rs78358410 (Ch 8, *LPL*,  $p = 6.8 \times 10^{-10}$ ), HDL-C ( $n = 2,879$ ) rs183130 (Ch 16, *CETP*,  $p = 3.1 \times 10^{-19}$ ), rs113932726 (Ch 11, *ZPR1-APOA5*,  $p = 2.2 \times 10^{-16}$ ). In summary, we observed the genetic differentiations among the island groups of the Ryukyu Archipelago. These genetic differences may reflect ancient migrations and admixtures of the people living in the Ryukyu Archipelago. We also identified several loci significantly associated with several metabolic traits in the Ryukyu population.

## PgmNr 2300: Analysis of runs of homozygosity segments in the Kuwaiti population.

### Authors:

S. John; P. Hebbar; A. Channanath; F. Al-Mulla; T. Thanaraj

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetics & Bioinformatics, Dasman Diabetes Institute, 15462, Dasman, Kuwait, Kuwait

---

**Background:** Runs of Homozygosity (ROH) are contiguous, homozygous segments of the genome having identical haplotypes inherited from the parents. The number and length of ROH provide useful information on the demographic history of populations and individuals, while the homozygosity burden could be a powerful tool to unravel the genetic architecture of complex diseases. Admixed populations carry the fewest ROH regions whereas consanguineous communities harbor very long ROH. Kuwait and the other Middle Eastern countries adorn a strategic place in the human migration route out of Africa. The Middle Eastern populations, characterized by a high proportion of consanguinity, becomes a preferential target for the study of nature and distribution of ROH. The aim of this study is to evaluate the patterns of Long ROH in the Kuwaiti population and characterizing regions marked by a high (hotspots) or low (coldspots) degree of homozygosity.

**Methods:** We carried out this analysis on a cohort of 291 healthy Kuwaiti individuals, whose exomes were sequenced. Computational detection and classification of ROH was done. We compared the number and length of ROH in Kuwait exomes with that of other global populations. Furthermore, we characterized the hotspot and coldspot regions in the shared ROH regions between individuals and annotated these regions for potentially deleterious variations.

**Results:** By comparing ROH in Kuwaiti exomes with global populations, in terms of numbers, we found that they lie somewhere in the middle between groups that are highly inbred and diverse. Lengthwise, they outdo most of other populations highlighting their consanguineous nature. Analysis of variants in the hotspot regions led to the identification of genes associated with metabolic disorders such as obesity, type 2 diabetes and triglycerides, and autosomal recessive disorders which are highly prevalent in this region.

**Conclusion:** Our study shows that exome sequencing, even on limited number of individuals, combined with ROH mapping can be a powerful tool to identify some of the potentially deleterious disease variants in Middle Eastern populations, which are otherwise poorly represented in the world genetic map.

## PgmNr 2301: GATTC1 haplotype in *CD36* gene may predict the risk of developing T2DM associated BMI and hypertension.

### Authors:

S. Gautam; M.D. Kaleem Ahmad; A. Ali Mahdi

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Biochemsistry, King George's Medical University, Luknocw, India

---

Type II diabetes mellitus (T2DM) is now considered to be a serious disease that is always associated with long-term and life-threatening complications, attributed deaths and economic burden to nations. Several Single nucleotide polymorphisms (SNPs) in *CD36* gene have been found to be associated with metabolic syndrome and HDL metabolism, both predictors of the risk of heart disease and T2DM. We studied eleven SNPs in entire *CD36* gene and their association with 100 each of control subjects and T2DM. The haplotypic analysis of few significant SNPs was carried out in individuals from families with diabetic history in order to evaluate its utility in disease prediction. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) was used for genotyping. Ten families with a family history of diabetes were identified and blood samples were collected from as many family members as possible. Genotyping of three SNPs viz. rs1761667 (G>A) in exon 1 A, rs3211938 (T>G) in exon 10 and rs3212018 (16 bp del) in exon 14 was performed in all samples. In our study on North Indian population SNP G>A (rs1761667) genotypes showed a highly significant association ( $P<0.001$ ) and 1264T>G (rs3211938) showed mild significant for T2DM ( $P=0.045$ ). Moreover, individuals having a 'GATTC1' haplotype might be at risk of developing T2DM ( $P<0.001$ ). In addition of it, 'A' 'G' and 'G' alleles of SNPs rs1761667 (G>A), rs3211938 (T>G) and rs1984112 (T>G) respectively tend to have increased BMI in families. Though their lipid profile were found to be normal in both patients as well as innondiabetic individuals of families but BMI and slightly enhanced blood pressure of Nondiabetic individuals were found to be high which is suggested that they may develop diabetes in near future. Such studies may be helpful for disease prediction in individuals at risk of T2DM and could be used as a genetic marker.

## PgmNr 2302: Comparison of genotypic divergence with phenotypic divergence for type 2 diabetes and related traits among continental ancestry groups.

### Authors:

R.L. Hanson<sup>1</sup>; A. Shuldiner<sup>2</sup>; W.C. Hsueh<sup>1</sup>; C. Van Hout<sup>2</sup>; M. Sinha<sup>1</sup>; L.J. Baier<sup>1</sup>; W.C. Knowler<sup>1</sup>; Regeneron Genetics Center

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Diabetes Epidemiology Clin Res, NIDDK, Phoenix, Arizona.; 2) Regeneron Genetics Center, Tarrytown, New York

---

Prevalence of type 2 diabetes (T2D) differs among ancestry groups, and many hypotheses invoke differential natural selection to account for these differences. To assess the potential role of differential natural selection across major continental ancestry groups for T2D and related traits, we analyzed genetic and phenotypic differences among 734 individuals from urban Phoenix (none of whom was more closely related to another than 3<sup>rd</sup> degree relatives), including 83 African Americans (AA), 523 American Indians (AI) and 128 European Americans (EA). Participants were not recruited based on T2D status or other traits. Body mass index (BMI) was calculated, and diabetes was diagnosed by a 75g oral glucose tolerance test. In those with normal glucose tolerance (n=434), fasting insulin (FI) and 30-minute post-load insulin ( $I_{30}$ ), adjusted for 30-minute glucose, were taken as measures of insulin resistance and secretion, respectively. Whole exome sequencing was performed using HI-SEQ (Illumina, San Diego, CA) resulting in 97,125 common (average minor allele frequency >5%) variants;  $F_{ST}$  was calculated across all markers as a measure of genetic divergence among populations. The phenotypic divergence index,  $P_{ST}$ , was also calculated from the phenotypic differences and heritability (which was estimated from genetic relatedness calculated empirically across all markers in 766 AI participants prior to the exclusion of close relatives). Under evolutionary neutrality, the expectation is  $P_{ST}=F_{ST}$ , while for traits under differential selection  $P_{ST}$  is expected to be much greater than  $F_{ST}$ . A bootstrap procedure was used to test the hypothesis  $P_{ST}=F_{ST}$ . With adjustment for age and sex, prevalence of T2D was 34.1% in AI, 12.4% in AA and 10.4% in EA ( $P=2.9 \times 10^{-10}$  for difference among groups). Mean BMI was 35.5, 32.4 and 31.8 kg/m<sup>2</sup> respectively ( $P=1.9 \times 10^{-7}$ ). Mean FI was 64.7, 47.3 and 44.0 pmol/l ( $P=9.2 \times 10^{-5}$ ), while mean  $I_{30}$  was 567.8, 539.0 and 349.6 pmol/l ( $P=5.7 \times 10^{-8}$ ).  $F_{ST}$  across all markers was 0.146, while  $P_{ST}$  for liability to T2D, adjusted for age and sex, was 0.147 ( $P=0.99$  for difference with  $F_{ST}$ ).  $P_{ST}$  was 0.094 for BMI ( $P=0.31$ ), 0.092 for FI ( $P=0.37$ ) and 0.219 ( $P=0.40$ ) for  $I_{30}$ . These analyses suggest that while T2D and related traits differ significantly among continental ancestry groups, the differences are consistent with expectations based on heritability and genetic distances; thus, differential natural selection may not be necessary to explain these differences.



## PgmNr 2303: Evidence of natural selection in Samoans is associated with BMI and the immune system.

### Authors:

E.M. Russell <sup>1</sup>; D.N. Harris <sup>2,3,4</sup>; J.C. Carlson <sup>5</sup>; J.Z. Zhang <sup>5</sup>; N.L. Hawley <sup>6</sup>; H. Cheng <sup>7</sup>; T. Naseri <sup>8</sup>; M. Sefuiva Reupena <sup>9</sup>; I.Y. Chen <sup>10</sup>; D.C. Rao <sup>11</sup>; A.C. Hsiung <sup>12</sup>; L.-M. Chuang <sup>13,14</sup>; W. Sheu <sup>15,16,17,18</sup>; D. Darbar <sup>19</sup>; R. Deka <sup>7</sup>; T.D. O'Connor <sup>2,3,4</sup>; S.T. McGarvey <sup>20,21</sup>; D.E. Weeks <sup>1,5</sup>; R.L. Minster <sup>1</sup>; TOPMed Population Genetics Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 2) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; 3) Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 4) Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 5) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 6) Department of Epidemiology (Chronic Disease), School of Public Health, Yale University, New Haven, CT; 7) Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, OH; 8) Ministry of Health, Government of Samoa, Apia, Samoa; 9) Lutia i Puava ae Mapu i Fagalele, Apia, Samoa; 10) The Lundquist Institute for Biomedical Innovation at Harbor UCLA, Pediatrics, Genomic Outcomes, Los Angeles, CA; 11) Washington University in St. Louis, School of Medicine, Division of Biostatistics, St. Louis, MO; 12) Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan; 13) Division of Endocrinology & Metabolism, Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 14) Institute of Preventive Medicine, School of Public Health, National Taiwan University, Taipei, Taiwan; 15) Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan; 16) School of Medicine, National Yang-Ming University, Taipei, Taiwan; 17) School of Medicine, National Defense Medical Center, Taipei, Taiwan; 18) Institute of Medical Technology, National Chung-Hsing University, Taichung, Taiwan; 19) Division of Cardiology, Department of Medicine, University of Illinois College of Medicine, Chicago, IL; 20) International Health Institute, Department of Epidemiology, School of Public Health, Brown University, Providence, RI; 21) Department of Anthropology, Brown University, Providence, RI

---

Samoa is part of the Polynesian Islands which were settled 3,000 years ago. Challenges faced by the settlers of Samoa may have led to positive adaptive selection when food was limited, especially for energy efficient alleles. Selection may also have occurred as a result of exposure to infectious diseases during migration or contact with outside populations. Genetic signatures of adaptation have been identified in other population isolates so, given Samoa's genetic isolation, they are likely to be identified in Samoa as well. We performed a SNP-based population branch statistic analysis (PBS) to simultaneously compare the genetic population differentiation between Samoa and a closely-related in-group and a distantly-related out-group. 419 unrelated Samoans were compared to 1,432 unrelated Taiwan Chinese and 1,083 unrelated Europeans. All individuals were sequenced as part of the TOPMed Program. We also tested for evidence of selective sweeps in 419 unrelated Samoans using  $n_S$ , a single-population haplotype-based analysis. Since  $n_S$  values are dependent on allele frequency,  $n_S$  was calculated once with  $MAF > 0.01$  and once with  $MAF > 0.05$ . The top 25 variants from the PBS,

$nS_{L(0.01)}$ , and  $nS_{L(0.05)}$  analyses were explored further. We performed gene ontology (GO) enrichment analysis on the PBS results. The top 25 variants from the PBS,  $nS_{L(0.01)}$ , and  $nS_{L(0.05)}$  analyses included 29 variants in/near genes (*ADH5P2*, *ARHGEF28*, *CLN3*, *GRIN2A*, *KIAA1217*, *LOC101929563*, *LRFN5*, *OFCC1*, *PDZD8*, *RNU5F-1*, *SCN3A*, *SEMAB*, and *TIGAR*) that have been associated with BMI and 17 variants in/near genes (*ARL17B*, *B4GALNT4*, *C22orf34*, *LOC101929563*, *MIR4289*, *MROH1*, *PFKFB3*, *PLXNC1*, and *PPP2R2D*) that have been associated with the immune system or respiratory disease. Genes (*ABC11*, *ARL17B*, *MCU*, and *TMEM132C*) which have been identified in other genome-wide scans of selection were also identified. Of the 10 gene sets that were significantly enriched in the GO enrichment analysis, 4 were associated with immune system-related biological processes. This suggests that variants associated with an energy efficient metabolism and the immune system may have been selected for in Samoans. Selection for BMI related alleles may have occurred during times of food scarcity. Selection for immune related alleles may have occurred in response to exposure to an infectious disease at some point in Samoan population history.



## PgmNr 2304: Exploring health, disease, and diet in Ancient Rome through paleogenomics.

### Authors:

H. Moots <sup>7</sup>; M. Antonio <sup>8</sup>; Z. Gao <sup>1</sup>; M. Lucci <sup>2</sup>; F. Candilio <sup>3</sup>; A. Sperduti <sup>4,5</sup>; L. Bondioli <sup>4</sup>; A. Nava <sup>4,6</sup>; M. Rubini <sup>9</sup>; S. Sawyer <sup>10</sup>; A. Coppa <sup>2</sup>; R. Pinhasi <sup>10</sup>; J. Pritchard <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics Department, Stanford University, Stanford, CA.; 2) Dipartimento di Biologia Ambientale, Sapienza Università di Roma; 3) Soprintendenza Archeologia, Sardinia; 4) Servizio di Bioarcheologia, Museo delle Civiltà, Roma, Italy; 5) Università L'Orientale Napoli, Italy; 6) DANTE Laboratory for the study of Diet and Ancient Technology, Sapienza Università di Roma; 7) Anthropology Department, Stanford University, Stanford, CA.; 8) Biomedical Informatics, Stanford University, Stanford, CA.; 9) Università di Foggia, SABAP-LAZ Ministero dei Beni e delle Attività Culturali; 10) Department of Anthropology, University of Vienna

---

Daily life in the Roman Empire has been studied in extensive depth across the humanities, social and natural sciences. Here we draw on this rich body of evidence to contextualize whole-genome sequence data from three well-documented Roman sites - Portus Romae, the port of Rome on the Tyrrhenian Sea, Crypta Balbi within the city of Rome, and Villa Magna, in the countryside south of Rome. This approach allows us to explore genetic changes in the context of changing mobility patterns, cultural practices and environmental conditions. We integrate existing isotopic data, dental calculus studies, and paleopathological evidence to better understand and interpret frequency changes we observe in functional alleles related to health and diet, such as malarial resistance and lactase persistence.

## PgmNr 2305: The human-specific *BOLA2* duplication modifies iron homeostasis and anemia predisposition in chromosome 16p11.2 autism patients.

### Authors:

G. Giannuzzi<sup>1</sup>; P.J. Schmidt<sup>2</sup>; E. Porcu<sup>1,3</sup>; G. Willemin<sup>1</sup>; K.M. Munson<sup>4</sup>; X. Nuttle<sup>4</sup>; R. Earl<sup>5</sup>; J. Chrast<sup>1</sup>; K. Hoekzema<sup>4</sup>; D. Risso<sup>4</sup>; K. Mannik<sup>1</sup>; P. De Nittis<sup>1</sup>; E.D. Baratz<sup>6</sup>; Y. Herault<sup>7</sup>; X. Gao<sup>8</sup>; C.C. Philpott<sup>6</sup>; R.A. Bernier<sup>5</sup>; Z. Kutalik<sup>3,9</sup>; M.D. Fleming<sup>2</sup>; E.E. Eichler<sup>4,10</sup>; A. Reymond<sup>1</sup>; 16p11.2 Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland; 2) Department of Pathology, Boston Children's Hospital, Boston, MA 02115, USA; 3) Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; 4) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; 5) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195, USA; 6) Genetics and Metabolism Section, Liver Diseases Branch, NIDDK, National Institutes of Health, Bethesda, MD 20892, USA; 7) Institut de Génétique et de Biologie Moléculaire et Cellulaire, 67400 Illkirch, France; 8) Model Animal Research Center, Collaborative Innovation Center for Genetics and Development, Nanjing Biomedical Research Institute, Nanjing University, Nanjing, 210061 China; 9) University Center for Primary Care and Public Health, 1010 Lausanne, Switzerland; 10) Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

---

Recurrent 600 kbp copy number variations with breakpoints BP4 and BP5 at human chromosome 16p11.2 are among the most frequent genetic causes of neurodevelopmental and psychiatric disorders. These pathogenic rearrangements are mediated by human-specific duplications through non-allelic homologous recombination. These duplications appeared at the beginning of the modern human lineage, rapidly increased in frequency, and are now nearly fixed in humans, suggesting that their expansion has a possible evolutionary advantage that outweighs the accompanying chromosomal instability. They are copy number polymorphic and include 3 to 8 copies of *BOLA2*, a gene involved in the maturation of cytosolic iron-sulfur proteins. To investigate the potential advantage provided by the rapid increase in the number of copies of *BOLA2*, we assessed hematological traits and anemia prevalence in 379,385 controls and individuals who have lost or gained copies of *BOLA2*: 89 chromosome 16p11.2 BP4-BP5 deletion and 56 reciprocal duplication carriers in the UK Biobank. We found that the 16p11.2 deletion is strongly associated with anemia (18/89 carriers, 20%,  $P=4e-7$ , OR=5), particularly iron-deficiency anemia. We observed similar enrichments in two clinical 16p11.2 deletion cohorts, with 6/63 (10%) and 5/19 (26%) unrelated individuals with diagnosis of anemia, low serum iron, or low blood hemoglobin. Upon stratification by *BOLA2* copy number, we found an association between low *BOLA2* dosage and anemia ( $P=2e-3$ ). In particular, 6/14 (43%) individuals with three copies were anemic. In parallel, we analyzed hematological traits in mice carrying the 16p11.2 orthologous deletion or duplication that include *Bola2*, as well as *Bola2*<sup>+/-</sup> and *Bola2*<sup>-/-</sup> animals. Consistent with human data, the 16p11.2 deletion mouse model and *Bola2*-deficient mice showed early evidence of iron deficiency, including a mild decrease in hemoglobin, lower plasma iron, smaller red blood cells, and an increased red blood cell zinc protoporphyrin to heme ratio. Our results indicate that *BOLA2* participates in iron homeostasis in

*vivo*. Its rapid expansion through segmental duplication might have evolved to protect humans against iron deficiency as our species successfully expanded its ecological range at the cost of increased predisposition to rearrangements associated with autism.

## PgmNr 2306: *HLA-G* gene associated with vitiligo in a Brazilian population sample.

### Authors:

M.L.G. de Oliveira <sup>1</sup>; L.C. Veiga-Castelli <sup>1</sup>; A.L.E. Pereira <sup>1</sup>; G. Debortoli <sup>1</sup>; L. Marcorin <sup>1</sup>; N.C.A. Fracasso <sup>1</sup>; G. Valle-Silva <sup>2</sup>; A.S. Souza <sup>3</sup>; J.D. Massaro <sup>4</sup>; A.L. Simões <sup>1</sup>; A. Sabbagh <sup>5</sup>; R.N. Cardili <sup>4</sup>; E.A. Donadi <sup>4</sup>; E.C. Castelli <sup>3</sup>; C.T. Mendes-Junior <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto-SP, Brazil; 2) Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto-SP, Brazil; 3) São Paulo State University (UNESP), Molecular Genetics and Bioinformatics Laboratory, School of Medicine, Botucatu, State of São Paulo, Brazil; 4) Departamento de Clínica Médica, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto-SP, Brazil; 5) UMR 216 MERIT IRD, Faculté de Pharmacie de Paris, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

---

Vitiligo is an autoimmune disease characterized by skin depigmentation due to failure of melanocytes function. It is the most frequent cause of depigmentation worldwide, affecting people from various ethnic backgrounds, notwithstanding, with no satisfactory treatment available. Genetic association studies have described 50 loci associated with vitiligo, many of them related to other autoimmune diseases. Classic *HLA* loci have already been associated with this condition, but no study investigated the association of *HLA-G* with vitiligo considering its whole genic region (exonic, intronic and regulatory regions – 5'URR and 3'UTR). We have identified the genetic diversity of *HLA-G* by next-generation sequencing in a population sample composed of 50 vitiligo cases and 393 healthy control individuals from São Paulo State, Brazil. Individual ancestry was determined by the SNPforID 34-plex ancestry informative marker (AIM) SNP panel included in this same assay. DNA libraries were prepared using the Haloplex Target Enrichment System (Agilent Technologies) and sequenced at the MiSeq platform (Illumina). CutAdapt, hla-mapper, and GATK software packages were used for trimming adaptor sequences, alignment and genotype calling, respectively. Missing alleles and haplotypes were inferred by using the PHASE method. Ancestry was estimated by STRUCTURE, considering EUR, AFR, and EAS groups from 1000 Genomes Project as reference. Case-control associations, adjusted by ancestry, were performed using logistic regression at PLINK v.1.9. We detected 105 variation sites distributed along the *HLA-G* locus fitting Hardy-Weinberg expectations ( $p \geq 0.01$ ). Comparing patients with controls, we detected associations with rs1625907\*G ( $p=0.026$ ; OR=1.639; 95%CI 1.061-2.529), and rs9380142\*G ( $p=0.040$ ; OR=1.590; 95%CI 1.020-2.479). The first one, rs1625907\*G, is associated with the G\*01:01:01 and G\*01:03 allele groups. The second one (rs9380142) is located at the 3'UTR, at position +3187, with allele Guanine composing the 3'UTR haplotype previously described as UTR-1. The rs9380142 has already been associated with different pathological conditions, as well as to mRNA stability. This is the first report that highlights the association of rs9380142 with vitiligo, suggesting a possible mechanism in the development of the disease. Financial Support: CAPES - Finance Code 001. CNPq (448242/2014-1), FAPESP (2013/15447-0), USP/COFECUB (2017.1.1337.1.5). M.L.G.O (141908/2016-5); C.T.M.J. (312802/2018-8).



## PgmNr 2307: Higher Native American ancestry proportion is associated with TB progression risk in the Peruvian population.

### Authors:

S. Asgari <sup>1,3</sup>; Y. Luo <sup>1,3</sup>; C. Huang <sup>1</sup>; R. Calderon <sup>2</sup>; L. Lecca <sup>2</sup>; J. Jimenez <sup>2</sup>; R. Yataco <sup>2</sup>; C. Contreras <sup>2</sup>; J. Galea <sup>1</sup>; D. Jones <sup>1</sup>; D.B. Moody <sup>3</sup>; M. Murray <sup>1,2</sup>; S. Raychaudhuri <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Harvard Medical School, Boston, Massachusetts.; 2) Socios En Salud Sucursal Peru, Lima, Peru.; 3) Brigham and Women's Hospital, Boston, Massachusetts.

---

A potential link between Native American ancestry (NAT) and tuberculosis (TB) risk has been extensively studied over the past two centuries (Jones, 2003), but the conclusions studies have been unsatisfactory and controversial due to the lack of adequate control for confounding factors. Here, we investigated the relationship between NAT and TB progression risk in a cohort of 3,980 genotyped admixed Peruvians from Lima. Upon admission of index cases, their household contacts (HHCs) were tested for TB at 2, 6, and 12 months and were recruited as cases if they developed active TB and as controls with latent TB if they were tuberculin skin test positive but did not develop active TB. We also collected extensive socio-demographic information for all participants. We merged our data with the data from Native American (Reich, 2012) and 1000 Genomes project populations to infer ancestry proportions using ADMIXTURE (K=4). NAT was associated with TB progression risk after correction for age, gender, household as a proxy for unmeasured environmental variables, and a genetic relatedness matrix to correct for relatedness and population structure (2,160 cases, 1,820 HHCs,  $p=1.82 \times 10^{-15}$ , OR for 10% increase in NAT ( $OR_{NAT0.1}$ )=1.24 (1.18-1.31)). Correction for 7 additional covariates including African and Asian ancestry, smoking, drinking, malnutrition, BCG vaccination, and socioeconomic status had no effect on this result ( $p=1.09 \times 10^{-6}$ , 1.26 (1.15-1.38)). To better control for factors related to transmission, we restricted our analysis to secondary cases (e.g. HHCs who developed TB during the follow-up, N=203) and their HHCs (N=253). This analysis resulted in an odds ratio similar to the one observed for the whole cohort ( $p=7.78 \times 10^{-4}$ ,  $OR_{NAT0.1}=1.30$  (1.12-1.52)). We next used admixture mapping to search for specific genomic regions that might explain some of this association. No genomic region reached the genome-wide significance threshold ( $< 5.68 \times 10^{-5}$ , lowest observe  $p=1.1 \times 10^{-3}$ ). Our results show that, relative to European ancestry, NAT is associated with TB progression risk. This association is unlikely to be due to confounding factors and is driven, at least in part, by genetics. This effect is most likely to follow a polygenic architecture. It is possible that the long, shared history of Europeans and TB has led to selective pressures that have mitigated TB risk, and that such pressures have not been present in the Americas.

## PgmNr 2308: Paternal age is significantly associated with higher *de novo* mutation rate in rhesus macaques, but not associated with sociality in male offspring.

### Authors:

R.J. Wang<sup>1</sup>; M. Raveendran<sup>2</sup>; J. Capitanio<sup>3</sup>; R.A. Harris<sup>2</sup>; G.W.C. Thomas<sup>1,4</sup>; P. Radivojac<sup>4,5</sup>; J. Rogers<sup>2</sup>; M.W. Hahn<sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Biology, Indiana University, Bloomington, IN; 2) Human Genome Sequencing Center and Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) California National Primate Research Center, University of California, Davis, CA; 4) Dept. of Computer Science, Indiana University, Bloomington, IN; 5) Khoury College of Computer Sciences, Northeastern University, Boston, MA

---

The nature and rate of *de novo* mutation is fundamental to our understanding of genome content, genome evolution and biomedical genetics. Studies in humans show that paternal age is a major factor determining the number of *de novo* mutations. Furthermore, evidence suggests advanced paternal age increases the risk of autism, possibly due to the elevated mutational burden from older fathers. To quantify *de novo* mutation in nonhuman primates and test for paternal age effects on behavioral variation, we generated whole genome sequences for three-generation families of rhesus macaques (*Macaca mulatta*) and used behavioral assessments of 203 male rhesus to ask whether offspring of older fathers differ in sociality compared to those with younger fathers. We sequenced the genomes (Illumina Nova-Seq) of 32 individuals, including 14 probands. We estimate the average mutation rate to be  $0.37 \times 10^{-8}$  per site per generation, with a strong association between paternal age and number of *de novo* mutations per offspring. For each additional year of paternal age offspring exhibit 1.3 additional *de novo* mutations ( $r^2 = 0.75$ ,  $p < 5 \times 10^{-5}$ ). The pedigree structure allowed phasing of new mutations, and 66% were derived from fathers. We found a weak relationship between maternal age and mutation rate, consistent with environmental effects. To investigate the effect of paternal age on social behavior, 8 behaviors involved in general social functioning were quantified across 203 male macaques. Principal component analysis yielded general scores summarizing social function. The first principal component (PC1) captures individual tendencies to interact with females versus males. PC2 captures overall contact with and proximity to both sexes and is correlated with observer ratings of sociability as a broad personality trait. We find no evidence of an effect of paternal age on social functioning. We conclude that, as in humans, *de novo* mutation in macaques is driven primarily by male mutation and that this rate increases with paternal age. The absolute frequency of macaque mutations fits the model of Thomas et al. (*Curr. Biol.* 28: 3193, 2018) which relates per generation mutation with reproductive longevity rather than simple generation length. We find no evidence for an effect of increased paternal age on social behavior among male rhesus macaques, consistent with hypotheses that the increased risk of neurodevelopmental disorder from parental age is not primarily due to *de novo* mutation.

## PgmNr 2309: Environmental and genetic drivers of telomere length variation in ethnically diverse Africans.

### Authors:

M. McQuillan <sup>1</sup>; S.C. Hunt <sup>2</sup>; M.E.B. Hansen <sup>1</sup>; S. Verhulst <sup>3</sup>; T.P. Lai <sup>4</sup>; G.G. Mokone <sup>5</sup>; S.W. Mpoloka <sup>6</sup>; D. Woldemeskel <sup>7</sup>; G. Belay <sup>7</sup>; T. Nyambo <sup>8</sup>; C.C. Abnet <sup>9</sup>; M. Yeager <sup>9</sup>; S.J. Chanock <sup>9</sup>; A. Ranciaro <sup>1</sup>; E. Mbunwe <sup>1</sup>; S. Fan <sup>1</sup>; M. Rubel <sup>1</sup>; W. Beggs <sup>1</sup>; S.M. Williams <sup>10</sup>; A. Aviv <sup>4</sup>; S.A. Tishkoff <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Genetic Medicine, Weill Cornell Medicine, Doha, Qatar; 3) Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, the Netherlands; 4) The Center of Human Development and Aging, New Jersey Medical School, Rutgers, Newark, NJ; 5) Department of Biomedical Sciences, University of Botswana School of Medicine, Gaborone, Botswana; 6) Department of Biological Sciences, University of Botswana, Gaborone, Botswana; 7) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia; 8) Department of Biochemistry, St. Joseph University College of Health Sciences, Dar es Salaam, Tanzania; 9) Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland; 10) Department of Population and Quantitative Health Science, Case Western Reserve University, Cleveland, Ohio, USA

---

Telomeres are repetitive non-coding sequences at the ends of chromosomes that maintain genomic stability. Telomeres progressively shorten at each cell division until a critical length is reached and cell division stops. Thus, telomere length is closely associated with aging and lifespan. In addition, longer telomeres are associated with increased risk of many cancers, while shorter telomeres are associated with elevated cardiovascular disease risk. Finally, telomeres can shorten as a result of many intrinsic and extrinsic factors, including chronic stress and infection status. However, relatively little is known about the genetic architecture underlying telomere length, or the environmental factors that influence telomere loss. Additionally, the evolutionary history of telomere length in humans, and whether selection has influenced telomere dynamics, remain to be explored. Here, we investigate the relationship between telomere length, genetics, and environmental factors in a set of ethnically diverse African people (n=1820) originating from populations in Botswana, Tanzania, Ethiopia, and Cameroon. We find significant variation in telomere length among populations after adjusting for age and sex, with the San hunter-gatherers from Botswana having the longest telomeres, and pastoralists from Cameroon having the shortest telomeres. After accounting for genome-wide ancestry and relatedness among individuals, we find that a large proportion of inter-individual variation in telomere length (>40%) is explained by genetic factors. Finally, telomere length varies significantly with environmental factors across Africa, as we find that altitude, UV-B radiation, and temperature explain small but significant amounts of variation in telomere length after adjustment for age, sex, and genome-wide ancestry. Ongoing work examines whether patterns of polygenic adaptation at genetic loci underlying telomere length vary with environmental factors globally and within Africa. This research will help elucidate the evolutionary forces driving telomere length variation in humans, which will provide insight into the basis of telomere-related disease risk.



## PgmNr 2310: Complete mitogenome sequencing of 605 CEPH family pedigree samples reveals spurious substitutions attributed to cell culture.

### Authors:

C. Marshall <sup>1,2,3</sup>; K. Sturk-Andreaggi <sup>1,2</sup>; J.D. Ring <sup>1,2</sup>; C.R. Taylor <sup>1,2</sup>; S.M. Barritt-Ross <sup>1,2</sup>; W. Parson <sup>3,4</sup>; T.P. McMahon <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory (AFMES-AFDIL), Dover Air Force Base, DE; 2) SNA International, Alexandria, VA, contractor supporting the AFMES-AFDIL; 3) Forensic Science Program, The Pennsylvania State University; 4) Institute of Legal Medicine, Medical University of Innsbruck, Austria

---

The present study evaluated mitochondrial genome sequence variability in the CEPH family pedigree cell lines with the intentions of developing interpretation guidelines for forensic casework. This involved an assessment of the inter-generational substitution rate, which can also inform clinical genetics and human evolutionary research. A total of 605 mitogenomes from 87 maternal lineages were produced using a long-range dual PCR amplification, KAPA HyperPlus library preparation, and Illumina sequencing workflow. Sequence data were analyzed with a minor variant detection threshold of 2% and a minimum read depth of 350. Then the number of point substitutions from mother to child was calculated for each of 516 generational events. A substitution was defined as a complete base change with no indications of heteroplasmy in either the mother or the child, which is unexpected given the multicopy inheritance of mitochondrial DNA in which several dozen are carried through oogenesis. The CEPH data showed 29 substitutions from mother to child, two of which were observed in a single sample, for an inter-generational substitution rate of 5.4%. This is in contrast to the 0% substitution rate reported in a similar study of non-cell line samples from mother-child pairs comprising 39 generational events. Of the 29 substitutions in the CEPH family pedigrees, 13 (45%) appeared to be spurious substitutions, defined as variants that were not observed in any other samples in the lineage and are also rare for the haplogroup. Raising the minor variant detection threshold to 10% increased the inter-generational substitution rate to 23.4%, and raised the number of spurious substitutions to 97, impacting 30 (5%) of the 605 CEPH family pedigree samples. If the same rate of spurious substitutions could be expected from the HGDP cell lines used for human evolutionary studies, this would erroneously speed up the molecular clock. Furthermore, the present results indicate an inflated heteroplasmy occurrence in the CEPH cell lines at an average of 1.85 point heteroplasmies per individual, in contrast to other serum, blood and buccal sample data at ~0.50 point heteroplasmies per individual. Overall, this study highlights the necessity for low-level heteroplasmy detection for accurate mitogenome reporting of cell line sequence profiles, and quantifies the expected rate of spurious substitutions and heteroplasmic sequence artifacts that may be reported from immortalized cell lines.

## PgmNr 2311: Multi-copy nuclear elements of mitochondrial DNA mimic paternal inheritance in humans.

### Authors:

W. Parson<sup>1,2</sup>; H. Niederstaeter<sup>1</sup>; J. Naue<sup>3,4</sup>; R. Koziel<sup>5</sup>; F. Yang<sup>6</sup>; T. Saenger<sup>3,4</sup>; G. Huber<sup>1</sup>; C. Berger<sup>1</sup>; R. Pflugradt<sup>7</sup>; C. Strobl<sup>1</sup>; C. Xavier<sup>1</sup>; M. Volleth<sup>8</sup>; S. Weiss<sup>9,2</sup>; J. Irwin<sup>10</sup>; E. Romsos<sup>11</sup>; P. Vallone<sup>11</sup>; T. Parsons<sup>12</sup>; P. Jansen-Duerr<sup>5</sup>; T. Liehr<sup>13</sup>; P. Lichter<sup>14</sup>; S. Pollak<sup>3,4</sup>; S. Lutz-Bonengel<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Tyrol, Austria; 2) Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Tyrol, Austria; 3) Forensic Science Program, The Pennsylvania State University, University Park, PA, USA; 4) Institute of Forensic Medicine, Medical Center, University of Freiburg, Forensic Molecular Biology, Freiburg, Germany; 5) Faculty of Medicine, University of Freiburg, Freiburg, Germany; 6) Institute for Biomedical Ageing Research, University of Innsbruck, Innsbruck, Austria; 7) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 8) State Investigation Department of Lower Saxony, Hannover, Germany; 9) Magdeburg University Hospital, Institute of Human Genetics, Otto von Guericke University, Magdeburg, Germany; 10) Institute of Experimental and Clinical Pharmacology and Toxicology, University of Freiburg, Freiburg, Germany; 11) DNA Support Unit, FBI Laboratory, Quantico, VA, USA; 12) U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, Gaithersburg, MD, USA; 13) International Commission on Missing Persons, The Hague, Netherlands; 14) Jena University Hospital, Institute of Human Genetics, Friedrich Schiller University, Jena, Germany; 14) German Cancer Research Center, Molecular Genetics, Heidelberg, Germany

---

Mammalian mitochondrial DNA (mtDNA) is generally believed to be inherited through the maternal lineage. A few studies claim paternal mtDNA inheritance mostly in patients suffering mitochondrial myopathies. We report the similar observation of a mixture of two phylogenetically distinct full mitogenomes in healthy individuals. Different tissues from a total of eight maternally related individuals spanning three generations gave to varying degrees mixtures of two mitotypes assigned to haplogroups V and U4c1a. While suggestive of paternal inheritance, the U4c1a contribution could instead be ascribed to a multi-copy mtDNA insert (NUMT) in chromosome 14, by fluorescence in situ hybridization and sequencing of DNA from cultured fibroblasts that were depleted of their mtDNA (p<sup>0</sup> cells). Droplet digital PCR and multiplex qPCR suggested 45–56 copies of the NUMT insert per p<sup>0</sup> cell. Our study casts doubt on earlier claims of paternal inheritance that lack explicit testing for the presence of NUMT sequences.

## PgmNr 2312: Heteroplasmy dynamics in individuals with biparentally-inherited mitochondrial DNA.

### Authors:

J. Slone<sup>1</sup>; W. Zou<sup>1,2</sup>; S. Luo<sup>1,3</sup>; E.S. Schmitt<sup>4,5</sup>; S.M. Chen<sup>5</sup>; X. Wang<sup>1</sup>; J. Brown<sup>1</sup>; M. Bromwell<sup>1</sup>; Y. Chien<sup>6</sup>; W. Hwu<sup>6</sup>; P. Fan<sup>7</sup>; N. Lee<sup>6</sup>; L. Wong<sup>4</sup>; J. Zhang<sup>4</sup>; T. Huang<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA; 2) Reproductive Medicine Center, Department of Obstetrics and Gynecology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022, Anhui, China; 3) Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region, Nanning 530003, Guangxi, China; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 5) Baylor Genetics, Houston, TX, 77030; 6) Department of Pediatrics and Medical Genetics, National Taiwan University Hospital, Taipei 10041, Taiwan; 7) Department of Pediatrics, National Taiwan University Hospital, Taipei 10041, Taiwan

---

With very few exceptions, mitochondrial DNA (mtDNA) in humans is transmitted exclusively from mothers to their offspring, suggesting the presence of a strong evolutionary pressure favoring the exclusion of paternal mtDNA. We have recently shown strong evidence of paternal mtDNA transmission. In these rare situations, males exhibiting biparental mtDNA appear to be limited to transmitting just one of their mtDNA species to their offspring, while females possessing biparental mtDNA populations consistently transmit both populations to their offspring at very similar heteroplasmy levels. The precise biological factors underlying these unusual transmission events remain unclear. Here, we have examined heteroplasmy levels in various tissues among individuals with biparental inheritance. Our results indicate that individuals with biparental mtDNA have remarkable inter-tissue variability in heteroplasmy level. At the single-cell level, paternal mtDNA heteroplasmy in sperm varies dramatically, and many sperm possess only one of the two mtDNA populations originally in question. Taken together, these results show a fundamental, parent-of-origin difference in how mtDNA molecules transmit and propagate. This helps explain how a single population of mtDNAs are transmitted from a father possessing two populations of mtDNA molecules, suggesting that some mtDNA populations may be favored over others when transmitted from the father.

## PgmNr 2313: The spectrum of mitochondrial genomic variation across 250,000 individuals.

### Authors:

N.L. Washington<sup>1</sup>; A. Bolze<sup>1</sup>; F. Mendez<sup>1</sup>; E.T. Cirulli<sup>1</sup>; S. White<sup>1</sup>; M. Rashkin<sup>1</sup>; W.J. Metcalf<sup>2,3</sup>; J.J. Grzymalski<sup>2,3</sup>; J. Lu<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Helix, San Carlos, California.; 2) Desert Research Institute, Reno, Nevada; 3) Renown Institute of Health Innovation, Reno, Nevada

---

Mitochondrial variants are causative for rare mitochondrial diseases (1 in 5,000 clinically affected adults), and are also linked to common health conditions such as obesity and Alzheimer's disease. For those with deleterious variants, differences in heteroplasmic levels can lead to varying phenotypic presentations of the same disease in a population. Understanding the genetic contribution of mitochondrial variation to human health is challenging due to the unique nature of mitochondrial biology and its coexistent, complementary, yet distinct inheritance from that of the nuclear genome.

Smaller mitochondrial-specific databases are biased due to their composition; our current understanding of human mitochondrial variation is largely informed by studies that recruit for patients with inherited mitochondrial disease, and by studies of human evolution and migration patterns. In both cases, there is often bias in recruitment, and baseline rates of variation may be skewed. Information on the frequency of a homoplasmic or heteroplasmic variant in the population would help develop, test, and refine or disprove hypotheses for causality.

Here we characterize the mitochondrial genomes of more than 250,000 individuals sequenced in the Helix clinical laboratory. This collection is unbiased towards individuals with a mitochondrial disorder. We also provide the community with a research resource of all mtDNA variants identified and their frequencies. Our database includes more than 10,000 mtDNA variants, with ~20% present as singletons, and ~54% mtDNA bases invariant in the population. While our variant count is similar to MitoMAP, our singleton rate is lower and invariant base fraction higher as is expected in an unbiased cohort with improved mtDNA variant frequency estimates. We find that heteroplasmic-only variants are enriched at sites of high conservation, supporting the presence of heteroplasmy as an indicator of pathogenicity. We show that this large population-based approach improves our ability to interpret disease-associated variants by analyzing those reported to be pathogenic for Leber's Hereditary Optic Neuropathy (LHON). We find that many reportedly pathogenic variants have a frequency far above the maximum plausible allele frequency for a LHON-causing variant, including the primary variant m.14484T>C. We further confirm these findings with matched medical records for participants in the UK Biobank and Healthy Nevada Project.

## PgmNr 2314: Dissecting the genetic make-up of the Garifuna population.

### Authors:

S. Chavan <sup>1</sup>; M. Daya <sup>1</sup>; T. Brunetti <sup>1</sup>; J. Shortt <sup>1</sup>; N. Rafaels <sup>1</sup>; C. Arehart <sup>1</sup>; C. Cox <sup>1</sup>; M. Preethi Boorgula <sup>1</sup>; M. Campbell <sup>1</sup>; K. G. Beaty <sup>2</sup>; N. F. Baldi <sup>3</sup>; A. B. Conley <sup>4</sup>; G. M. Belbin <sup>5</sup>; I. King Jordan <sup>4</sup>; E. E. Kenny <sup>5</sup>; D. A. Mejía Mejía <sup>6</sup>; T. H. Beaty <sup>7</sup>; A. Mayorga <sup>8</sup>; R. A. Mathias <sup>7,9</sup>; C. Gignoux <sup>1</sup>; E. Herrera-Paz <sup>10</sup>; K. Barnes <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, University of Colorado, Denver, USA; 2) Department of Anthropology, University of Oklahoma, Norman, USA; 3) Centro de Investigaciones Antropológicas (CIAN), Universidad de Costa Rica, Costa Rica; 4) School of Biological Sciences, Georgia Institute of Technology, Atlanta, USA; 5) Center for Genomic Health, Icahn School of Medicine at Mount Sinai, USA; 6) Universidad Tecnológica Centro Americana UNITEC, Honduras; 7) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, USA; 8) Centro de Neumología y Alergias, San Pedro Sula, Honduras; 9) Department of Medicine, Johns Hopkins University, Baltimore, USA; 10) Facultad de Medicina, Universidad Católica de Honduras, San Pedro Sula, Honduras

---

The Garifuna is a Caribbean population that originated in the mid 17th century, with origins tales of shipwrecks off the coast of the island of St. Vincent. In addition to admixture with local Caribs, the Garifuna experienced a number of bottlenecks, periods of migration and exponential growth. In this study, we aimed to better understand the history of the Garifuna by leveraging genotype data available through Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA).

Samples were genotyped on Illumina's MEGA chip. After quality control, 520 samples were available for analysis from 8 communities (Alfonzo Lacayo n=43; Bajamar n=64; Corozal n=105; Sambo Creek n=122; San Juan n=28; Tornabe n=43; Travesia n= 95 and Triunfo de la Cruz n=20). We estimated frequencies of mitochondrial haplogroups (using Haplogrep2), Y- chromosome haplogroups (SNAPPY) and HLA genotypes (HIBAG), and compared frequencies among different communities. We phased genotype data and estimated segments of genetic material shared identical-by-descent (iLASH), to infer patterns of effective population size in current and past generations (IBDNe).

The most common mitochondrial haplogroups observed are L (83%; continental African), C1 and B2 (4.7% and 0.6%; continental American). The C1 American haplogroup was observed more frequently in Travesia (17.8%) and Bajamar (7.9%); was rare in Sambo Creek (2.5%); and absent in other communities. The most common Y haplogroups observed in the Garifuna are E1b1a1 (64.5%; continental African), Q1a2a1a1 (29.5%; continental American) and R1b1a2a1a (2.2%). Contrary to the mitochondrial data, the American origin Q1a2a1a1 haplogroup was common in all communities. The most common HLA alleles are A\*23:01 (21.3%), C\*04:01 (19.2%), DQA1\*01:02 (19.2%) and DPB1\*01:01 (32%) with corresponding African American frequencies of 10%, 20.5%, 30.7% and 27% respectively (obtained from allele frequencies.net). Consistent with previous analysis based on whole genome sequence data of n=41 subjects, our estimates of effective population size points to a

bottleneck event 10-15 generations ago, corresponding to the period when slaves of African ancestry first settled in this region.

For future work, we plan to genotype additional samples from coastal communities to expand our analysis. We will also generate ~100 whole genome sequences of indigenous Caribbean populations from Honduras, which will enable us to characterize admixture events in the Garifuna.

## **PgmNr 2315: *Deciphering the genetic mechanism of epilepsies using a founder population structure.***

### **Authors:**

S.L. Girard <sup>1</sup>; C. Moreau <sup>1</sup>; R.-M. Rebillard <sup>2</sup>; S. Wolking <sup>2</sup>; V. Cloutier <sup>2</sup>; J. Michaud <sup>3</sup>; B. Minassian <sup>4</sup>; P. Cossette <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Centre Intersectoriel en Santé Durable, Université du Québec à Chicoutimi, Chicoutimi, Quebec, Canada; 2) Axe Neurosciences, Centre de recherche du CHUM, Université de Montréal, Montreal, Quebec, Canada; 3) Centre de recherche du CHU Ste-Justine, Département de pédiatrie et de neurosciences, Université de Montréal, Montreal, Quebec, Canada; 4) Department of pediatrics and neurology and neurotherapeutics, UT Southwestern Medical Center, Irving, Texas, USA

---

### **Introduction**

Epilepsy is defined as a group of neurological disorders characterized by epileptic seizures defined as brief episodes of symptoms that are caused by abnormal or excessive neuronal activity in the brain. Although monogenic forms of the disease have been reported, they represent less than 2% of epilepsy cases. Until recently, most studies on epilepsies did not focus on a specific type of epilepsy, probably due to small sample sizes and lack of appropriated methods.

### **Methodology**

For the present study, we used 522 French-Canadians suffering from various types of epilepsies in addition to 954 French-Canadian control individuals. The French-Canadian population is well-known for its founder effect and its elevated frequency of some population-specific genetic disorders. All patients and controls underwent whole-genome genotyping and 212 epilepsy patients had their whole-genome sequenced. Additionally, all patients and controls had their genealogies reconstructed to identify a common pool of ancestors dating back to the 17<sup>th</sup> century (arrival of French colons in America). We imputed whole-genome sequences using genotypes to perform a GWAS and to calculate polygenic risk scores (PRS) and heritability on more specific types of epilepsy as described in other studies. We also used the genotypes to infer identity-by-descent (IBD) shared segments to try to identify shared haplotypes that could be related to the disease. Following this, we conducted a haplotype-based genome-wide association study in addition to enrichment analyses.

### **Results**

We were able to replicate several associations identified by the ILAE GWAS study. The same study was used to establish heritability profiles as well as significant PRS regression analysis (using the pairwise IBD matrix as covariates) for several types of epilepsy. These results show different profiles for cases and controls. We found several IBD segments with significantly higher frequency for specific types of epilepsy when compared to the control individuals.

### **Perspective**

Our comprehensive study on the genetics of epilepsy in the French-Canadian population shows the multifactorial nature of the disease as well as the need to use complimentary strategies in order to unravel the genetic mechanisms of the disease. It also shows the importance of refining the individual diagnosis of each patient and that having a good knowledge of the population structure will help to decipher the genetic mechanisms of epilepsies.



## PgmNr 2316: Fishing for function in the evolutionary gene pool: A zebrafish model of human duplicated gene *SRGAP2*.

### Authors:

M.Y. Dennis <sup>1,2</sup>; A. Colon-Rodriguez <sup>1</sup>; B. Radke <sup>1</sup>; K. Weyenberg <sup>1</sup>; G. Kaya <sup>1</sup>; E. Ferino <sup>1</sup>; A. Gelli <sup>3</sup>; P. Lein <sup>4</sup>; L. Jao <sup>5</sup>; J.M. Uribe-Salazar <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Center, MIND Institute, University of California, Davis, Davis, CA; 2) Integrative Genetics and Genomics Graduate Group, University of California, Davis, CA; 3) Department of Pharmacology, University of California, Davis, CA; 4) Department of Molecular Biosciences, School of Veterinary Medicine, University of California, Davis, CA; 5) Department of Cell Biology and Human Anatomy, University of California, Davis, CA

---

Gene duplication is a rich source of species innovation contributing to novel phenotypic features. Comparison of great ape genomes has identified millions of base-pairs uniquely duplicated affecting >30 gene families that may contribute to neurological features in humans. *SRGAP2*, one notable example, has duplicated uniquely in the *Homo* lineage to produce three partial paralogs present in modern humans. Previous studies show that expression of truncated human paralog *SRGAP2C* in mice impacts cortical neuronal migration and synaptogenesis by antagonizing the function of the full-length ancestral ortholog. Considering most human duplicated genes are uncharacterized, we propose using zebrafish (*Danio rerio*) to quickly test gene function. Zebrafish are advantageous over other vertebrate model systems based on their rapid development, small size, fecundity, transparent bodies, and high efficiencies of gene editing.

As proof of principle, we used CRISPR to generate a stable knockout of zebrafish ortholog *srgap2*. RNA-seq and *in situ* hybridizations show that *srgap2* expression peaks at ten hours post fertilization when postmitotic neurons begin forming in the developing zebrafish embryo. Consistent with previous findings that link *SRGAP2A* dysfunction with epilepsy (human) and synaptic connectivity (mouse), zebrafish *srgap2* mutants exhibit seizure-like behavior when exposed to GABA antagonist pentylenetetrazol (median distance moved in 18 hours, homozygous=42.3 m, wildtype=22.4 m, P=0.037). Experiments are ongoing to assay additional neurodevelopmental defects. Preliminary data show that *srgap2* mutants exhibit deviations in head size at two days post fertilization, suggesting its conserved function in neuronal migration. Based on the high conservation of *srgap2* between zebrafish and human (74% amino acid identity) and previous results in mouse models, we hypothesize that expression of human truncated paralog *SRGAP2C* in zebrafish will impede the function of endogenous *srgap2* and, thus, phenocopy defects identified in our knockout mutants. While a work in progress, promisingly, we observe significant levels of *SRGAP2C* transcripts retained in zebrafish two days post injection of *in vitro* transcribed RNA. Ultimately, if successful, these experiments will provide a roadmap to quickly assay the function of additional duplicated genes to understand if/how they contribute to the evolution of novel neurological traits and diseases unique to humans.



## PgmNr 2317: Diversity and directionality of a large human-specific repeat expansion.

### Authors:

M.M. Course <sup>1</sup>; K. Gudsnuk <sup>1</sup>; A. Sulovari <sup>2</sup>; S. Smukowski <sup>2</sup>; K. Winston <sup>1</sup>; N. Desai <sup>1</sup>; J. Ross <sup>3</sup>; P. Dion <sup>3</sup>; E.E. Eichler <sup>2</sup>; G.A. Rouleau <sup>3</sup>; P.N. Valdmánis <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, University of Washington School of Medicine, Seattle, WA, USA; 2) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 3) Montreal Neurological Institute, McGill University, Montreal, QC, Canada

---

So far, over 30 tandem repeat expansions in the human genome are known to cause neurological disease. However, the mechanism by which these repeats expand remains uncertain, largely because most disease-causing repeats are composed of an uninterrupted sequence tract, making it difficult to identify where the repeat arises within the expansion. We have identified a variable number tandem repeat (VNTR) in the last intron of the gene *WDR7*, which is predicted to fold into a stable hairpin structure and exhibits striking nucleotide variability in humans. This VNTR is 69 nucleotides long, with six variable nucleotide positions within each repeat unit. It expands from 1 copy in phased non-human primate genomes to 6 copies in the reference human genome, without adjacent repetitive elements that could facilitate this expansion. By performing multiplexed long-read sequencing in 144 individuals, we found that the *WDR7* repeat ranges anywhere from 1-86 repeat copies, with a mean of 16 copies. The internal sequence is even more complex, a feature we are leveraging to understand the origin and mechanism of its expansion. Alignment of the complete repeat sequence for each allele reveals that the 3' region is fixed, while variability increases towards the 5' end, opposite the direction of transcription. Evaluating the repeat in 1000 Genomes Project samples reveals that while super-populations share a similar distribution of length and internal repeat unit composition, certain repeat units are highly enriched in specific groups. Notably, these rare repeat units are biased to the 5' edge of the repeat. Analysis of repeat matches from short-read sequences provides evidence of two repeat copies in the Neanderthal genome, while the Denisovan genome has a mean of 27 repeat copies and a composition similar to modern-day humans. We also note several instances of large duplications within the repeat in several individuals, and that the majority of new sequence occurs in combinations of repeat units divisible by two. Two repeat units roughly match the length of an Okazaki fragment, which could account for this curious periodicity. Finally, we have confirmed that the unique hairpin structure of the repeat can generate microRNAs, suggesting a functional role for this repeat. Overall, our analysis of this complex intronic repeat expansion allows us to evaluate the roles of VNTRs in human evolution and disease, and uncover their mechanism of expansion.

## PgmNr 2318: Evolutionary patterns of conservation at schizophrenia GWAS loci reveal varying histories of natural selection.

### Authors:

M. Sankoh <sup>1,2</sup>; K.J. Wade <sup>2</sup>; S.-A. Bacanu <sup>3</sup>; B.T. Webb <sup>3</sup>; B.P. Riley <sup>3</sup>; B.C. Verrelli <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Integrative Life Sciences Doctoral Program, Virginia Commonwealth University, Richmond, VA; 2) Center for Life Sciences Education, Virginia Commonwealth University, Richmond, VA; 3) Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA

---

Schizophrenia (SCZ) is a common, complex, severely debilitating psychiatric disorder likely due to many genomic loci of small effect. Sequencing-based rare variant discovery efforts often focus on coding regions or assume a highly deleterious nature to the variants. While Mendelian disease mutations usually occur at coding nucleotide sites with high evolutionary conservation, this may not be true for alleles underlying common complex traits with persistent prevalence across time and populations. Disorders such as SCZ with reduced reproductive fitness in modern societies may instead fit a model where recent evolutionary constraint is detectable but constraint over deep phylogenetic time is not. Here, we used a GWAS meta-analysis of 65,205 SCZ cases and 87,919 controls that identified 248 independent loci as a dataset to determine the relationship between evolutionary site conservation and potential etiology. We applied the phylogenetic measure of evolutionary conservation PhyloP, which represents the likelihood of the observed conservation site score across phylogenetic time, for each of the different periods spanning vertebrate (n=100), mammalian (n=30), and primate (n=17) evolution. We also employed an analysis using a composite of tests of natural selection across the 248 loci within populations from the 1000 Human Genome Dataset. The PhyloP site distribution for the 248 GWAS variants showed no significant differences across the three time periods. However, there are significant outliers within each period with ~50% of the SCZ associated variants showing little conservation. To identify deviations across depths, we standardized the PhyloP scores within each locus at each depth and found significant heterogeneity. Several sites underlying GWAS signals show low conservation at deep vertebrate evolutionary time, but significantly higher evolutionary conservation within the primate lineage. We also identified loci with high vertebrate evolutionary conservation but with human population-specific positive natural selection. The observation that a majority of SCZ GWAS variants show little site conservation in deep time, and that this pattern changes over evolutionary time, is consistent with the prediction of a spectrum of functional variation expected for disorders with mild impact on reproductive fitness. These results demonstrate that GWA studies should consider evolutionary histories associated with complex traits to identify underlying disease variants.

## PgmNr 2319: Genome-wide detection and characterization of *de novo* repeat expansions in the general population.

### Authors:

I. Mitra <sup>1</sup>; N. Mousavi <sup>2</sup>; R. Yanicky <sup>3</sup>; M. Gymrek <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics Program, University of California San Diego, La Jolla, CA; 2) Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA; 3) Department of Medicine, University of California San Diego, La Jolla, CA; 4) Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA

---

Tandem repeats (TRs) consist of 1-20bp sequence motifs repeated consecutively. TRs have mutation rates that are orders of magnitude higher than SNPs, indels, or CNVs, and thus represent one of the largest sources of human genomic variability. Dozens of TRs have been associated with a range of phenotypes, including Huntington's Disease, Fragile X Syndrome, cancer risk, and schizophrenia. So far, population-scale analyses of TR mutations have been limited to several hundred loci due to technical challenges. Here, we use deep WGS of families and our recently developed TR genotyper, GangSTR, to conduct the first genome-wide (GW) population analysis of *de novo* TR mutations. GangSTR is capable of genotyping nearly 2 million TRs GW, ranging from short homopolymer repeats to large pathogenic repeat expansions.

We applied GangSTR to profile TRs from WGS of 1,916 quad families from the Simons Simplex Collection collected for autism research. We then developed a novel method, CookieMonSTR, to identify germline *de novo* TR mutations in WGS data of parent-offspring trios. CookieMonSTR combines genotype likelihood scores and estimated TR mutation rates to determine the posterior probability of a mutation at each locus. We identified an average of 300 *de novo* TR mutations per child and validated a subset of mutations using Sanger sequencing and capillary electrophoresis. Per-locus mutation rates were highly concordant with previous studies. Similar to other class of *de novo* variants, TR mutation rates were strongly correlated with the age of the father and had on average a 3x increase in mutational burden from fathers vs. mothers. Our results confirm a known mutational bias in direction dependent on the repeat length of the parental allele: alleles longer than the population mean are far more likely to contract, whereas shorter alleles are more likely to expand. We characterized the effects of replication timing, recombination, local SNP mutation rate, repeat unit base composition, recombination rate, local GC content, and modifier mutations on GW TR mutational processes.

Lastly, we develop a novel metric for prioritizing putative pathogenic TR mutations by combining per-locus constraint scores and population allele frequencies for each mutation. We apply our metric to identify a candidate set of TRs potentially involved in autism. Overall, GW identification and prioritization of TR mutations is an important tool for enabling the discovery of novel pathogenic loci.

## PgmNr 2320: Grave goods, mortuary practices, and ancient DNA at the Yukisma Mound in Central California.

### Authors:

K.G. Beaty<sup>1,2</sup>; E. Lenci Jr.<sup>3</sup>; A. Leventhal<sup>4</sup>; M.V. Arellano<sup>3</sup>; C. Nijmeh<sup>3</sup>; R. Cambra<sup>3</sup>; C. Monroe<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratories of Molecular Anthropology and Microbiome Research, University of Oklahoma, Norman, OK; 2) Department of Anthropology, University of Oklahoma, Norman, OK; 3) Muwekma Ohlone Tribe of the San Francisco Bay, CA; 4) School of Social Science, San Jose State University, San Jose, CA

---

The archaeological record indicates that the end of the Late Period (900 BP-250 BP) in Central California was marked with shifts in settlement and mortuary patterns which was influenced by environmental and social change. Mortuary treatment and distribution of grave goods suggest either a reduction in social inequality with wealth concentrated in a distinct few, or a shift toward a corporate group identity that was associated with kinship. These interpretations were tested using data collected from 200 individuals from Yukisma Mound (CA-SCL-38), including a combination of mitochondrial DNA (mtDNA) sequences and restriction length polymorphisms, demographic information including sex and age, the distribution and type of burial site, and the number and type of grave goods associated with each burial. Thirty-six haplotype lineages belonging to Native American haplogroups A (5.5%), B (22%), C (29.5%), and D (43%) were identified at Yukisma Mound. Seventy-eight percent of these burials were associated with grave goods, with the number and type of good varying widely between burials. Spatial autocorrelation showed that the distribution of these haplotypes throughout the site were not correlated ( $r \approx 0$ ) between maternal relatedness and burial type, or between maternal lineage and the quantity or diversity of grave goods, indicating that group identity was not associated with maternal kinship. The distribution and diversity of mtDNA lineages throughout the site may instead show its use as a cemetery for wealthy elites that lived throughout the Bay area.

## PgmNr 2321: Identification of introgression events between archaic hominins and modern human populations.

**Authors:**

Z. Jin; H. Watanabe

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genome Information Science Laboratory, Graduate School of Information Science and Technology, Hokkaido University , Sapporo, Japan

---

With the discovery of fossils and the accelerated accumulation of genome sequence data, increasing evidence supports the hybridization and introgression between the modern human lineage and archaic human lineages, e.g., Neanderthals and Denisovans, and between those archaic lineages. So far, a number of introgressed fragments have been suggested via different introgression identification methods.

In this study, we searched for genomic regions containing population-specific SNPs shared with archaic hominins and identified different introgression events, including unreported ones. The number of the genomic regions supporting a specific introgression event varies significantly among events, probably due to the timing of the event and natural selection. We will discuss the unreported events.

## PgmNr 2322: *DCDC2* READ1 regulatory element: How perception shapes language.

### Authors:

M.M.C. DeMille <sup>1</sup>; K. Tang <sup>2</sup>; J.C. Frijters <sup>3</sup>; J.R. Gruen <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Yale University School of Medicine, New Haven, CT 06520; 2) Department of Linguistics, Zhejiang University, Hangzhou, 310058 Zhejiang, China; 3) Child and Youth Studies, Brock University, St. Catherine's, ON L2S 3A1, Canada; 4) Department of Genetics, Yale University School of Medicine, New Haven, CT 06520

---

We previously showed an association between the READ1 regulatory element in *DCDC2* and the number of consonants in the languages of 43 populations distributed across five continents. We hypothesized that this association was due to the high precision required by the auditory cortex to discriminate between consonants and the role of *DCDC2* in modulating temporal precision of neuron firing. READ1 is associated with normal variation in phonological processing necessary for receptive language. This current study was designed to provide a link between the population frequency of the RU1-1 group of READ1 alleles with the neurophysiology of consonant perception. In particular, we focus on five consonant classes that differ in terms of their manner of articulation called stops, affricates, fricatives, nasals and approximants. We show that only stops and nasals associate with RU1-1. The nature of these consonant classes and the direction of their relationships with RU1-1 supports an account of how the sound system evolved to maintain phonemes that are easy to perceive.

## PgmNr 2323: Human adaptation to rice cultivation: Genetic evidence from contemporary China.

### Authors:

C. Zhu <sup>1</sup>; T. Talhelm <sup>2</sup>; G. Chen <sup>3</sup>; Y. Li <sup>3</sup>; G. Kong <sup>4</sup>; J. Wang <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) China Agricultural University, Beijing, Beijing, China; 2) University of Chicago, Booth School of Business, Chicago, Illinois, USA; 3) WeGene, Shenzhen, Guangdong, China; 4) Renmin University of China, Beijing, China

---

It is widely accepted that agriculture, one of the most important transitions in human history, has strongly changed the way people lived and eventually made us who we are today. Following the initial domestication in the lower Yangtze River valley 9,400 years ago, rice farming became widespread throughout China, changing patterns of food consumption, labor supply, and reproduction among Neolithic populations. Here we show that, the advent of rice domestication and cultivation has shaped the humans not only culturally but also genetically.

Using a dataset of 4,538 individuals recently collected from mainland China, we test whether genetic variants associated with various phenotypes have been under selection of the rice cultivation history. We leverage recent findings from molecular genetics to construct a number of polygenic scores that use individual genetic information to predict their height, body mass index (BMI), age at first birth, time discounting, risk tolerance, educational attainment, depression, and alcohol tolerance. We then examine associations between these polygenic scores and proportion of cultivated paddy rice field of the individual's birthplace at the county level.

Our results imply that rice cultivation may have exerted selective pressures in favor of earlier reproduction and higher levels of alcohol intolerance. Under such selection, the beneficial genes became more common among populations in rice farming areas as their carriers have more descendants. These results are robust after we control for individual ethnic makeup, regional differences (i.e. average temperature, historical population density, herding history, regional economic development, adjusted river length, historical pathogen epidemics), and potential endogeneity of rice farming. Our results provide evidence that human evolution occurred over the past generations due to culture and lifestyle inherited from a rice-cultivating society, shedding light on the potential role of rice domestication/cultivation in the history of human evolution for the first time.

## PgmNr 2324: The first intron of genes is functionally relevant: Analysis of epigenetic chromatin marks in the first intron.

**Authors:**

S.S. Choi; S.S. Jo; H. Choi

[View Session](#) [Add to Schedule](#)

**Affiliation:** Kangwon National University, Chuncheon, Gangwon-do, Korea, Republic of

---

The first intron of genes exhibits several interesting characteristics not seen in other introns: 1) it is the longest intron on average in almost all eukaryotes, 2) it presents the highest number of conserved sites, and 3) it exhibits the highest density of regulatory chromatin marks. Here, we expand on the aspects by integrating various multiomics data, leading to further evidence supporting the functionality of sites in the first intron. First, we show that trait-associated single-nucleotide polymorphisms (TASs) are significantly enriched in the first intron. We also show that, in the first intron, epigenetic chromatin signals are densely located near TASs. Furthermore, we reveal that the distribution of epigenetic chromatin signals is not independent of the width or level of gene expression. Moreover, genes carrying multiple first intron TASs are shown to interact with each other within a large protein-protein interaction network connected to the *UBC* protein. We believe that our results will shed light on the functionality of first introns as the genomic entity performing gene expression regulation.



## PgmNr 2325: Whole genome sequence data reveal global demographic dynamics respond to the last ice age.

**Authors:**

Liu

[View Session](#) [Add to Schedule](#)**Affiliation:** USF Genomics, College of Public Health, University of South Florida, Tampa, Florida.

---

It has long been hypothesized that human populations experienced a shrinkage during the last ice age and then recovered and boomed following the deglaciation. However, previous population genetics studies only provide scattered evidence and most of those are from certain European populations. In this study, we analyzed 27 whole genome sequence data sets from 23 populations worldwide (5 African, 5 South Asian, 6 European and 7 East or Southeast Asian) using the Stairway Plot method and inferred their demographic histories. Our results show that many European, East Asian and Southeast Asian populations experienced population size decrease during the last glacial maximum between 30,000 to 10,000 years ago, while most of the African and South Asian population growth stagnated between 30,000 to 20,000 years ago. Beginning from 20,000 years ago, we observed population growth of some South Asian and African populations possibly responding to the deglaciation. Another wave of population growth of some European and East Asian populations began around 10,000 years ago, which is likely influenced by both the deglaciation and agriculture development. Finally, most of the populations experienced dramatic population growth between 6,000 to 3,000 years ago, which are likely the results of agriculture development.

## PgmNr 2326: Dimensionality reduction in population genetics: New methods for understanding population structure.

### Authors:

A. Ioannidis <sup>1,2</sup>; J. Blanco Portillo <sup>3</sup>; J.E. Rodriguez Rodriguez <sup>3</sup>; A. Pourshafeie <sup>2</sup>; C. Quinto Cortes <sup>3</sup>; A. Moreno Estrada <sup>3</sup>; C. Bustamante <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Computational and Mathematical Engineering, Stanford, CA, USA; 2) Department of Biomedical Data Science, Stanford University, Stanford, CA, USA; 3) National Laboratory of Genomics for Biodiversity, Irapuato, Guanajuato, Mexico

---

Dimensionality reduction methods, such as PCA and MDS, have long been relied upon as both filtering and visualization tools to interpret the genetic structure present in extremely large dimensional genomic datasets. Recently, new dimensionality reduction methods, including tSNE and UMAP, have been released, but their application to genetic structure interpretation is not common. Meanwhile, algorithms for ancestry-specific PCA and MDS have also been developed, but these are not widely used or understood, and currently suffer from a few serious flaws, which we will discuss.

We will present a novel multi-array MDS algorithm, dimensionality reduction methods particularly applicable to large datasets of ancient DNA, where each sample is missing many different sites, dimensionality reduction methods for accounting for disparate noise across samples or for uneven sampling, and novel, highly accurate, methods for ancestry-specific application of a variety of dimensionality reduction techniques. We will present procedures for determining which of these novel, or existing, algorithms are appropriate for a given application, discuss widespread misunderstandings and pitfalls (including inadvertent supervised learning and unbalanced sampling), and discuss proper methods, seldom applied, for determining dimensionality of a genomic dataset. We will present guidelines and best practices, as well as theoretical interpretations for each of these different modalities of dimensionality reduction in terms of traditional statistics for population substructure and genetic drift.

## PgmNr 2327: Taiwan View: A Taiwanese genomic variation, HLA genotyping and methylation status database.

**Authors:**

M.W. Su; C.W. Lin; H.W. Chu; T.C. Lee

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of Biomedical Science, Academia Sinica, Taipei, Taiwan Biobank, Taiwan

---

Taiwan Biobank intends to conduct a large-scale prospective cohort studies and case-control studies on local diseases. The cohort study will call for 200,000 volunteers, while the case-control study will invite 100,000 patients with the 10 to 15 most common diseases. There are currently over 114,533 participants are recruited in the cohort study and 27,751 participants have TWB1 genotyping array data. Taiwan View (<https://taiwanview.twbiobank.org.tw>) consists of the relational database and the web interface. The database includes single nucleotide variations (SNVs) frequency, mean methylation  $\beta$  value, and HLA genotyping frequency. The web interface consists of functions to search SNVs and explore the chromosome region surrounding an SNV. In the third release, this database contained frequency data of SNVs among 1,517 whole genome sequence samples and 24,625 Affymetrix Axiom TWB1 genotyping array samples from the Taiwan Biobank cohort study. There also 2,091 blood samples have DNA methylation status at > 850,000 CpGs sites. The DNA methylation status was examined using the Illumina Infinium MethylationEPIC BeadChip. At each targeted site, the ratio of the fluorescence signal was measured using a methylated probe relative to the sum of the methylated and unmethylated probes, the so-called  $\beta$  value, which ranges from 0.00 to 1.00, reflecting the methylation level of an individual targeted site. HLA genotyping data of 1,103 Taiwan biobank cohort study samples were generated with NXType™ Ion S5 NGS for Class I (HLA-A, -B, and -C) and Class II (HLA-DR, HLA-DQ, and HLA-DP).

## PgmNr 2328: Changes in ancestry-related assortative mating in the United States across multiple generations.

### Authors:

R. Sebro; S. Murray

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics & Radiology, Univ Pennsylvania, Philadelphia, Pennsylvania.

---

### Abstract

**Introduction:** The United States of America (USA) is a cosmopolitan, multiracial, multiethnic country with over 800 different ancestries. Ancestry-related assortative mating is required for population substructure to persist. The aim of this abstract is to evaluate the rate of the change of endogamy across several generations in the United States and to evaluate how this affects genetic studies.

**Materials and Methods:** The data were extracted from the 2000 and 2010 decennial USA Censuses and the 2011 to 2016 American Community Survey (ACS) Census. Data collected included individual age, sex, race, ancestry, and marital status. Spouse-pairs were considered to be from the Silent Generation (1928-1945); Baby Boomer Generation (1946-1964); Generation X (1965-1981); or Millennials (1982-2000). Estimates of endogamy in the 50 most common ancestries were calculated using phi-coefficient. McNemar's odds ratio was used to evaluate asymmetry in assortative mating. Statistics were all two-sided and P-values<0.05 were considered statistically significant.

**Results:** Over 3 million spouse-pairs were available for analysis (1.5 million spouse-pairs from Census 2000, and over 1.5 million spouse-pairs from Census 2010). The data show strong ancestry-related assortative mating based on self-reported ancestry, which decreased across each generation. The ancestry-related assortative mating was weakest among European-derived populations ranging from 0.24 (P<0.05) (between Danish and Germans) to 0.989 (P<0.05) (between Danish and Eastern Europeans). Ancestry-related assortative mating within Hispanics ranged from 0.816 (P<0.05) (between Mexicans and Central Americans) and 0.989 (P<0.05) (between Mexicans and Dominicans); within Asians ranged from 0.925 (P<0.05) to 0.995 (P<0.05) (between Japanese and Vietnamese); within Native Hawaiian and/or Pacific Islanders (NHPI) ranged from 0.888 (P<0.05) (between Samoans and Hawaiians) and 1.00 (between Samoans and Guamanians). Asymmetry was largest within Europeans between the Italian and Spanish (OR=1.35); largest within Hispanics between the Cubans and Dominicans (OR=2.19); largest within Asians between Asian-Indians and Koreans (OR=7.75); and largest within NHPI between Samoans and Hawaiians (OR=4.00).

**Conclusion:** Ancestry-related assortative mating decreased across each generation consistent with decreased endogamy. Significant asymmetric mating was noted, which will have implications for diseases affected by genetic imprinting.

## PgmNr 2329: Gene loss and functional decline of chitinases and chitinase-like proteins in dog.

### Authors:

F. Oyama <sup>1</sup>; A. Kashimura <sup>1</sup>; E. Tabata <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Chemistry and Life Science, Kogakuin Univ, Hachioji, Tokyo, Japan; 2) Research Fellow of Japan Society for the Promotion of Science (DC1), Tokyo, Japan

---

Chitin is a linear  $\beta$ -1, 4-linked polymer of *N*-acetyl-D-glucosamine (GlcNAc) and it is a major component of fungal cell walls and exoskeletons of invertebrates. Although humans and mice do not synthesize endogenous chitin, they express two active chitinases, chitotriosidase (Chit1) and acidic chitinase (Chia). In addition, they also produce chitinase-like proteins, chitinase 3-like-1 (Chi3L1) and chitinase 3-like-2 (Chi3L2), which are structurally homologous to chitinases but lack the ability to degrade chitin. Chia has attracted considerable attention due to its increased expression under specific pathological conditions including asthma and allergic inflammation. We have reported that chitinolytic activity of human Chia is significantly lower than that of the mouse counterpart and that human Chia became an inactive pseudogene during the course of evolution. Recently, we also showed that Chia is highly expressed in mouse, chicken, and pig stomach tissues and it can digest chitin in the respective gastrointestinal tracts. In this study, we focus on dog, a major domestic and carivorous animal. Chia mRNA level was significantly lower in the dog stomach than those in mouse. In addition, the chitinolytic activity of dog Chia enzyme was lower compared with that of the mouse Chia enzymes. Furthermore, dogs do not possess both Chit1 and Chi3L2 genes in the genome, although it possess Chi3L1 gene. These results suggest that dog lost their Chit1 and Chi3L2 genes and that Chia is likely to be a gene undergoing loss of function during the course of evolution.

## PgmNr 2330: Functional recovery of dog acidic chitinase.

**Authors:**

E. Tabata <sup>1,2</sup>; A. Kashimura <sup>1</sup>; F. Oyama <sup>1</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Chemistry and Life Science Dept, Kogakuin Univ, Hachioji, Tokyo, Japan; 2) Research Fellow of Japan Society for the Promotion of Science (DC1), Tokyo, Japan

---

Chitin, a polymer of *N*-acetyl-D-glucosamine (GlcNAc), functions as a major structural component in crustaceans, insects and fungi and is the second most abundant polysaccharide in the nature. Chitin-containing organisms such as insects and fungi have been suggested as novel animal feed resources. The drawback of using these organisms as livestock diets has been considered the high content of chitin which has long been thought to be indigestible fibers. We showed that Chia functions as a protease-resistant digestive enzyme in mouse, chicken and pig (omnivores), whereas the chitinolytic activity were lower in the bovine (herbivores) and dog (carnivores). Here, we show that dog Chia is activated by introducing specific regions of mouse Chia. To identify regions responsible for the reduced chitinolytic activity in dog Chia, we expressed chimeric dog-mouse Chia proteins in *E. coli* and measured the chitinolytic activity. Chimera with the region encoded by dog exon 6 and 7 regions exhibited weak chitinolytic activity toward the fluorogenic substrate and similar to that of the dog enzyme. In contrast, chimera with region encoded by mouse exon 6 and 7 showed strong chitinolytic activity. Thus, the region encoded by exon 6-7 in dog Chia significantly reduced chitinolytic activity. These results suggest that dog Chia integrated some amino acids in the region leading to the reduced chitinolytic activity during evolution.

## PgmNr 2331: A functional genomics investigation reveals the molecular basis for adaptation to diving in Sea Nomads.

### Authors:

M. Ilardo <sup>1,2</sup>; R. Nielsen <sup>1</sup>; B. Leibold <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Integrative Biology, UC Berkeley, Berkeley, CA 94720, USA; 2) The University of Utah Molecular Medicine Program, University of Utah, Salt Lake City, UT 84132, USA

---

We present molecular mechanisms underlying the recently discovered genetic adaptation to diving in a group of Sea Nomads. The Bajau people were shown to have evolved enlarged spleens to enhance their diving capabilities. In all mammals, including humans, the spleen contracts during diving to infuse a bolus of oxygenated red blood cells into systemic circulation. Therefore, a larger spleen has been proposed to prolong dive time in this population. The large spleen phenotype observed in the Bajau is associated with a genetic variant in the gene *PDE10A*, hypothesized to affect spleen size by stimulating thyroid hormone production. However, the precise molecular mechanisms underlying the relationship between thyroid hormones and spleen size are yet unknown. We used a functional genomics approach to demonstrate a significant increase in spleen size ( $p = 0.049$ ) in mice that pharmacologically simulate the Sea Nomad *PDE10A* genetic variant. Using flow cytometry, immunocytology, and hematological measurements, we have identified hematopoietic variation potentially underlying changes in spleen size. Further, we link these hematological phenotypes to Genome Wide Association Study (GWAS) data from human populations. Our results indicate that the large spleen phenotype observed in the Bajau is inducible in mice via manipulation of *PDE10A* and that the underlying mechanism involves red blood cell production and turnover, which could have important implications for hematological disorders.

## PgmNr 2332: Uncovering the evolutionary history of the Turkana, a desert pastoralist tribe.

### Authors:

A. Lea <sup>1,2</sup>; D. Martins <sup>1,3</sup>; J. Kamau <sup>4</sup>; J. Ayroles <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Ecology and Evolution, Princeton University, Princeton 08544, New Jersey, USA; 2) Lewis Sigler Institute for Integrative Genomics, Princeton University, Princeton 08544, New Jersey, USA; 3) Mpala Research Centre, Nanyuki 10400, Kenya; 4) Institute of Primate Research, National Museums of Kenya, Nairobi 00502, Kenya

---

In a blink of evolutionary time, humans have explored every corner of this planet and have shown an amazing capacity to adapt to extreme conditions. However, few well-understood examples of selection pressures, adaptive genetic variation, and phenotypic outcomes exist for a single population, limiting our knowledge of how genetic variation arises and gives rise to complex traits in our species. This gap is largely due to difficulties in sampling populations of interest, generating large genome-wide datasets, and connecting loci identified by selection scans with their phenotypic effects.

We are addressing this problem by studying the Turkana of northern Kenya, a semi-nomadic pastoralist tribe that inhabits one of the most arid regions in East Africa. In addition to living in an extreme desert climate, the Turkana subsist on a very high protein diet (>70-80% of the diet is animal-derived). In fact, daily protein intake exceeds the FAO/WHO requirements by >300%, despite general caloric intake being limited (1,300–1,600 kcal/day for adults).

To uncover loci associated with desert living and a pastoralist lifestyle, we collected DNA samples and extensive biomarker data from several hundred Turkana, as well as individuals from 10 nearby tribes. We compiled a large dataset comprised of 31 high (>20x) and 270 low (~6-7x) coverage genomes, as well as genome-wide array data for a subset of individuals for validation (n=188). Using this dataset, we imputed missing genotypes in the low coverage samples from the high coverage reference set with ~98% accuracy. Admixture analysis identified many regions of shared ancestry between the Turkana and other nearby groups, confirming the complex demographic history of this tribe. Finally, we identified dozens of genes that exhibit signatures of positive selection, and may contribute to the Turkana's ability to tolerate extreme desert conditions and a high protein diet. In particular, we identified an enrichment of genes involved in fatty acid metabolism, amino acid metabolism, and vasoconstriction/vasodilation, which we are currently linking to our phenotypic data. Together, this work provides a comprehensive picture of how natural selection shapes genetic and phenotypic variation in a unique population, providing insight into evolutionary processes and present day human diversity.



## PgmNr 2333: HLA allele frequency distribution and evolutionary relationships of ethnically diverse African populations.

### Authors:

E. Mbunwe <sup>1</sup>; J. Duke <sup>2</sup>; D. Ferriola <sup>2</sup>; T. Mosbrugger <sup>2</sup>; G. Damianos <sup>2</sup>; A. Dinou <sup>2</sup>; I. Kotsopoulou <sup>2</sup>; A. Ranciaro <sup>1</sup>; S. Thompson <sup>1</sup>; W. Beggs <sup>1</sup>; S.W. Mpoloka <sup>3</sup>; G.G. Mokone <sup>4</sup>; T. Nyambo <sup>5</sup>; D.W. Meskel <sup>6</sup>; G. Belay <sup>6</sup>; C. Fokunang <sup>7</sup>; A.K. Njamnshi <sup>8</sup>; T.L. Lenz <sup>9</sup>; M. Carrington <sup>10</sup>; M. Maier <sup>11</sup>; S. Tishkoff <sup>1,12</sup>; D. Monos <sup>2,13</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, Univ Pennsylvania, Philadelphia, Pennsylvania.; 2) Immunogenetics Laboratory, Division of Molecular Diagnostics, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA.; 3) University of Botswana, Biological Sciences, Gaborone, Botswana; 4) University of Botswana, Faculty of Medicine, Gaborone, Botswana; 5) Muhimbili University of Health and Allied Sciences, Biochemistry, Dar es Salaam, Tanzania.; 6) Addis Ababa University, Department of Biology or Department of Microbial Cellular and Molecular Biology, Addis Ababa, Ethiopia.; 7) Department of Pharmacotoxicology and Pharmacokinetics, Faculty of Medicine and Biomedical Sciences, The University of Yaoundé I, Yaoundé, Cameroon; 8) Department of Neurology, Central Hospital Yaoundé; Brain Research Africa Initiative (BRAIN), Neuroscience Lab, Faculty of Medicine and Biomedical Sciences, The University of Yaoundé I, Yaoundé, Cameroon.; 9) Research Group for Evolutionary Immunogenomics, Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, Ploen, Germany; 10) Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, MD, United States and Ragon Institute of the Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Boston, MA, United States.; 11) Bioinformatics Research, National Marrow Donor Program/Be The Match, Minneapolis, MN, USA and Biomedical Informatics, Center for Blood and Marrow Transplant Research, Minneapolis, MN, USA; 12) Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA 19104, USA.; 13) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA,

---

Genes of the immunologically important human leukocyte antigen (HLA) are key candidates for both adaptive divergence and introgression among human populations, depending on lifestyle and environment. This project investigated whether population evolutionary relationships inferred from highly polymorphic HLA genes in chromosome 6p21.1-21.3 differed from those inferred from genome-wide neutral loci. Volunteers (N=231) were sampled from 12 ethnically diverse sub-Saharan populations (≥13 participants per population) from Eastern (Ethiopia, Tanzania), Central (Cameroon) and Southern (Botswana) Africa. These populations practice traditional subsistence lifestyles including agro-pastoralism (4 populations), pastoralism (3 populations) and hunting-gathering (5 populations). All individuals were fully characterized at 11 HLA loci (*HLA-A*, *-B*, *-C*, *-DPA1*, *-DPB1*, *-DQA1*, *-DQB1*, *-DRB1*, *-DRB3/4/5*) by targeted next generation sequencing (Holotype HLA, OMIXON). Overall, 291 alleles, including 260 known and 31 previously unknown alleles (representing 6 synonymous and 25 non-synonymous point mutations), were identified. Clustering analysis using population allele frequencies per population was performed separately for 8 classical HLA loci. The first axis of Principal

Component analysis (PCA1) was able to explain between 16% and 25% of inter-population variation per locus. Heat maps based on allele frequencies confirmed the pattern of clustering at each locus. Neighbor joining phylogenetic trees were constructed using genetic distance matrices. In accordance with PCA, there was great variation in population clustering between loci; for some HLA loci, phylogenies reflected population distances based on neutral data whereas they did not for others. These observations are consistent with contrasting evolutionary trajectories of HLA loci (non-neutral) versus neutral loci used in genome-wide analysis. In addition, they suggest that some HLA alleles may be more driven by selection than others, and that in some cases the similarities between divergent populations could reflect similar selection pressure. Computational modeling and examination of shared selective forces (e.g. infectious disease exposure) could be useful for distinguishing the role of demography and natural selection influencing patterns of variation at the HLA region in ethnically diverse African populations.

## PgmNr 2334: Evolutionary dynamics of sex-biased gene expression in mammalian tissues.

### Authors:

S. Naqvi <sup>1,2</sup>; A.K. Godfrey <sup>1,2</sup>; J.F. Hughes <sup>1</sup>; M.L. Goodheart <sup>1,3</sup>; R.N. Mitchell <sup>4</sup>; D.C. Page <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Whitehead Institute, Cambridge MA 02142; 2) Department of Biology, Massachusetts Institute of Technology, Cambridge MA 02139; 3) Howard Hughes Medical Institute, Whitehead Institute, Cambridge MA 02142; 4) Brigham and Women's Hospital, Harvard Medical School, Boston MA 02115

---

**Introduction:** Sex differences are widespread in both humans and other mammals. As one example, the distribution of height or body size is shifted upwards in males relative to females. However, little is known about how gene expression differs between the sexes in a broad range of mammalian tissues and species. A catalogue of such sex-biased gene expression could serve to better understand phenotypic sex differences. Assessing the extent to which sex-biased gene expression is conserved across the body could also have important implications for the use of non-human models of sex differences.

**Rationale:** To identify both conserved and lineage- or species-specific sex differences in gene expression, we sequenced RNA from male and female samples in twelve tissues in each of four non-human mammals (macaque, mouse, rat, dog), and analyzed these data jointly with publicly available data from post-mortem male and female human tissues. To assess the impact of sex-biased gene expression on the sex difference in mean human height, we applied methods that integrate the effects of genetic variation on both gene expression and phenotype (height in this case). We sought to understand which transcription factors (TFs) contribute to evolutionary changes in sex bias by analyzing motifs gained or lost concurrently with lineage- or species-specific changes in sex bias.

**Results:** We identified ~3,000 genes with conserved (species-shared) sex bias in gene expression, most of which was tissue-specific. The cumulative effects of conserved sex bias explain ~12% of the sex difference in mean human height. However, most sex-biased gene expression (~77%) was specific to single species or subsets of species, implying that it arose more recently during evolution. We estimate that gains and losses of motifs for sex-biased TFs can account for ~27% of lineage-specific changes in sex bias.

**Conclusion:** By conducting a 12-tissue, five-species survey of sex differences in gene expression, we found that while conserved sex bias in gene expression exists across the body, most sex bias has been more recently acquired during mammalian evolution. Height is likely subject to opposing selective pressures in males and females; our study thus documents how such selective forces can result in sex-biased expression which can lead to trait distributions shifted between the sexes. Our findings also suggest that, in many cases, non-human mammals may not recreate molecular sex differences observed in humans.

## PgmNr 2335: Paternal diversity in different ethnic groups of Jammu and Kashmir.

### Authors:

E. Rai <sup>1</sup>; I. Sharma <sup>1</sup>; V. Sharma <sup>1</sup>; G. Singh <sup>2</sup>; P. Kumar <sup>3</sup>; M. Vilar <sup>4</sup>; S. Sharma <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Biotechnology, Shri Mata Vaishno Devi University, Katra, 182320, India; 2) Department of human Genetics, Guru Nank Dev University, Amritsar; 3) Institute of Human Genetics, University of Jammu, Jammu and Kashmir, 180006, India.; 4) Human Genographic Project, National Geographic Society, USA.

**Background:** Jammu and Kashmir is located on the crossroads of Eurasia, it has been believed that this region might have acted as corridor to various migrations and immigrations to and from mainland India, Eurasia or northeast Asia. These migrations and immigrations have impacted the cultural and social structure of society. Ethnically and linguistically this region is diverse.

**Objective:** To find out the paternal diversity among different ethnic groups of Jammu and Kashmir.

**Methodology:** 133 markers of non-recombining (NRY) region of Y chromosome belonging to major haplogroups from A to R, were genotyped for 291 samples. Genotyping was done by using Agena Massarray Platform.

**Results:** A total of twenty one different haplogroups: C, G2, H1, H1a, H1b, J, J2, J2b, K, L, P, Q, R, R1a1a, R1a1a1, R2, E1, F, G and H1a3 were found after genotyping. The frequent haplogroup found was R1a1a1, H1b was the second frequent and R2 was the third frequent haplogroup found in the present study. Huge paternal diversity is indicated in the population studied by the presence of diverse paternal haplogroups.

**Conclusion:** Multiple lines of paternal ancestry and geographic origins is highlighted by the presence of more than one haplogroup in each of the ethnic groups in the present study. Even these ethnic groups are presently endogamous, yet their genetic architecture suggests their genetic relations. This unifies populations of the region genetically as the presence of haplogroups was not observed as caste specific.

## PgmNr 2336: Genetic and environmental contributions to variations in intellectual development: A twin study.

**Authors:**

A. Panghal; B. Dhanda

[View Session](#) [Add to Schedule](#)

**Affiliation:** HDFS, CCS HAU, Hisar-125004, Hisar, India

---

We analyzed how the effects of genetic and environmental factors on intellectual development of twins from 6-8 years. The data were collected by assessment, interview, observation and questionnaire method from the Hisar twins (N=110) and Bhiwani twins (N=90) at 6-8 years of age and from their parents for environmental circumstances by using the WISC-R (Wechsler, 1974) and HOME Inventory by Bradley *et al.* (1988). The data analyzed by using heritability estimates, mean, standard deviation, z-test, correlation coefficient and chi square test. The correlation coefficient of monozygotic twins was  $r=0.93$ ,  $r=0.95$  at the age group 6-7 years and 7-8 years respectively and the correlation coefficient was dizygotic twins was  $r=0.64$  (6-7 years) and  $r=0.69$  (7-8 years). The heritability estimate explained the variation in intellectual development of twins was attributed to genetic factors ( $h^2=58\%$ ) at 6-7 years and ( $h^2=51\%$ ) at 7-8 years. Simultaneously, the remaining variations in intellectual development of twin were due to environmental factors (42%, 49% at 6-7 years and 7-8 years respectively). These findings suggest that the correlation coefficient of monozygotic twins was more than the dizygotic twins i.e. monozygotic twins share 100% genetic material whereas the dizygotic twins share 50% genetic material. On the conclusion of the heritability estimates showed that reverse trend with regard to the genetic influence on intellectual development of twins as decreasing of scores on heritability estimates with increasing age of the twins. It clearly reveals that the influence of genetic material decrease as the age of twins increase and the contribution of environmental circumstances increase the age of twins increase.

## PgmNr 2337: A catalogue of human ancestry components from population studies.

### Authors:

C. Barbieri <sup>1,2</sup>; V. Bajic <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland; 2) Department of Linguistic and Cultural Evolution Max Planck Institute for the Science of Human History, Jena, Germany; 3) Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

---

The latest studies of ancient and modern DNA are revealing a high-resolution picture of human history and migration, by improving the sample geographic coverage and the methodological power. New open questions are concerning the appropriate framing of the results, and the impact of those for the society at large.

With genomic datasets routinely analyzed, the field of population genetics is shifting from using discrete units of variations (such single genetic markers, or uniparental haplogroups) to other complex forms of ancestry, often identified with ethnolinguistic references, cultural attributes, or geographic descriptors. This level of information is sometimes difficult to translate to non-experts. Cultural studies, forensic applications, and commercially available ancestry tests brought examples of how combining genetics with human history can generate a problematic narrative bordering genetic determinism and identity marking. How much do such ancestry components define the history of actual people, or distinct populations?

Here we examine the state-of-the-art in the studies of admixture components retrieved by ancient and modern DNA. We perform Admixture analysis on a global dataset of more than 3000 individuals and evaluate clustering and frequencies of components in different macro regions. This analysis returns a catalogue of discrete ancestries which we assemble in a hierarchical fashion, and confront against the structure of known haplogroup phylogenies. We then discuss the frequency and distribution of these components as shaped by human migration and admixture. By reconsidering the history of our genetic structure, this work ultimately aims at deconstructing the strict association between ancestry components and human populations.

## PgmNr 2338: Exploring the demographic history of tribal and caste groups from West Maharashtra (India) using dense genome-wide data.

### Authors:

M. Jonnalagadda <sup>1</sup>; C. Abbatangelo <sup>2</sup>; G. Debortoli <sup>2</sup>; C. Fortes-Lima <sup>3</sup>; H. Norton <sup>4</sup>; S. Ozarkar <sup>5</sup>; E. Parra <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Symbiosis School for Liberal Arts, Symbiosis International (Deemed University), Pune, India; 2) Department of Anthropology, University of Toronto at Mississauga, Canada; 3) Human Evolution Program, Department of Organismal Biology, EBC, Uppsala University, Sweden; 4) Department of Anthropology, University of Cincinnati, Ohio, USA; 5) Department of Anthropology, Savitribai Phule Pune University, Pune, Maharashtra, India

---

The South Asian subcontinent has a complex history of human migrations and interactions, which is reflected in an extensive cultural and genetic diversity. One of the defining characteristics of the Indian human landscape is the presence of endogamous tribal and non-tribal groups. Additionally, non-tribal groups are hierarchically arranged in social classes termed castes. Recent studies based on genomic data for ancient and modern South Asian populations have indicated that the genetic diversity in this region is derived primarily from three population sources: *i)* Ancient Ancestral South Indians (AASI), which represent an early South Asian hunter-gatherer lineage, *ii)* Iranian agriculturalists, which started to expand from West Asia to South Asia via the Iranian Plateau in the 7<sup>th</sup> and 8<sup>th</sup> millennia BCE, and *iii)* Steppe pastoralists, which arrived in the region in the 2<sup>nd</sup> millennium BCE. In this study, we characterized the population diversity and demographic history of six populations from West Maharashtra (WM), India using dense genome-wide SNP data. The dataset includes 331 individuals from four Indo-European tribal groups (Kokana, Warli, Bhil and Pawara) and 125 individuals from two castes (Deshastha Brahmins and Kunbi Marathas). Comparisons of the six WM samples with other available ancient and modern DNA samples indicate that the WM tribal groups are characterized by some of the highest proportions of AASI ancestry reported in South Asia. The relative contributions of Steppe pastoralists and in particular Iranian agriculturalists have been very small in the tribal groups. In contrast, in the two caste samples both the Steppe and agriculturalist contributions are substantially higher, in line with what has been described in other Indian caste groups. The Deshastha Brahmin sample has a higher Steppe contribution than the Kunbi Maratha sample, also in agreement with previous reports indicating relatively high Steppe proportions in Brahmins. Overall, this study highlights the complex demographic histories of tribal and caste groups from India.

## PgmNr 2339: A population-based approach for genes prioritization in the diagnosis of Mendelian and complex diseases.

### Authors:

M. Mezzavilla <sup>1</sup>; M. Cocca <sup>1</sup>; F. Guidolin <sup>3</sup>; P. Gasparini <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) IRCSS 'Burlo Garofolo', Trieste, TS, Italy; 2) University of Trieste, Trieste, TS, Italy; 3) University of Padova, Padova, PD, Italy

---

Gene prioritization is the process of determining which variants or genes identified in any genetic analysis are likely to cause a disease or phenotype. The advent of high throughput technologies followed by increased data production and GWAS studies results lead to the necessity of rapid prioritization of different genes. We should note that for many of them neither in vitro nor in vivo testing are available, thus assessing their pathogenic role could be difficult, leading to possible false positive or false negative results. Although several approaches have been developed to prioritize genes, most of them do not consider the role of natural selection and the genomic characteristic of the reference population of the patients. Here we propose an innovative score of gene prioritization based on the population of interest. Firstly, we introduce the concept of singleton-cohort variants (SC variant), which is a variant that has allele count equal to one in the cohort under study. Our hypothesis assumes that if a gene harbours any level of constraints (in a specific cohort), due to its function or from the presence of natural selection, the amount of SC variants in the coding region should be lower than the amount of SC variant in the non-coding regions. Then the difference between the normalized count of SC variants in the coding region respect the normalized count of SC variant in the non-coding region should give us a hint of the level of constraints for that gene in a specific population. This scoring system (from now DSC score) is negative when there are constraints that allow the presence of SC variants only in the non-coding region, on the contrary, is positive when there are no constraints. We calculated the DSC score for all genes in the different subpopulation of 1000 Genome. The DSC score is significant different (t.test p-value<0.00001) between genes deemed essential (median=-0.87) and non-essential (median=0.27) in multiple cultured cell lines based on CRISPR/Cas data. Our methodology showed a high level of constraining for genes such as *USP34*, *PRPF8*, *FBXO11* in all subpopulations; whereas some genes showed high negative score only in specific populations, for examples *HHAT* in Europeans, *BCKDHB* in East Asians, *PARD3B* in South Asian, and *NUP210* in Africans. Our aim is the definition of a gene ranking score that should be based on a population closely related to the patients under study.



## PgmNr 2340: Towards greater equity in genetics research.

**Authors:**

A.J. Shastri; M.E. Moreno; E.F. Bullis; G.D. Poznik; J. O'Connell; S. Das; S. Shringarpure; J. Zhan; the 23andMe Research Team

[View Session](#) [Add to Schedule](#)

**Affiliation:** 23andMe, Inc., Mountain View, California.

---

Genetics research has been far from equitable in terms of engaging representative numbers of non-European participants. 23andMe, with one of the largest research cohorts in the world, aims to increase the diversity of its genetic data and improve its analysis methods so that new discoveries and products benefit people of all racial and ethnic backgrounds. Through internal projects and collaboration with the scientific community, we aim to improve our ancestry and health reports for all customers; facilitate novel discovery in underrepresented populations; and better understand barriers to participation in online genetics research.

Through our newest program, the Populations Collaborations Program, 23andMe partners with researchers engaging underrepresented populations in genetics research globally, providing genotyping and funding for project costs. The goals of this Program are to improve our Reference Data Panel, better understand the extent and patterns of human genetic diversity, and infer human migration and population history. In a complementary initiative, the Global Genetics Program, 23andMe invites individuals from regions underrepresented in our database to participate in research. Together, the data from these efforts has resulted in more granular ancestry reports for non-European customers.

Several internal projects focus on improving GWAS in non-European populations. Through the African American Sequencing Project, 23andMe has sequenced ~2300 consented customers and utilized the data to construct an imputation panel designed to improve imputation for individuals with African ancestry. We evaluated imputation performance within African American research participants and have demonstrated improved imputation accuracy over publicly available reference panels such as Haplotype Reference Consortium and 1000 Genomes. The reference panel is being made publicly available via dbGaP. We have used the panel to perform large-scale GWAS on ~100,000 African American research participants and over 300 phenotypes, yielding novel GWAS hits not previously identified via European GWAS.

Finally, to understand research participation by members of minority populations, we have conducted preliminary qualitative interviews with African Americans on perceptions of genetics and health. Future efforts will build off of these initial investigations with further qualitative and quantitative studies to understand barriers to research participation across populations.

## **PgmNr 2341: Novel way of building reference panels enables finer resolution of ancestry inference among indigenous people of the Americas.**

### **Authors:**

S. Song <sup>1</sup>; D.A. Turissini <sup>1</sup>; A. Sedghifar <sup>1</sup>; Y. Wang <sup>1</sup>; R.E. Curtis <sup>1</sup>; A. Anderson <sup>1</sup>; J. Byrnes <sup>1</sup>; E.L. Hong <sup>1</sup>; K.G. Chahine <sup>2</sup>; C.A. Ball <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Ancestry.com DNA, LLC, San Francisco, California.; 2) AAncestry.com DNA, LLC, Lehi, Utah

---

Ethnicity estimation involves creating reference panels often based on single-origin individuals with roots tracing back to a particular geographic location. However, many modern populations in the Americas are characterized by historic admixture between American, European, and African ancestors. Using research-consented individuals with deep roots in geographic regions, we found substantial genetic substructure across the Americas. We then developed a novel method for creating reference panels using only the inferred indigenous genomic segments from admixed individuals and demonstrate improved resolution and accuracy of ethnicity inference for geographic populations across the Americas.

## PgmNr 2342: Draft assembly of an Armenian genome.

**Authors:**

H. Barseghyan<sup>1</sup>; S. Bhattacharya<sup>1,3</sup>; J. LoTempio<sup>1,2</sup>; B. Dufault<sup>1</sup>; M. Almalvez<sup>1</sup>; E. Vilain<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Center for Genetic Medicine Research, Children's Research Institute, Children's National Medical Center, Washington, DC 20010, USA; 2) Institute for Biomedical Sciences, School of Medicine and Health Sciences, George Washington University Washington, DC 20052, USA; 3) Genomics and Precision Medicine, School of Medicine and Health Sciences, George Washington University Washington, DC 20052, USA

---

Commercialization of short-read sequencing (SRS) in the past decade has allowed characterization of millions of small genetic variations across thousands of human genomes for genetic diagnosis and disease susceptibility associations. However, due to innate methodological limitations of SRS, identification of large structural variants (SVs), such as deletions, insertions, inversions, translocations and copy number variants has proven to be challenging. Like smaller DNA variations, SVs are an important class of human-to-human genetic variability and have been shown to be implicated in disease. New technological advancements in genomics (such as optical genome mapping (OM) or long-read sequencing (LRS)) allow for better assessment of SVs and variants found in highly repetitive regions. However, most variant identification tools utilize a single human genome reference for different types of variant calling ranging from SNVs to SVs. The validity of identified variants in different populations is therefore not accurate due to underrepresentation of other populations in the reference genome. Hence, there is a need for a hybrid-based approach for *de novo* population-specific human genome assembly (reference quality) for variant calling. It is now possible for researchers to survey the diploid architecture of the human genome using multiple genomic technologies for a fraction of the cost of the original Human Genome Project. Here, we present a draft assembly of Armenian genome. The assembly was generated by integrating OM (Bionano Genomics) and LRS (Oxford Nanopore) data sets. OM and LRS molecule N50 values were >200kbp and >50kbp respectively. Coupling of both technologies allowed for an assembly of highly contiguous scaffolds, which in many cases covered entire chromosome arm lengths. Direct comparison of the assembly with the human genome reference GRCh38 identified thousands of previously unseen SVs. Epigenetic modifications were also detected from LRS analysis and mapped across the scaffolds, generating the epigenetic profile of the assembly. This work builds upon other published population-specific genome references necessary for accurate variant characterization in diverse populations. As such, this assembly of the previously unmapped Armenian genome provides a baseline for genomic studies serving the Armenian population and research of population-specific diseases such as Familial Mediterranean Fever, which has a higher frequency in the Armenian population.

## **PgmNr 2343: Understanding the biological meaning of the human genome sequence by studying diverse indigenous populations in East Asia and Southeast Asia.**

**Authors:**

S. Xu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Population Genomics, CAS-MPG Partner Institute for Computational Biology, Shanghai, Shanghai, China

---

Our current biological understanding of the human genome sequence is still very much in its infancy. The functions and phenotypic effects of the majority of our genome remain unknown. There are three typical approaches to understand the biological meaning of our genome sequence. The medical approach relies on measured phenotypes and can not access many lower level genomic functions, while the experimental approach can only be done in non-human organisms or cell lines. Evolutionary analysis of diverse populations can access subtle consequences of individual mutations in humans, thus providing an economic and efficient way to sensitively probe the function of our genome sequence. Here we build a reference genomic resource of 977 diverse populations, including 1,009 newly sequenced genomes representing 16 indigenous populations living in unusual environments (e.g., tropical forests and highlands) in East Asia and Southeast Asia. This effort is expected to i) enable studies of variants that are rare or not existing in well-studied populations; ii) improve the interpretations of putative causal loci associated with Mendelian diseases and adaptations to local environments in global populations; and iii) help advance our understanding of the biological meaning of the human genome sequence.

## PgmNr 2344: Genome-wide association study identifies new regions associated with iris surface features in individuals of European ancestry.

### Authors:

F. Lona Durazo; M. Edwards; D. Cha; E.J. Parra

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Anthropology, University of Toronto, Mississauga, Ontario, Canada

---

The human iris is one of the most complex structures on the outside of the human body, in which several patterns influence the physical appearance and color of the eye. The two top cell layers of the iris, which arise during embryological development from the neural crest, give rise to the iris surface features, due to their composition and arrangement. These features are highly heritable, and they ultimately cause differences in eye appearance among individuals. However, research on the genetic basis of iris surface features is limited to a handful of studies, which have highlighted the association of a few genes involved in neurodevelopment. Here, we conducted a genome-wide association study to investigate the genetic basis of three iris surface features: Fuchs' crypts, Wolfflin nodules, and pigment spots, in a sample of individuals of European ancestry ( $N = 541$ ), using high-resolution photographs processed in a web-based application for phenotype characterization. We identified a region in chromosome 2 that is significantly associated with Wolfflin nodules (non-synonymous polymorphism: rs3739070,  $p$ -value =  $4.55 \times 10^{-8}$ ), overlapping *TRAF3IP1*, a gene which has been previously associated with the frequency of contraction furrows. Additionally, we identified two suggestive signals associated with Wolfflin nodules: a region in chromosome 12 overlapping the 3'UTR of *SRGAP1*, which is a gene involved in 3D collagen migration (top marker: rs1067638,  $p$ -value =  $5.43 \times 10^{-7}$ ), and a region in chromosome 2 overlapping *ERBB4*, which is a gene involved in neuronal migration and axon guidance (top marker: rs839500,  $p$ -value =  $3.12 \times 10^{-6}$ ). We also identified suggestive signals of association for pigment spots and Fuchs' crypts, which overlap genes involved in neurodevelopment, such as *DTNB* (top marker: rs9789762,  $p$ -value =  $6.49 \times 10^{-6}$ ) and *LSAMP* (top marker: rs4487267,  $p$ -value =  $3.67 \times 10^{-6}$ ), respectively. We are currently working on a meta-analysis in order to increase the power of our analyses and to possibly reinforce our current results.

## PgmNr 2345: Patterns of ROH across species: From a dog to a human.

### Authors:

A. Anderson; J. Miglo; S. Hoque; S. Smieszek; C. Polymeropoulos; M. Polymeropoulos

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanda Pharmaceuticals, WASHINGTON, D.C..

---

Animals are used as preclinical models for human diseases in drug development. Dogs, especially, are used in preclinical research to support the clinical safety evaluations during drug development, however, comparison of regions of homozygosity (ROH) patterns and phenotypes between dog and human are not well known. A ROH is defined as a continuous stretch of DNA sequence without heterozygosity in the diploid state (min ROH  $\geq 1.5$  Mb). We conducted a genome-wide homozygosity analysis (GWHA) in human and dog genomes.

We calculated ROH patterns across distinct human cohorts: the Amish, IGSR 1000 genomes, Welllderly, Vanda 1k genomes, and ADNI Alzheimer's data. The Amish provided a large cohort of extended kinships allowing for in depth family oriented analyses whereas Welllderly and IGSR 1000 genomes represented healthy aging populations. The remaining human cohorts served as statistical references. We then calculated ROH across different dog breeds (EMBARC project) with emphasis on the beagle, dog that is the preferred breed in drug development.

Vanda 1k and the Welllderly ROH analysis was conducted on whole genome sequencing data. Out of 5 human cohorts we reported the highest mean ROH in the Amish population. When we considered both the location and the allelic form of the ROHs, we were able to separate the populations by PCA, demonstrating that ROHs contain information on the demographic history and structure of a population. We next calculated the extent of the genome covered by ROH ( $F_{ROH}$ ) (human 3.2Gb, dog 2.5Gb). In our dataset,  $F_{ROH}$  differed significantly between the Amish and the 1000 genomes, and between the human and the beagle genomes. The mean  $F_{ROH}$  per 1Mb was  $\sim 16$ kb for Amish,  $\sim 0.6$ kb for Vanda 1k, and  $\sim 128$ kb for beagles. This result demonstrated the highest degree of inbreeding in beagles, far above that of the Amish, one of the most inbred human populations.

ROH can contribute to inbreeding depression if they contain deleterious variants that are fully or partially recessive. The differences in ROH characteristics between human and dog genomes question the applicability of dog models in preclinical research, especially when the goal is to gauge the subtle effects on the organism's physiology produced by candidate therapeutic agents. We should hesitate to generalize from dog to human, even if human and beagle are relatively close species phylogenetically.

## PgmNr 2346: Patterns of cross ancestry constraint in regulatory regions.

### Authors:

T.J. Hayeck <sup>1</sup>; N. Stong <sup>1</sup>; E.H. Baugh <sup>1</sup>; R. Dhindsa <sup>1</sup>; T. Turner <sup>2</sup>; I. Ionita-Laza <sup>1</sup>; D.B. Goldstein <sup>1</sup>; A.S. Allen <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Genomic Medicine, Columbia University, BROOKLYN, New York.; 2) University of Washington Department of Genome Sciences 3720 15th Ave NE, Seattle WA 98195-5065; 3) Duke Center for Statistical Genetics and Genomics, Duke University, Durham, NC 27710, USA

---

Evolutionary conservation is often used to identify functional elements in genomes. It relies on isolating regions with fewer differences than would be expected given divergence times between species. This makes elements that are of functional consequence in only one species difficult to detect. Recent studies have shown that estimates of human specific constraint using standing population variation improves the detection of pathogenic variants over cross species conservation. When applied to noncoding sequence where functional annotations are often unavailable, these constraint estimation approaches require a theoretical or empirical neutral model to identify regions where variation is depleted relative to neutral expectation. These neutral models are distorted by demography, we find substantial differences in the neutral site frequency spectrum (SFS) across populations. For example, SFSs generated from sampling a million intergenic bases from the genome aggregation database (gnomAD) 4,368 African/African-American and 7,509 Non-Finish European samples, demonstrate significant differences in their SFSs (permutation based logrank test:  $p < 1e-8$ ). This suggests that looking for shifts in the SFS without taking into account ancestry may produce population specific artifacts and a potential loss of power. Here we improve constraint estimates by explicitly considering demographic differences due to ancestry. Specifically, we developed a non-parametric statistical framework to test for shifts in the SFS across the genome relative to neutral intergenic variation, which is stochastically sampled within ancestry groups fitting on 15,496 whole genome samples from gnomAD. we show that by explicitly accounting for ancestry we substantially improve classification of pathogenic variation, even in non-coding sequence. For example, in classifying non-coding ClinVar pathogenic variants versus a million randomly sampled common variants taken from 62,784 whole genome Trans-Omics for Precision Medicine (TOPMed projects) samples, AUCs are significantly improved when ancestry is adjusted for (AUC=90%) relative to a similar unadjusted approach (AUC=80%). It also substantially outperforms a previously proposed approach for classifying noncoding sequence (CDTS: AUC=76%). Our results demonstrate the importance's of explicitly considering demographic factors in constraint estimation and the value of diverse large population samples of standing variation across world populations.

## **PgmNr 2347: A reference free method to identify introgressed hominin sequences reveals insights into Neanderthal ancestry in African and non-African populations.**

### **Authors:**

L. Chen <sup>1</sup>; A. Wolf <sup>1,2</sup>; W. Fu <sup>3</sup>; J. Akey <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Ecology and Evolutionary Biology, PRINCETON UNIVERSITY, PRINCETON, New Jersey.; 2) Department of Genome Sciences, University of Washington, Seattle, WA.; 3) Microsoft, Redmond, WA

---

Admixture has played a prominent role in shaping patterns of human genomic variation, including gene flow with now extinct hominins like Neanderthals and Denisovans. We describe a novel probabilistic method called IBDmix to identify introgressed hominin sequences, which unlike existing approaches, does not use a modern reference population. We applied IBDmix to 2,504 geographically diverse populations to identify and analyze Neanderthal sequences segregating in modern humans. Strikingly, we find African individuals have substantially more Neanderthal ancestry (0.3%) than previously thought and show this can be explained by migrations back to Africa, predominately from ancestral Europeans. Moreover, IBDmix recovers comparable amounts of Neanderthal sequence among non-African populations, in contrast to notable differences observed using reference-based methods. Our results refine our understanding of Neanderthal ancestry in African and non-African populations, simplify models of archaic hominin admixture, and demonstrate that remnants of Neanderthal genomes survive in every modern human population studied to date.



## PgmNr 2348: Natural selection shapes codon usage in the human genome.

### Authors:

R.S. Dhindsa <sup>1</sup>; B.R. Copeland <sup>1</sup>; S. Petrovski <sup>2</sup>; A.M. Mustoe <sup>3</sup>; D.B. Goldstein <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Genomic Medicine, Columbia University Medical Center, New York, NY.; 2) Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK; 3) Department of Chemistry, University of North Carolina, Chapel Hill, NC, USA

---

Synonymous mutations change the sequence of a gene without altering the encoded protein sequence. Though these mutations are typically considered neutral, changes in codon usage can impact multiple molecular processes. In particular, optimal codons, which correspond to tRNAs that are highly abundant in the cytoplasmic pool, increase translational efficiency, resulting in increased mRNA half-life and protein output. In this study, we leveraged the unprecedented amount of sequencing data available in two population reference cohorts—gnomAD and TopMED—to demonstrate that natural selection optimizes codon content in protein-coding regions in the human genome. We devised two scores to rank genes by their intolerance to synonymous mutations. The first metric, synRVIS, measures human-specific constraint against changes in codon optimality. The second metric, synGERP, reflects phylogenetic conservation at fourfold degenerate sites across the mammalian lineage. Consistent with the known role of codon optimality in regulating gene expression and translation efficiency, we find that both of these scores are predictive of dosage sensitivity. tRNA levels are dynamically regulated in stress conditions and during the cell cycle. Accordingly, we find that genes in the DNA damage response pathway and periodically expressed cell cycle genes display intolerance patterns that reflect these dynamic expression patterns. Finally, we demonstrate that both synRVIS can predict oncogenes and tumor suppressor genes. Importantly, we find that single synonymous variants that reduce codon optimality in *BRCA1* may result in significantly reduced protein function. Overall, our work provides some of the first direct evidence that natural selection influences human codon usage and has strong implications for the interpretation of synonymous variants in the human genome.

## PgmNr 2349: The genetic consequences in the Americas of the transatlantic slave trade.

### Authors:

S. Micheletti <sup>1</sup>; K. Bryc <sup>1</sup>; S. Ancona Esselmann <sup>1</sup>; W. Freyman <sup>1</sup>; M. Moreno <sup>1</sup>; A. Shastri <sup>1</sup>; S. Beleza <sup>2</sup>; J. Mountain <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) 23andMe, 899 W Evelyn Ave, Mountain View, CA 94041, US; 2) University of Leicester, University Rd, Leicester LE1 7RH, UK

---

The transatlantic slave trade marks the largest forced migration event in history, involving the deportation of an estimated 12.5 million Africans. Occurring between the 15th and 19th centuries, the slave trade involved nearly every country with an Atlantic coastline and has continued to impact the world more than a century after its abolition. Even though shipping records document the number of enslaved people disembarking across the Americas, there have been very few broad-scale genetic studies of populations across all of the Americas with ancestral ties to the transatlantic slave trade. In this study, we utilized high-density genotype data from over 25,000 23andMe research participants with western African ancestry to determine how the transatlantic slave trade has shaped populations across twenty-eight regions within the United States, Latin America, and the Caribbean. Using 23andMe's Ancestry Composition algorithm, which detects ancestry from multiple subregions of western Africa, we show that individuals with African ancestry in the Americas tend to have connections with multiple historical slave trading regions in western Africa. However, the most frequently inferred western African ancestries differ across the Americas. For instance, individuals in the United States and Caribbean tend to have connections with all historical slave trading regions in western Africa, while individuals from Latin America tend to have ancestry primarily from either west central Africa or Senegambia. To estimate the temporal connection between populations in the Americas and West Africa, we identified DNA identical by descent (IBD) between all twenty-eight disembarkment regions and seven historical embarkment regions in western and southwestern Africa. We used the distribution of IBD segment lengths to determine the probability that shared IBD between regions were inherited from a common ancestor who embarked from western Africa during the 400-year period of the transatlantic slave trade. Results suggest that genetic estimates of time to most recent common ancestor are generally concordant with documented embarkment dates from western Africa. Overall, this study provides an unprecedented investigation of the transatlantic slave trade, establishing genetic links between populations in the Americas to populations in western Africa, which has implications for individuals' understanding of their African roots.

## PgmNr 2350: Multi-omics analysis of Qatar Biobank reveals causal functional consequences of Middle Eastern genetic variation and their impact on human traits and disease.

### Authors:

Y. Mokrab<sup>1,3,4</sup>; R.M. Razali<sup>1</sup>; H. Naeem<sup>1</sup>; F.R. Vempalli<sup>1</sup>; H. Almabrazi<sup>1</sup>; R. Temanni<sup>1</sup>; N. James<sup>1</sup>; V. Mattei<sup>1</sup>; M. Al Hashmi<sup>1</sup>; K. Wang<sup>1</sup>; L. Mathew<sup>1</sup>; S. Lorenz<sup>1</sup>; R. Al Ali<sup>1</sup>; X. Estivill<sup>1</sup>; D. Chaussabel<sup>1,4</sup>; S. Tomei<sup>1</sup>; A. Akil<sup>1</sup>; K.A.M. Fakhro<sup>1,3,4</sup>; S.B. Montgomery<sup>2</sup>; Qatar Genome Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Precision Medicine, Research Branch, Sidra Medicine, Doha, Qatar; 2) Stanford University School of Medicine, Stanford University, USA; 3) Weil Cornell Medicine, Qatar; 4) Hamad Bin Khalifa University, Qatar

---

A major goal of human population genetics studies is to unveil genetic polymorphisms and understand their influence on phenotype in health and disease. Sequencing projects e.g. HapMap, 1KG and recent national efforts have documented extensive genetic variation and LD in various populations. Yet, our understanding of its origins, global distribution, and functional consequences is far from complete. In particular, regulatory impact on gene expression has mostly been studied in small population sizes, and breadth of transcriptome is yet to be revealed on large scales mirroring the sequencing efforts. Also, studies so far lack enough diversity for regions such as Middle East & North Africa.

Qatar Genome Project aims to obtain population-level insight into genetic variation in Qatar and the wider Arabian Peninsula. So far, it has sequenced whole genomes of >10k subjects using Illumina HiSeq X ten at 30x coverage, identifying ~ 30m novel variants. These predominately-healthy subjects were enrolled at Qatar Biobank and were deeply phenotyped with diverse data on demographics, health, lifestyle, physical measurements and >66 biochemical tests. With this in hand, we present analysis of the pilot phase RNAseq (mRNA & small RNA) from whole blood of 3k subjects from this cohort and combine with genetic and phenotypic data to characterize the polymorphism discovered and its interplay with molecular and clinical phenotypes at a population scale.

Hamilton-extracted total RNA had RNA Quality Number >7.5 and yield > 1 µg for >70% of samples. 1,076 globin-depleted libraries were sequenced at 20-100m 150 pair-end reads on Illumina HiSeq 4000. High quality reads (average phred score > 30) were mapped using STAR (Dobin et al, 2013) showing >90% mapping rate followed by TMM-based normalisation (Robinson et al, 2010) of the feature counts (Liao et al, 2014). ~70 % of genes are expressed with ≥0.1 TPM in ≥20% samples. We run hidden factor analysis to identify confounding factors and correlated with available co-variables. We are running eQTL/sQTL/aseQTL analysis for various genetic variation classes, identifying novel regulatory rare variants e.g. for KCNJ2 associated with lipid levels. Also, we are applying colocalization approaches integrating genetic associations using previously applied approaches (Liu et al, 2019). Together, our study provides a comprehensive view on how regulatory and splicing variation impacts traits and diseases in a Middle Eastern population sample.

A- A+

## **PgmNr 2351: Visualizing dominant patterns of allele frequency variation as a means to understand population structure in genetic data.**

### **Authors:**

J. Novembre; A. Biddanda

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, Univ Chicago, Chicago, Illinois.

---

Typical approaches for showing population structure (e.g. PCA, Structure/Admixture, and EEMS) are limited in that they do not reveal how allele frequencies vary across populations. Most methods show how individuals are related to one another and help reveal the existence of sub-structure – However, they do a poor job of showing how typical alleles are distributed. This is problematic because it is easy for viewers to observe, for example, clusters in a PCA plot and falsely assume there is 'deep' differentiation, when in fact the average allele frequency varies little between populations. This has impacts in research settings as it affects how investigators understand their study populations, and in pedagogical settings for how students or the lay public understand human genetic diversity. Here, we present a simple way of showing the dominant allele frequency patterns among a set of SNPs and have applied it to the 1000 Genomes data and a few other datasets. The visualizations give a different perspective on human genetic structure, that helps viewers answer more immediately important empirical questions such as: How often should a common GWAS variant found in one population be expected to be found in a second population? How typical is it to see an allele that is highly differentiated among human groups? Overall, we argue that we have a useful tool to provide a variant-centric view of population structure, that we expect will be of broad utility to the population genetic community.

A - A+

## **PgmNr 2352: Reconstructing the peopling of old world South Asia: From modern to ancient genomes.**

### **Authors:**

N. Rai

[View Session](#) [Add to Schedule](#)

**Affiliation:** BIRBAL SAHNI INSTITUTE OF PALAEOSCIENCES, LUCKNOW, UTTAR PRADESH, India

---

South Asia was one of the first geographic regions to be peopled by modern humans after their African exodus. Today, the diverse ethnic groups of South Asia comprise an array of tribes, castes, and religious groups, who are largely endogamous and have hence developed complex, multi-layered genetic differentiation. From such a complex structure, several questions have stood out from the research of our group and others that are only beginning to be resolved using modern sequencing techniques and targeted sampling of populations and archaeological specimens. In this forum, I will discuss the complex population structure of the Indian subcontinent and future research directions to understand the deep ancestry components of South Asians and Eurasians in the global context, patterns of admixture, and migrations, as revealed by the study of ancient and modern genomes.

## PgmNr 2353: Admixture mapping of TB susceptibility in a five-way admixed southern African population.

**Authors:**

Y. Swart; C. Uren; M. Möller; E.G. Hoal

[View Session](#) [Add to Schedule](#)

**Affiliation:** DST/NRF Centre of Excellence for Biomedical TB Research, SAMRC Centre for Molecular Biology and Cellular Biology, Division of Molecular Biology and Human Genetics, Stellenbosch University, Cape Town, SA

---

Tuberculosis (TB), caused by the human pathogen *Mycobacterium tuberculosis* (*M.tb*), is ubiquitous in almost all populations residing in southern Africa. A number of studies indicated a disparity in the rates that different ethnic groups are infected with the TB bacterium and progress to active TB disease. The difference is not merely a reflection of socio-economic circumstances. Centuries of exposure to TB in Europe may have resulted in adaptive selection for TB resistance, in comparison to groups in sub-Saharan Africa that were only relatively recently exposed to the virulent European *M.tb* strains. Previous investigations of a highly admixed southern African population not only revealed indigenous KhoeSan ancestry in the region, but also indicated that African ancestry are associated with an increased risk to progress to active pulmonary TB rather than remaining latently infected. However, these studies relied on a limited number of controls, inadequate sample size for reference populations and low SNP density. This study focuses on localizing and investigating ancestry-specific genetic regions associated with TB susceptibility in a unique complex five-way admixed South African Coloured population. Admixture mapping of 820 individuals (413 cases and 407 controls) was done using data generated from the Illumina Multi Ethnic Genotyping Array, a more appropriate tool for our diverse and highly admixed population. Our local ancestry results identify the ancestral origins of distinct chromosomal segments with a high degree of accuracy and allow for the identification of genetic regions associated with TB susceptibility. The adjusted allelic association results point to several candidate regions.

## PgmNr 2354: Identification of potentially protective genomic variants in the Lithuanian population group.

### Authors:

G. Žukauskaitė; L. Ambrozaitė; I. Domarkienė; T. Ranevičiūtė; V. Kučinskis

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Human and Medical Genetics, Institute of Biomedical Sciences, Vilnius University, Vilnius, Lithuania

There is a growing number of reported genomic variants shown to be protective, lowering the risk of certain diseases and conditions. The identification of such variants may serve as new potential drug targets and provide new therapeutic opportunities as well as deepen our understanding of the mechanisms of microevolution. The aim of this study was to determine protective genomic variants (PGVs) in the general Lithuanian (LT) population. This study is the first of its kind to analyze PGVs in the LT population.

Exome sequencing (*AB SOLiD*) was performed for 98 self-reported healthy individuals of the LT descent. A catalog of 123 known PGVs was constructed using *ClinVar*, *OMIM* databases as well as scientific publications. The catalog was used to sort out the PGVs that are common in the LT population from the exome sequencing data – 54 PGVs were identified. Frequencies of genotypes of these variants were determined and compared to the frequencies of the general European population from 1000 Genomes Project. Statistical analysis ( $\chi^2$  or F-test,  $\alpha=0.05$ ) was performed using *Rstudio* software. Four variants were significantly more frequent in the study group and may have protection against **alcohol dependence** (*ADH1C*, rs698,  $p=0.05$ ), **type 2 diabetes** (*PPARG*, rs1801282,  $p=0.005$ ), **obesity** (*SH2B1*, rs7498665,  $p=0.03$ ), **esophageal cancer** (*PLCE1*, rs2274223,  $p=0.034$ ). *ADH1C* encodes the class I alcohol dehydrogenase. Our results show that 50% of the LT study group carry a genomic variant (NM\_000669.4:c.1048A>G), which is denoted as protective against alcohol dependence in different databases. However, studies suggest that this variant is associated with slower ethanol metabolism and therefore people carrying it have a higher risk of heavy and excessive drinking (Tolstrup J S, 2008). These conclusions support the statistics of World Health Organization, which reports that Lithuania was the second country in Europe by consumption of alcohol in 2018. Significantly higher frequencies in our study group might partly explain such behavior of alcohol consumption. Therefore, we hypothesize that this variant may be more causal than protective and an understanding of the molecular mechanisms involved in excessive alcohol consumption is still unresolved. Other three identified variants are considered as potentially protective according to bioinformatic methods applied (Butler J M, 2017). However, functional impact of all these variants should be confirmed and additional analysis is needed.

## PgmNr 2355: Parameter reduction of sequence-context mutation models.

**Authors:**

C.J. Adams<sup>1</sup>; B.F. Voight<sup>2,3</sup>

[View Session](#)[Add to Schedule](#)**Affiliations:**

1) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 2) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 3) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

---

Statistical models that predict where, what type, and how frequently polymorphisms occur across the human genome have become essential to addressing basic questions in population genetics, including identifying regions under selective constraint, discovering functional genomic elements, and estimating divergence times. Basic knowledge of the mutational process is also of critical importance for phenotypic studies; for example, characterizing the parts of the genome that harbor a burden of rare variation contributing to complex disease, or nucleotide sequences that are selectively constrained and most likely to create phenotypic consequences if mutated (i.e., pLI scores).

A fundamental challenge to advancing these models is simply that the number of parameters grows exponentially as the window size of sequence context considered and number of genomic features increase. This property limits statistical power and increases error for biological inference in existing models. To address this challenge, we present an algorithm to cluster mutational parameters within models of nucleotide sequence context. We demonstrate for the first time that nonanucleotide ('9-mer') sequence-context windows explain patterns of variation better than the previously reported heptanucleotide ('7-mer') context model ( $P < 1 \times 10^{-100}$ ). We further demonstrate that a collapsed form of the 9-mer sequence context model recovers 99.99% of the likelihood using <1% of the parameters, facilitating novel hypotheses to be tested by "freeing" these parameters. Application of our approach enables systematic tests of novel hypotheses, including: (i) asymmetric sequence context models where different sizes of flanking nucleotides around the polymorphic site are considered, (ii) models with genomic features, like replication timing, nested within a condensed model, and (iii) genome-wide tabulation of scores to quantify local reduction of polymorphism levels – consistent with the action of background selection – that can be utilized in phenotypic studies.



## PgmNr 2356: Familial language history links with population structure and admixture patterns.

**Authors:**

N. Telis <sup>1</sup>; H. Arbel <sup>1</sup>; E. Sorokin <sup>1</sup>; A.R. Girshick <sup>1</sup>; M. Zhang <sup>1</sup>; J.M. Granka <sup>1</sup>; E.L. Hong <sup>1</sup>; K.G. Chahine <sup>2</sup>; C.A. Ball <sup>1</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) DNA Science, Ancestry, San Francisco, California; 2) DNA Science, Ancestry, Lehi, UT

---

Language and population structure are intricately connected in human history. Past genetic studies have found that language can roughly mirror population genetics, though the two are not exactly identical. We pilot an approach using linguistic change to approximate recent fine-scale population structure. In a large cohort with rich pedigree information, we examine the relationship between genetics and recent immigration patterns, which we study using reported grandparent and reported individual language. We identify subtle differences in genetic ancestry composition, likely determined by both subpopulation structure and admixture patterns. We leverage maternal/paternal language information to further probe these differences in ancestry, identifying likely recent sex-biased admixture through linguistic differences. Using simulations, we estimate the impact of sex-biased admixture on sex chromosome local ancestry, and downstream effects of this impact on association statistics through population structure confounding. This suggests linguistic patterns may be a reasonable proxy for substantially more complex population structure, and that this population structure can have a quantitative impact on measures of genetic architecture admixed populations.

## PgmNr 2357: Ancient retroviruses drove the rapid evolution of mammalian APOBEC3 family genes.

### Authors:

J. Ito <sup>1</sup>; R. Gifford <sup>2</sup>; K. Sato <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Systems Virology, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan; 2) MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, UK; 3) CREST, Japan Science and Technology Agency, Saitama, Japan

---

The AICDA/APOBEC family genes (AIDCA and APOBEC1-4) encode cellular cytidine deaminases that target nucleic acids and catalyze cytosine-to-uracil (C-to-U) mutations. In the case of retroviral replication, APOBEC3 proteins induce C-to-U mutations in minus-stranded viral DNA, which results in guanine-to-adenine (G-to-A) mutations in the viral genome. Previous studies have implied that the expansion and rapid evolution of mammalian APOBEC3 genes has been driven to by an arms race with retroviral parasites. However, this possibility has not been thoroughly investigated. Endogenous retroviruses (ERVs) are the ancient relics of retroviral infections. These sequences sometimes bear the hallmarks of APOBEC3-mediated mutations, and therefore serve as a record of the ancient conflict between retroviruses and A3 genes. Here we systematically investigated the sequences of ERVs and APOBEC3 genes in mammals to reconstruct details of the evolutionary conflict between them. We identified 1,561 AICDA/APOBEC family genes in a comprehensive screen of mammalian genome. Of the AICDA/APOBEC family genes, APOBEC3 genes have been selectively amplified in mammalian genomes, and disclose evidence of strong positive selection – whereas the catalytic domain was highly conserved across species, the structure “loop 7”, which recognizes viral DNA/RNA substrates, was shown to be evolving under strong positive selection. Comparative analysis of mammalian ERV sequences revealed that G-to-A mutations were accumulated in ERVs, and that the sequence signature of the G-to-A mutation sites on ERVs was concordant with the target preferences of APOBEC3 proteins. Importantly, the number of APOBEC3 genes was significantly correlated with the frequency of G-to-A mutations in ERVs, suggesting that the amplification of APOBEC3 genes led to enhance the attacks on ERVs by APOBEC3 proteins. Furthermore, the numbers of APOBEC3 genes and ERVs in mammalian genomes were positively correlated, and the timings of APOBEC3 gene amplification and ERV invasions in primates were highly concordant. Our findings suggest that conflict with ancient retroviruses and ERVs was a major selective pressure driving the rapid evolution of APOBEC3 genes in mammals.

## PgmNr 2358: Rome as a genetic melting pot: Population dynamics over 12,000 years.

### Authors:

M. Antonio <sup>1</sup>; Z. Gao <sup>2,3</sup>; H. Moots <sup>4</sup>; M. Lucci <sup>5</sup>; F. Candilio <sup>6,7</sup>; S. Sawyer <sup>8</sup>; V. Oberreiter <sup>8</sup>; K. Devitofranceschi <sup>8</sup>; R. Aikens <sup>1</sup>; S. Aneli <sup>9</sup>; D. Cotter <sup>2</sup>; D. Fernandez <sup>8</sup>; F. La Pastina <sup>5</sup>; G. Matullo <sup>9</sup>; A. Nava <sup>10</sup>; R. Schmidt <sup>11</sup>; A. Sperduti <sup>12,13</sup>; M. Rubini <sup>14</sup>; L. Bondioli <sup>12</sup>; A. Coppa <sup>5</sup>; R. Pinhasi <sup>8</sup>; J. Pritchard <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical Informatics Program, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA; 3) Howard Hughes Medical Institute, Stanford University, Stanford, CA; 4) Department of Anthropology, Stanford University, Stanford, CA; 5) Dipartimento di Biologia Ambientale, Sapienza Università di Roma; 6) Soprintendenza Archeologia, Sardinia; 7) University of Pennsylvania, Philadelphia, Pennsylvania; 8) University of Vienna, Vienna, Austria; 9) University of Torino, Turin, Italy; 10) DANTE Laboratory for the study of Diet and Ancient Technology, Sapienza Università di Roma; 11) University College Dublin, Dublin, Ireland; 12) Servizio di Bioarcheologia, Museo delle Civiltà, Roma, Italy; 13) Università L'Orientale Napoli, Italy; 14) Department of Archaeology, Foggia University, Foggia, Italy

---

Nearly 2000 years ago, Rome was the largest urban center of the ancient world and the capital of an empire with over 60 million inhabitants. Although Rome has long been a subject of archaeological and historical study, little is known about the genetic history of the Roman population. To fill this gap, we performed whole genome sequencing on 127 individuals from 29 sites in and around Rome, spanning the past 12,000 years. Using allele frequency and haplotype-based genetic analyses, we show that Italy underwent two major prehistoric ancestry shifts corresponding to the Neolithic transition to farming and the Bronze Age Steppe migration, both prior to the founding of the Roman Republic. As Rome expanded from a small city-state to an empire controlling the entire Mediterranean, the city became a melting pot of inhabitants from across the empire, harboring diverse ancestries from the Near East, Europe and North Africa. Furthermore, we find that gene flow between Rome and surrounding regions closely mirrors Rome's geopolitical interactions. Interestingly, Rome's population remains heterogeneous despite these major ancestry shifts through time. Our study provides a first look into the dynamic genetic history of Rome from before its founding, into the modern era.

## PgmNr 2359: Genomic game of thrones: Ancient DNA analysis of European and Asian royal dynasties.

### Authors:

O. Balanovsky <sup>1,2,3</sup>; M. Allentoft <sup>4</sup>; W. Haak <sup>5</sup>; D. Pezhemsky <sup>6,7</sup>; V. Volkov <sup>8</sup>; M. Zhabagin <sup>9</sup>; V. Zaporozhchenko <sup>2</sup>; E. Lukyanova <sup>1</sup>; T. Gunther <sup>10</sup>; E. Balanovska <sup>2,3</sup>; J. Krause <sup>5</sup>; E. Willerslev <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vavilov Institute of General Genetics, Moscow, Russian Federation; 2) Research Centre of Medical Genetics, Moscow, Russia; 3) Biobank of North Eurasia, Moscow, Russia; 4) Center for Geogenetics, Copenhagen University, Copenhagen, Denmark; 5) Max Planck Institute for the Science of Human History, Jena, Germany; 6) Institute and Museum of Anthropology, Moscow State University, Moscow, Russia; 7) Paleoethnology Research Center, Moscow, Russia; 8) Tomsk State Pedagogical University, Tomsk, Russia; 9) Center for Bioengineering, Nur-Sultan, Kazakhstan; 10) Uppsala University, Uppsala, Sweden

In most societies, aristocracy has a marriage pattern different from that of the general population. Dynasties of rulers often had foreign origin and thus might be genetically distant. Moreover, they tend to marry their relatives (which genetically would result in elevated inbreeding) and partners from other countries (genetically counteractive). A genomic analysis of royal families is therefore interesting not only for the sake of accurate historical identification, but also for addressing population genetic questions on an “aristocratic subpopulation”.

We analyzed genomic data representing three dynasties: the medieval North European Rurikid dynasty, the 19<sup>th</sup> century Russian Romanov dynasty, and Genghis Khan’s descendants.

The chronicles indicate that Rurikid princes, which ruled Russia over seven centuries, descent from the Viking Rurik. The genetic-genealogical analysis of the present-day princes identified the core genetic lineage rooted in Scandinavia. We analyzed the ancient DNA of medieval princes including Vladimir (11th century) who is placed in traditional genealogy just five generations apart from the Rurik, and compared the ancient and modern members of the Rurikid dynasty.

The second study was performed in parallel by the two ancient DNA labs which were approached by the Russian Church requesting kinship analyses of the anonymous bone samples. The analysis was made as a blind test, the investigation was paralleled, and the results matched. The mtDNA profiles coincided with the published data on the last Russian emperor and empress (Rogaev et al., 2009), but the new results included also the preceding generations of the dynasty. The genomic results corroborate previous PCR-based findings and will settle the remaining doubts around the authenticity of Royal family remains. Genetically, these individuals are consistent with being Central Europeans, in agreement with the genealogical records, as they were close relatives of the Danish king, Queen Victoria of the United Kingdom, and other Central European dynasties.

Finally, we analyzed Y-chromosomes of several clans claiming their origin from Genghis Khan.

Surprisingly, the genetic variation in this sample was even higher than in general population. While the Nyru’un tribe (Genghis Khan belonged to) in general carried high frequency of the “star-cluster” (C-F3796), most studied aristocratic clans from Kazakhstan, Mongolia, and China carried two other lineages (C-F1756 and C-M407).

## PgmNr 2360: Linkage disequilibrium and haplotype patterns in ancient DNA: Theory and applications.

### Authors:

A. Biddanda <sup>1</sup>; M. Steinrücken <sup>2</sup>; J. Novembre <sup>1,2</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Univ Chicago, Chicago, Illinois.; 2) Department of Ecology and Evolution, Univ Chicago, Chicago, Illinois

---

The rise of ancient DNA studies in population genetics has been revolutionary – revealing insights into demographic history and recent positive selection. However, most studies to date have ignored the covariance between variants on haplotypes (i.e. linkage disequilibrium, LD). Basic properties of LD in samples of mixed ages are still not well understood, in part because much of population genetic theory was developed without an expectation of time-serial sampling. Here, we derive several results for summary statistics of haplotypic variation. We derive results for the: 1) The expected number of pairwise differences between time-staggered samples ( $\pi_{\{\Delta t\}}$ ) and the correlation of this statistic between two loci, in models with and without strict population continuity; 2) The time-staggered analog of the LD-metric  $\sigma_D^2$ , which can be interpreted as a measure of haplotypic similarity between a modern and ancient sample; 3) The backward-in-time evolution of a set of LD-based moments defined by Hill and Robertson (1968) – these results can be used to predict aspects of LD in the past, based on observations in the present; and 4) The expected switch rate in a haplotype copying model, which suggests how to improve and use haplotype copying models with ancient samples and modern reference panels. Overall, these results provide a characterization of how several haplotypic patterns are affected by the time-gap between sampling, recombination rates, and population size. We expect these results will help guide thinking and methodological development for the analysis

## PgmNr 2361: Imputing gene regulation in ancient humans.

**Authors:**

L.L. Colbran <sup>1</sup>; M.R. Johnson <sup>2</sup>; J.A. Capra <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, Tennessee.; 2) Vanderbilt School for Science and Math, Vanderbilt University, Nashville, Tennessee; 3) Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee.

---

As anatomically modern humans (AMHs) migrated out of Africa and spread throughout the world, they were exposed to many different environments, including changes in climate, diet, and lifestyle. Given that closely related species are often very similar in their protein-coding sequences (e.g., human and chimpanzee are ~99% identical), and that most variants associated with disease in modern humans are non-coding, we expect most of the functional genetic changes that have occurred in recent evolutionary history to influence gene regulation. Recent technological advances in genotyping ancient DNA provide a novel avenue for studying when genetic changes occurred, and in what populations they arose, as well as the ability to correlate those changes with environmental and demographic shifts.

Because of the age of the samples involved, it is not possible to directly assay gene expression levels in ancient human samples or the effects of regulatory variants in ancient genomes. Here, we explore the feasibility of using PrediXcan, a method for imputing gene regulation based on genotypes, to study ancient human gene regulation. We have previously shown the utility of PrediXcan in studying gene regulatory patterns in archaic hominins with high-coverage sequenced genomes. However, most ancient human genotype data is very low-coverage, rendering existing gene regulatory models unstable due to high rates of missingness. To address this and assess the feasibility of using PrediXcan on data of this type, we trained new models in 48 tissues, optimized for use with an ancient human dataset. We will report the results of applying these models in 429 ancient humans from Eurasia ranging from 318 to 45000 years ago. We also compare the patterns of gene regulation by lifestyle and time period, as well as the potential phenotypic effects of the genes showing divergence in regulation. Altogether, our results assess the feasibility of applying gene expression imputation models in low-coverage data and demonstrate the potential of these methods for understanding patterns of gene regulation in ancient humans.

## PgmNr 2362: Comparing archaic to modern human-derived sites in the genome reveals the evolution of human facial morphology.

### Authors:

C. Liu <sup>1</sup>; M. Lee <sup>2</sup>; M. Marazita <sup>1,2</sup>; S. Weinberg <sup>1,2,3</sup>; J. Shaffer <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA; 3) Department of Anthropology, University of Pittsburgh, Pittsburgh, PA

---

Facial morphology is characterized by high heritability and rapid recent evolution in modern humans. The fossil record shows that modern humans are characterized by changes in facial size and shape compared to archaic humans. The availability of Neanderthal and Denisovan genomes, along with recent genome-wide association studies (GWAS) of modern human facial variation, provides data to investigate the genetics underlying recent evolution in facial morphology. We examined two sets of evolutionary markers: (1) Modern human-derived (MD) sites, defined as SNPs for which the derived allele is present at high frequency (>90%) in modern humans, and the ancestral allele was shared by chimpanzee and archaic humans; and (2) Archaic human-derived (AD) sites, defined as SNPs for which the ancestral allele is present at high frequency (>90%) in modern humans and is shared with chimpanzee, with the derived allele present in archaic humans. We contrasted the effects of 48,589 MD and 28,929 AD sites in a recent GWAS of 24 quantitative facial soft-tissue measures derived from the 3D surface images of 3118 individuals of European ancestry. All facial traits were adjusted for age, sex, height, weight, principal components of ancestry, and facial size (calculated as the geometric mean of the linear distance measures). Differences in the effects of AD and MD sites were most pronounced for Lower Vermilion height. Among AD and MD sites showing nominal association with facial traits (i.e.,  $p < 0.05$  in GWAS), the median effect sizes were in the opposite directions, with 63% of AD alleles having an elongating effect on lower vermilion height, and 63% of MD alleles having a shortening effect (Wilcoxon test  $p$  value =  $3E-144$ ). When a stricter significance threshold was applied (i.e.,  $p < 5E-05$  in GWAS), the corresponding percentages were 100% and 97%, respectively. Similar results were observed for lower lip height, upper vermilion height and nasal width. In contrast, features related to facial height showed AD vs. MD effects in the opposite directions, with the majority of MD alleles having an elongating effect, especially for nasal height and upper facial height. Our results suggest that the features around oral region decreased in the modern human lineage, whereas the vertical height of the midface increased. These findings are based on soft tissue measurements in modern humans, and have yielded important new insights into the genetic evolution of facial morphology.

## PgmNr 2363: Multiple waves of peopling northeast Europe: An ancient DNA study through Mesolithic to the Middle Ages.

### Authors:

M. Metspalu <sup>1</sup>; L. Saag <sup>1,2</sup>; K. Tambets <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Genomics, University of Tartu, Estonia; 2) Institute of Cell and Molecular Biology, University of Tartu, Estonia

---

aDNA studies have resketched the peopling story of Europe highlighting the major role of three distinct mass immigrations in the Mesolithic, Neolithic and Early Bronze Age periods. Here we look at these processes in Estonia representing northeast Europe. One key difference of this region from the rest of Europe is high (30-40%) frequency of chr Y haplogroup N3a, which is otherwise spread in Siberia. Genome wide the Siberian component makes up only a few percent of the genepool of Northeast Europeans. We sample each major cultural transition in the region and follow the accompanying demographic changes. The earliest Mesolithic samples cluster together with the Western hunter-gatherers of Europe (WHG). The arrival of Comb Ceramics Culture (CCC) is coupled with Eastern hunter-gatherer (EHG) genetic ancestry. The arrival of Corded Ware Culture (CWC) in Late Neolithic/Early Bronze Age, is part of the massive immigration to Europe from the Steppes of contemporary Russia, and installs a new genepool.

We then turned to study the genetic ancestry of individuals (N=56) from more recent but as yet genetically unstudied cultural traditions in Estonia: i) the Late Bronze Age (EstBA) (1200-400 BC), ii) the Pre-Roman Iron Age (EstIA) (800/500 BC-50 AD), and iii) Middle Ages samples from Estonia (EstMA) (1200-1600 AD). Compared to the preceding CWC people the genetic ancestry of the EstBA samples shifts back toward that of WHG. While it is the EstIA where we see minute fraction of the Siberian ancestry component for the first time. Moreover, the respective chr Y variant - N3a - is also detected in EstIA samples and not in bearers of the older cultures. This eastern/Siberian ancestry reached the coasts of the Baltic Sea no later than the mid-first millennium BC; i.e. in the same time window as the diversification of west Uralic/Finnic languages. Furthermore, phenotypic characteristics often associated with modern Northern Europeans like light eyes, hair and skin as well as lactose tolerance can be traced back to the Bronze Age in the Eastern Baltic.

During the Middle Ages the Estonian population is socially structured into subpopulations of landlords of likely mostly German origin and the common people. In the pilot phase of the study we have confirmed, that burials from high class cemetery resemble genetically contemporary Germans, while rural and urban cemeteries of common people yield aDNA samples genetically reminiscent of the current Estonian population.



## PgmNr 2364: Evolutionary modeling of the differential contribution of Neanderthal ancestry to complex traits provides insights into selective forces that shape trait variation.

### Authors:

C.R. Robles <sup>1</sup>; S. Sankararaman <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, UCLA, Los Angeles, CA; 2) Department of Computer Science, UCLA, Los Angeles, CA

---

We recently developed a methodology to assess whether Neanderthal ancestry is over- or under-represented in the genetic component of complex phenotypes compared to random genetic variation. Based on 500,000 individuals from the UK Biobank, we found the estimated contribution of Neanderthal alleles to phenotypic variation (NIM heritability) is significantly depleted in the great majority of the phenotypes. This is consistent with the observation that in general, natural selection has acted to remove Neanderthal alleles since introgression. On the other hand, we have found that Neanderthal alleles were significantly over-represented in their contribution to a handful of traits.

To understand the evolutionary models that could explain these observations, we performed forward-in-time population genetic simulations to model the evolution of Neanderthal and non-Neanderthal alleles according to a demographic model relating modern humans and Neanderthals. We chose parameters used in a previous study (Petr PNAS 2019) analyzing the fitness cost of Neanderthal introgression. Specifically, an ancestral population of size 10,000 diploid individuals splits into a human population and a Neanderthal population, each one evolves separately before a single pulse of Neanderthal admixture followed by subsequent random mating. Under this demography, we modeled evolution of phenotypes subject to different forces including directional, stabilizing, and disruptive selection. We estimated a *NIM heritability Z-score*, a measure of whether NIM heritability deviates significantly from the background alleles. We found under most models of selection, the NIM heritability Z-score is near zero or negative, indicating NIM heritability is neutral or depleted. Interestingly, we were able to recreate a positive NIM heritability Z-score, indicating an elevated Neanderthal contribution to heritability in two separate models of stabilizing and directional selection. In the stabilizing selection model, the optimal value of the trait is decreased in the human branch during the split between humans and Neanderthals leading to a positive NIM heritability Z-score. We also observe a positive NIM heritability Z-score in a directional selection model in which the parameter that couples SNP effect size and fitness is reduced after introgression. This observation highlights possible mechanisms for how complex traits evolved in human history by examining the genetic contribution of Neanderthal ancestry.

## PgmNr 2365: Lactase persistence prevalence from genetic risk allele frequency.

### Authors:

B.E. Graham <sup>1</sup>; S.M. Williams <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Systems Biology and Bioinformatics, Case Western Reserve University, Cleveland Heights, Ohio.; 2) Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland Heights, Ohio.

---

Under simple architecture (i.e. additivity), a Genetic Risk Score (GRS) using frequencies of phenotype associating variants, should predict relative trait prevalence among populations. This should be true for monogenetic traits, of which our understanding is relatively comprehensive with respect to risk loci/alleles. As a proof of principle, we used lactase persistence (LP), the continued production of the lactase enzyme into adulthood, to test whether a GRS predicted population prevalence. LP is a monogenic trait caused by mutations affecting the expression of *LCT*, the gene encoding lactase, and is reasonably well understood in some populations. The Lactase enzyme is produced by our bodies to breakdown lactose, the sugar found in milk. Lactase production usually decreases after weaning, in some cases leading to an intolerance of lactose. To determine the relationship, if any, between population prevalence of LP and the GRS of its associated alleles, we used a polygenic weighted additive risk score based on the allele frequency distributions of trait associated alleles. The GRS was based on 11 variants with known population allele frequencies. In Europe, two alleles upstream of the *LCT* gene, -13910\*T (rs4988235) and -22018\*G (rs182549), confer LP. In non-Europe populations, other alleles may be responsible for LP, where it exists. We accessed risk allele frequency from the 1000 Genomes project and genomAD through the Ensembl browser. The estimated prevalences of LP were taken from literature. We found a strong relationship between the GRS and the population prevalence of LP world-wide ( $r^2 > 0.6$ ), and especially among European populations ( $r^2 > 0.9$ ). This trend held for both South Asian and Amerindian populations, though to a lesser extent ( $r^2 \sim 0.5$  and  $0.2$ , respectively). However, in East Asian and African populations, the GRS failed to predict the relationship between the known LP alleles and the population prevalence. This indicates that, while we have identified most variants that confer LP in European populations, and to a lesser extent Amerindian and South Asian populations, explained heritability for both African and East Asian populations of Lactase Persistence is effectively missing. Therefore, the variants responsible for LP in African and East Asian populations have yet to be discovered or there is an unknown environmental factor. Using this GRS comparative approach can inform us of the completeness of the genetic bases for human traits.

## PgmNr 2366: Differential complex trait architecture across humans: Epistasis identified in non-European populations at multiple genomic scales.

### Authors:

M.C. Turchin <sup>1,2</sup>; I. Tung <sup>1,3</sup>; L. Crawford <sup>1,4,5</sup>; S. Ramachandran <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Computational Molecular Biology, Brown University, Providence, RI; 2) Department of Ecology and Evolutionary Biology, Brown University, Providence, RI; 3) Department of Computer Science, Brown University, Providence, RI; 4) Department of Biostatistics, Brown University, Providence, RI; 5) Center for Statistical Science, Brown University, Providence, RI

---

Genome-wide association (GWA) studies have identified thousands of significant genetic associations in humans across a number of complex traits. However, the vast majority of these studies use datasets of predominantly European ancestry (Popejoy & Fullerton 2016). It has generally been thought that complex trait genetic architecture should be transferable across populations of different ancestries, but recent work has shown a number of differences in trait architecture across human ancestries, including heterogeneity in both the identified causal variants and estimated effect sizes (Martin et al. 2017, Wojcik et al. 2017). Here, we report further evidence that complex trait genetic architecture is fundamentally different among human ancestries by jointly leveraging pathway and epistasis analysis.

Under the assumption that a given complex trait may have differential polygenic architectures across human ancestries, we hypothesize that human populations may also be enriched for differences in epistatic effects. However, since polygenic traits tend to have smaller GWA effect sizes, combining variants via pathway analysis may allow us to better reveal these signals. To accomplish this, we extend the concept of identifying marginal epistasis, moving from testing single variants (Crawford et al. 2017) to testing groups of variants for nonlinear association with a trait of interest.

We apply our new method to multiple ancestries present in the UK Biobank (Sudlow et al. 2015) and explore multiple pathway-related interaction models. Using morphometric traits we find evidence for genome-wide epistasis in African and other non-European populations. We also find evidence that these trends exist on the SNP and gene levels as well. Results also indicate this may be due to increased heterozygosity in non-European populations. This suggests that non-European populations may be well-suited for identifying non-additive effects in human complex trait architecture; this also suggests further evidence that European populations -- predominantly used for epistasis studies -- may indeed be limited and inaccurate proxies for all human ancestries in complex trait research.

## PgmNr 2367: Putting RFMix and ADMIXTURE to the test in a complex admixed population.

**Authors:**

C. Uren; E.G. Hoal; M. Möller

[View Session](#) [Add to Schedule](#)

**Affiliation:** DST-NRF Centre of Excellence for Biomedical TB Research, South African Medical Research Council Centre for TB Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, SA

---

Global and local ancestry inference in admixed human populations can be performed using computational tools implementing distinct algorithms, such as RFMix and ADMIXTURE. The accuracy of these tools has been tested on populations with relatively straightforward admixture histories (e.g. African Americans) but little is known about how well they perform in more complex admixture scenarios. Using simulations, we show that RFMix outperforms ADMIXTURE in determining global ancestry proportions in a complex 5-way admixed population. In addition, RFMix correctly assigns local ancestry with an accuracy of 89%. The increase in reported local ancestry inference accuracy in this population (as compared to previous studies) can largely be attributed to the recent availability of large-scale genotyping data for more representative reference populations. The ability of RFMix to determine global and local ancestry to a high degree of accuracy, allows for more reliable population structure analysis, scans for natural selection, admixture mapping and case-control association studies. This study highlights the utility of the extension of computational tools to become more relevant to genetically structured populations, as seen with RFMix. This is particularly noteworthy as modern-day societies are becoming increasingly genetically complex and some genetic tools are therefore less appropriate. We therefore suggest that RFMix be used for both global and local ancestry estimation in complex admixture scenarios.

## PgmNr 2368: An overview of genetic diversity and signatures of natural selection observed in HLA-G and its ligands KIR2DL4, LILRB1/ILT2, and LILRB2/ILT4, in an admixed Brazilian sample.

### Authors:

E. Castelli <sup>1</sup>; M.L.G. Oliveira <sup>2</sup>; E. Weiss <sup>3</sup>; A.S. Souza <sup>3</sup>; T.H.A. Lima <sup>3</sup>; H.S. Andrade <sup>3</sup>; M.R.S. Passos <sup>1</sup>; N.S.B. Silva <sup>1</sup>; C.F.B. Castro <sup>4</sup>; L. Marcorin <sup>2</sup>; G. Debortoli <sup>2</sup>; A.L.E. Pereira <sup>2</sup>; N.C.A. Fracasso <sup>2</sup>; G.V. Silva <sup>5</sup>; A. Sabbagh <sup>6</sup>; C.T. Mendes-Junior <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) São Paulo State University (UNESP), School of Medicine, Botucatu, State of São Paulo, Brazil; 2) Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, 14049-900, Ribeirão Preto-SP, Brazil; 3) São Paulo State University (UNESP), Bioscience Institute, Botucatu, State of São Paulo, Brazil; 4) Centro Universitário Sudoeste Paulista – UniFSP, Brazil; 5) Departamento de Química, Laboratório de Pesquisas Forenses e Genômicas, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 14040-901, Ribeirão Preto-SP, Brazil; 6) UMR 216 IRD MERIT, Faculté de Pharmacie de Paris, Université Paris Descartes, Paris, France

*LILRB1/ILT2* and *LILRB2/ILT4* encode important Natural Killer (NK) and T cell inhibitory receptors, while *KIR2DL4* encodes an activator NK receptor. These receptors are targeted by HLA-G, an immunomodulatory molecule expressed mainly on the trophoblast. Polymorphisms at both ligand and receptor might influence their binding and expression. We assessed *HLA-G*, *ILT2*, and *ILT4* exonic diversity in 410 Brazilian samples, and *KIR2DL4* in a subset of 126 samples. Ancestry composition of the sample comprehended 71% European, 21% African, and 8% Amerindian/Asian. We used Haloplex (Agilent) for *HLA-G*, *ILT2*, and *ILT4*, and Nextera XT (Illumina) for *KIR2DL4*. Libraries were sequenced using the Miseq system in 2x300 runs. The alignment was performed with an adapted version of hla-mapper that considers known HLA, KIR, and LILR sequences. Genotyping and haplotyping were performed combining GATK HaplotypeCaller, ReadBackedPhasing, and Beagle 4. Linkage Disequilibrium (LD) was evaluated by means of Haploview and Genepop. The dn/ds ratio and Tajima's D were calculated using Mega7. DNA sequences were translated into proteins using emboss. HLA-G presented 5 frequent proteins, and the most frequent one was G\*01:01 (65%). We have found 5 frequent KIR2DL4 proteins. KIR2DL4\*001 and \*008 presented a summed frequency of 60%, and they differ by two amino acid exchanges and the latter is truncated due to a premature stop codon. ILT2 presented 12 frequent proteins, but one of them was very common (43%). ILT4 presented 10 frequent proteins, three of them with a summed frequency of 75%. According to the human genome draft hg38, there are 350 Kb between *LILRB2* and *LILRB1*, and 165Kb between *LILRB1* and *KIR2DL4*. Despite that, we detected no LD among these genes, with different combinations of *ILT2*, *ILT4*, and *KIR2DL4* receptors encoded per chromosome. In fact, there are many recombination hotspots between these genes (HapMap project). It is not clear whether receptor polymorphisms influence HLA-G binding. Nevertheless, our results indicate that *HLA-G* and *ILT2* are under purifying selection. Moreover, the number of individuals presenting some combinations, such as G\*01:03 and *ILT2* (with Lys at position 625), is higher than expected. Whether this is due to chance or not needs to be further investigated. In conclusion, here we present the genetic diversity of *HLA-G* and its ligands. *HLA-G* and

*ILT2* might be under purifying selection, and some HLA-G/*ILT2* combinations are more frequent than expected.

## **PgmNr 2369: Wavelet analysis of recombination rates: Population derived linkage maps have limited value when applied to different human populations.**

### **Authors:**

C. Horscroft <sup>1,2</sup>; R.J. Pengelly <sup>1,2</sup>; A. Collins <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Human Genetics and Genomics, University of Southampton, Duthie Building (MP808), Tremona Road, Southampton, Hampshire, United Kingdom, SO16 6YD; 2) Institute for Life Sciences, University of Southampton, Life Sciences Building (85), Highfield, Southampton, United Kingdom, SO17 1BJ

---

Recombination is the process in which pairs of homologous chromosomes cross-over during meiosis. These cross-over events are beneficial as they stop negative mutations building up along one haplotype by breaking down linkage between alleles. Recombination rates are known to be variable across the genome, with events clustering in hotspots. In the study of evolution and natural selection, it is important to understand the underlying recombination map for the population being analysed. The aim of this work is to calculate recombination rates in African and European populations and use wavelet analysis to investigate the differences on multiple scales.

Recombination rates were estimated using the widely implemented LDhat software for four human genomic datasets: two European and two African. The resulting recombination maps were then analysed using wavelets. Wavelet analysis allowed the recombination maps to be analysed at multiple scales simultaneously, rather than using a fixed window size, from the fine scale (2 Kb) through to the wide scale (16 Mb). This meant it could be ascertained whether the variance in the original scale was contained in short-term trends, or in longer ranging changes. Wavelet coherence analysis also allowed the comparison of the correlations of recombination rate changes between pairs of datasets, both globally and locally along the region.

The highest proportion of variance in the recombination rates was found at around the 16-64 Kb range across all the datasets, accounting for around a third of the total variance. The correlations between the recombination rate changes increased with scale over all pairs of datasets. The differences between African and European recombination rate changes were greatest at the fine scale. The recombination rate changes were found to be non-uniform across the region analysed, regardless of scale.

This work provides evidence that a recombination map built from one human population will not be representative of all other populations, especially at the fine scale. This has implications for methods designed to find evidence of natural selection in the genome that rely on linkage disequilibrium calculations.

## PgmNr 2370: Measuring recessive selection using summary statistics of the site frequency spectrum.

### Authors:

D.M. Jordan <sup>1</sup>; D.J. Balick <sup>1,2</sup>; S.R. Sunyaev <sup>2</sup>; R. Do <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Charles Bronfman Institute for Personalized medicine, Icahn School of Medicine at Mt Sinai, New York, NY; 2) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

---

With the growth of large population sequencing datasets, measures of population constraint are becoming increasingly popular tools to predict gene essentiality or morbidity. These measures include, most notably, the constraint or loss of function intolerance score (pLI) from the Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomAD) datasets. A recent perspective by Fuller et al. noted that many of these measures perform poorly as estimators of the strength of selection, and that they are particularly insensitive to the difference between strong recessive selection and dominant or additive selection at a much weaker strength, in some cases even approaching neutrality. We have developed a population genetic method that is able to distinguish between recessive and additive selection in both simulated and observed genes. While previous likelihood-based methods that consider the entire site frequency spectrum (SFS) have had success in simulated models, they have been difficult to apply to actual large-scale human population sequencing data because they are highly sensitive to noise and demographic history. In contrast, our method uses summary statistics of the SFS that are chosen to be relatively robust to these factors: a measure of heterozygous burden and a logarithmic transformation of the SFS. Within the two-dimensional space defined by these summary statistics, we compare the density of known recessive genes or simulations to the density of non-recessive genes or simulations to identify regions that are typical of genes experiencing recessive selection. In addition to performing well on simulations, when this method is applied to a dataset of over 30,000 non-Finnish European human exomes derived from ExAC, we are able to identify genes implicated in autosomal recessive disease and genes suspected of being under recessive selection based on studies of consanguineous populations. Applying this method broadly to the entire exome produces potentially important insights into the biology of recessive selection. This method also has potential applications for the discovery of new recessive disease genes and clinical genetic diagnostic pipelines.



## PgmNr 2371: Gene expression in the testis shapes germline mutation signatures and modulates gene evolution rates.

### Authors:

B. Xia <sup>1</sup>; Y. Yan <sup>1</sup>; M. Baron <sup>1</sup>; F. Wagner <sup>1</sup>; D. Barkley <sup>1</sup>; M. Chiodin <sup>1</sup>; S. Kim <sup>5</sup>; D. Keefe <sup>3</sup>; J. Alukal <sup>3</sup>; J. Boeke <sup>2,4</sup>; I. Yanai <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Computational Medicine, NYU Langone Health, New York, NY; 2) Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY; 3) Department of Obstetrics and Gynecology, NYU Langone Health, New York, NY; 4) Institute for Systems Genetics, NYU Langone Health, New York, NY; 5) Department of Pathology, NYU Langone Health, New York, NY

---

The testis expresses the largest number of genes of any mammalian organ, a finding that has long puzzled molecular biologists. Analyzing our single-cell transcriptomic maps of human and mouse spermatogenesis, we provide evidence that this widespread transcription serves to maintain DNA sequence integrity in the male germline by correcting DNA damage through “transcriptional scanning”. Supporting this model, we find that genes expressed during spermatogenesis display lower mutation rates on the transcribed strand and have low diversity in the population. Moreover, this effect is fine-tuned by the level of gene expression during spermatogenesis. Analyzing the germline mutation datasets, we further identified novel rules governing the transcription-coupled repair of germline DNA damage. The unexpressed genes, which in our model do not benefit from transcriptional scanning, diverge faster over evolutionary time-scales and are enriched for sensory and immune-defense functions. Collectively, we propose that transcriptional scanning modulates germline mutation rates in a gene-specific manner, maintaining DNA sequence integrity for the bulk of genes but allowing for fast evolution in a specific subset.

## PgmNr 2372: Great ape-specific ATF4 retrocopies may act to regulate wild type ATF4 activity and viral infections.

### Authors:

H.M. Dalton; N.C. Elde; C.Y. Chow

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, United States of America

---

Multiple stress signals trigger the integrated stress response (ISR) such as endoplasmic reticulum stress and viral infection. Dysregulation of the ISR is implicated in diseases such as cancer and diabetes. The ISR activates kinases that phosphorylate eIF2alpha, causing strong inhibition of protein synthesis machinery while repair occurs. Despite this, certain genes contain specialized upstream Open Reading Frame (uORF) regulatory sequences in their 5'UTR region that allow, or even upregulate, translation during the ISR. One of the most upregulated uORF-containing transcripts is ATF4, a highly conserved transcription factor that activates pro-survival or apoptotic genes.

We have found conservation of multiple ATF4 retrocopies – labeled in humans as pseudogenes *ATF4P1-P4*. *hATF4P1* and *P2* are short, truncated retrocopies that date to the common ancestor between apes and old world monkeys, while *hATF4P3* and *P4* are ape-specific. *hATF4P3* is a full length copy that maintains all domains and uORF regulatory regions with 95% amino acid (AA) identity with the parent *ATF4* gene. The 5% AA that are diverged from the parent gene are primarily in the N-terminal side, with no differences in the C-terminal DNA binding domain. Perhaps the most interesting is *hATF4P4* which has a truncation event removing its DNA binding domain, yet maintains 90% AA identity with the parent gene in the truncated half. *hATF4P4* is well conserved in humans, chimps, bonobos, and gorillas (Homininae) and ancestral to their most common ancestor. Strikingly, independent retrocopies with similar truncation events have occurred in orangutan, gibbon, and gorilla genomes (an extra gorilla copy). In addition, all Homininae have lost their uORF regulatory sequences, while the independent truncation retrocopies maintain them.

Strong conservation and independent events in different lineages suggests that each of these retrocopies has the potential to be functional. For example, all truncated *hATF4P4*-like retrocopies maintain the transcription-enhancing P300 protein binding domain. We hypothesize that *hATF4P4*-like retrocopies may regulate the parent gene by sequestering proteins like P300 away from ATF4. In addition, viruses, like HIV, can hijack important machinery during the ISR, suggesting an alternative hypothesis where *hATF4P4* may buffer ATF4 to inhibit viral infection while not affecting ATF4-driven transcription. We will present functional analyses that will distinguish between these possibilities.

## PgmNr 2373: Prioritizing natural selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders.

### Authors:

L. Deng<sup>1</sup>; C. Zhang<sup>1,2</sup>; K. Yuan<sup>1,2</sup>; Y. Gao<sup>1,2,3</sup>; Y. Pan<sup>1,2</sup>; X. Ge<sup>1,2</sup>; Y. He<sup>4</sup>; Y. Yuan<sup>1</sup>; Y. Lu<sup>1</sup>; X. Zhang<sup>1,2,3</sup>; H. Chen<sup>1,2</sup>; H. Lou<sup>1</sup>; X. Wang<sup>1,2</sup>; D. Lu<sup>1,2</sup>; J. Liu<sup>1,3</sup>; L. Tian<sup>1,2</sup>; Q. Feng<sup>1,2</sup>; A. Khan<sup>1</sup>; Y. Yang<sup>5</sup>; Z. Jin<sup>6</sup>; J. Yang<sup>6,7</sup>; F. Lu<sup>6</sup>; J. Qu<sup>6</sup>; L. Kang<sup>8</sup>; B. Su<sup>4,9</sup>; S. Xu<sup>1,2,3,9,10,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; 2) University of Chinese Academy of Sciences, Beijing 100049, China; 3) School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China; 4) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 5) State Key Laboratory of Genetic Engineering and Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai 200433, China; 6) The Eye Hospital, School of Ophthalmology & Optometry, Wenzhou Medical University, Wenzhou, 325027, China; 7) Institute for Molecular Bioscience, Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia; 8) Key Laboratory for Molecular Genetic Mechanisms and Intervention Research on High Altitude Disease of Tibet Autonomous Region, School of Medicine, Xizang Minzu University, Xianyang Shaanxi 712082, China; 9) Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China; 10) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China; 11) Human Phenome Institute, Fudan University, Shanghai 201203, China

---

Human genetic adaptation to high altitudes (>2500 m) has been extensively studied over the last few years, but few functional adaptive genetic variants have been identified, largely owing to the lack of deep-genome sequencing data available to previous studies. Here, we build a list of putative adaptive variants, including 63 missense, 7 loss-of-function, 1,298 evolutionarily conserved variants, and 509 expression quantitative traits loci. Notably, the top signal of selection is located in *TMEM247*, which is a transmembrane protein coding gene and harbors one high-frequency (76.3%) missense variant, rs116983452 (c.248C>T; p.Ala83Val), in Tibetan highlanders. The rs116983452-T is derived from archaic ancestry and carried by >94% of Tibetans but is absent or in low frequencies (< 3%) in non-Tibetan populations. This Tibetan-enriched allele is strongly and positively correlated with altitude, and is significantly associated with reduced hemoglobin concentration ( $p = 5.78 \times 10^{-5}$ ), red blood cell count ( $p = 5.72 \times 10^{-7}$ ), and hematocrit ( $p = 2.57 \times 10^{-6}$ ). These results were further validated in 1,160 replicate Tibetan samples collected from four different altitudes. In particular, *TMEM247*-rs116983452 showed greater effect size and better predicts the phenotypic outcome than any *EPAS1* variants in the association with several adaptive traits in Tibetans. In addition, modeling the interaction between *TMEM247*-rs116983452 and *EPAS1* variants showed weak but statistically significant epistatic effect. We argue that more than one variant may jointly deliver the fitness on the Plateau, and that the

adaptive evolution of the Tibetan people involves a large number of variants, of which interactions or epistatic effects might be existing.

## PgmNr 2374: Association of common genetic variants in human accelerated regions with craniofacial and subcortical morphology.

### Authors:

A.C. Driscoll <sup>1</sup>; M.K. Lee <sup>2</sup>; M.L. Marazita <sup>2,3</sup>; P. Claes <sup>4</sup>; S.M. Weinberg <sup>2,5</sup>; J.R. Shaffer <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA; 2) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA; 3) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 4) Department of Electrical Engineering & Department of Human Genetics, KU Leuven, Leuven, Belgium; 5) Department of Anthropology, University of Pittsburgh, Pittsburgh, PA

---

Human accelerated regions (HARs) are small genomic regions (100-200 base pairs) that, although highly conserved through primates, exhibit an unusually high number of mutations from the ancestral state in humans. This suggests that they may impact traits that are both functionally necessary and differentiate humans from other primates. For example, HARs have been associated with enhancers that regulate neural development. Additionally, gene ontology enrichment analyses suggest that HARs are connected to key developmental processes impacting craniofacial structure such as the regulation chondrocyte differentiation and cartilage development. We therefore determined whether common variants in HARs were associated with measures of craniofacial morphology and subcortical brain volumes. A list of 2,737 HARs compiled from previous publications by the authors of a 2016 study was used to obtain the genomic locations of HARs. Then we extracted results for 2,314 SNPs located in these HARs from recent genome-wide association studies of craniofacial morphology (N=2,329, 3D facial norms consortium) and 1809 SNPS from a recent genome-wide association study of subcortical brain volumes (N=8,064, ENIGMA consortium). We used the Bonferroni p-value threshold, adjusting for 2,314 tests for the craniofacial analysis and 1,809 tests for subcortical volumes, to declare statistical significance. We identified two SNPs within HARs that were significantly associated with craniofacial morphology, with effects concentrated in the nasal area. One of these was located within an intergenic region adjacent to *PAX3*, a gene which controls the embryonic development of bones comprising the face and has been associated with nasal morphology in several prior GWAS of human facial morphology. We observed no evidence that SNPs within HARs were associated with subcortical brain volumes. Finally, we found no evidence that HARs were enriched for SNPs associated with craniofacial morphology or any measure of subcortical volume. Future work should further explore the relationship between human accelerated regions and cortical brain structures, which are more directly relevant to human evolution.

## PgmNr 2375: Using a combination of GWAS methods and haplotype-based selection statistics to infer patterns of positive selection in Oceania.

### Authors:

S. Kaewert <sup>1</sup>; G. Wojcik <sup>2</sup>; E. Sorokin <sup>2</sup>; J. Homburger <sup>3</sup>; C. Quinto-Cortés <sup>4</sup>; A. Ioannidis <sup>5</sup>; J. Portillo <sup>4</sup>; J. Rodriguez <sup>4</sup>; K. Auckland <sup>6</sup>; A. Severson <sup>3</sup>; J. Blanc <sup>7</sup>; G. Belbin <sup>8</sup>; S. Oppenheimer <sup>6</sup>; T. Parks <sup>6</sup>; M. Phipps <sup>9</sup>; K. Robson <sup>6</sup>; E. Kenny <sup>8</sup>; A. Hill <sup>6</sup>; A. Mentzer <sup>6</sup>; C. Bustamante <sup>2</sup>; A. Moreno-Estrada <sup>4</sup>; C. Gignoux <sup>1,10</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Colorado Center for Personalized Medicine, University of Colorado, Aurora, CO; 2) Department of Biomedical Data Science, Stanford University, CA; 3) Department of Genetics, Stanford University, CA; 4) National Laboratory of Genomics for Biodiversity, Irapuato, Mexico; 5) Computational and Mathematical Engineering, Stanford University, CA; 6) University of Oxford, U.K.; 7) University of California Davis, CA; 8) Icahn School of Medicine at Mount Sinai, New York, NY; 9) Monash University, Malaysia; 10) Department of Biostatistics and Informatics, University of Colorado, CO

---

The history of the settlement of the Pacific was a complex demographic event involving multiple migrations and exposure to a wide variety of environmental pressures on new islands, including the unique disease dynamics present in isolated, high-density populations with intermittent contact. These challenges resulted in opportunities for local adaptation to new conditions. However, Oceania remains largely underrepresented in studies of genetic selection. Using data from the Oceanian Genome Variation Project (OGVP), involving 980 individuals from 88 populations measured on the MEGA platform at over 1.7M sites along the genome, we searched for evidence of positive selection using the machine learning model SWIFr. Our inputs into the model were three haplotype-based statistics (iHS, nSL, and xp-EHH) and the p-values from a mixed effects model GWAS analysis using population membership as the phenotype to quantitatively assess overall differentiation. Together, these account for local population structure and suggest genome-wide significant levels of potential selection in multiple regions, including in HLA-DRB1 and -DQA1 in West Polynesia. We also find significant enrichment in East Polynesia in top scoring genes involved in monosaccharide metabolism and response to xenobiotic stimulus. The results suggest local adaptation to infectious disease and diet.

## PgmNr 2376: Distinct patterns of genetic ancestry and adaptation in the islanders of Taiwan.

### Authors:

W.Y. Ko <sup>1</sup>; Y.H. Lo <sup>1</sup>; H.C. Cheng <sup>1</sup>; H.Y. Wang <sup>1,2</sup>; C.W. Peng <sup>1,2</sup>; C.Y. Chen <sup>3</sup>; C.N. Hsiung <sup>3</sup>; C.H. Chen <sup>3</sup>; H.W. Chu <sup>3</sup>; C.Y. Shen <sup>3</sup>; S.M. Chaw <sup>2</sup>; M. Lin <sup>4</sup>; J.H. Loo <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Life Sciences and Institute of Genome Sciences Taipei, Taiwan; 2) Biodiversity Research Center, Academia Sinica, Taipei, Taiwan.; 3) Institute of Biomedical Science, Academic Sinica, Taipei, Taiwan; 4) Mackay Memorial Hospital, Taipei, Taiwan.

---

Taiwan is an island and home of multiple ethnic groups including Taiwanese Han and sixteen officially recognized aboriginal tribes. However, little is known about the levels of genetic structure/admixture and patterns of evolutionary adaption among these populations. Here, we analyzed the whole-genome SNP genotyping data from 14,496 individuals of Taiwanese people. The PCA first showed a clear discrepancy between the Taitung (dwelling in the Eastern Taiwan) and Taiwanese Han people. The ADMIXTURE analysis further revealed that the genomes of Taiwanese Han contain three major components with distinct geographic distributions while the Taitung people contain one major genetic component that contributes to many of the Austronesian populations in a great proportion. To scan for the genomic signatures of recent positive selection, we calculated integrated haplotype score (*iHS*) and subsequently identified 26 and 12 candidate loci in the Taiwanese Han and Taitung populations, respectively. Surprisingly, none of these candidate loci was shared between them except for the *HLA* genes. We further constructed a fine-scale LD map and calculated the iSAFE (Integrated Selection of Allele Favored by Evolution) and CADD scores (Combined Annotation Dependent Depletion) to pinpoint the causal variants by analyzing a subset of individuals (772) whose whole-genome sequencing data are available. As a result, we discovered some selection-targeted variants with high deleterious scores (e.g., *NFKBIA*, *WNT7A* and *NEO1* genes that play a role in either inflammatory or tumor suppressor signalling pathway). Together, our findings reveal important insights into the evolutionary history of Taiwanese people and provide an alternative approach to identify genetic basis that underlie disease susceptibility.

## PgmNr 2377: Ongoing purifying and overdominant selection in the human genome.

### Authors:

Wei <sup>1</sup>; Z. Pan <sup>1,3</sup>; R. Nielsen <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UC Berkeley, Berkeley, CA.; 2) University of Copenhagen; 3) Sun Yat-Sen University

---

The release of 500,000 genomes from the UK Biobank (UKB) provides unprecedented opportunities for studying ongoing selection using deviations from Hardy-Weinberg Equilibrium (HWE). Compared to the HWE expectation, we observe increased heterozygosity in British ancestry individuals in the UKB cohort genome-wide. This observation could potentially be caused by genotyping errors. However, we find that nonsynonymous SNPs with low Minor Allele Frequency (MAF  $< 5\%$ ) are enriched for excess heterozygosity compared to other SNPs of the same MAF. This observation cannot be explained by genotyping errors, but is likely due to purifying selection against recessive deleterious alleles. Perhaps surprisingly, low MAF archaic SNPs from Neanderthal admixture show depletion of excess heterozygosity, a pattern that can be replicated in simulations incorporating selection and addition, by mutation, of new deleterious alleles, for many generations after the time of admixture. We also find that high MAF (30-50%) nonsynonymous SNPs are enriched for excess heterozygosity compared to other SNPs of the same MAF. These SNPs are enriched for genes previously identified to be under balancing selection. They also show evidence of overdominance from decreased all-cause mortality in heterozygous individuals. Population genetic simulations under realistic parameter settings can recapitulate these observations. Our study demonstrates that analyzing patterns of deviations from HWE can be a powerful way to detect selection in large cohorts and that ongoing selection in humans is common.



## PgmNr 2378: Selection of 28 AIMs analysed *in silico* and tested in Native and Mestizo Mexican samples to infer the Mexican regional ancestry.

### Authors:

R. Montero Delgado <sup>1</sup>; G. Noris <sup>2</sup>; C. Santana <sup>2</sup>; R. Gomez <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Toxicology, CINVESTAV-IPN, Mexico City, Mexico City, Mexico; 2) Biología Molecular Diagnostica S.A. de C.V., Queretaro, Qro, Mexico.

---

Mexican population emerged from the admixture among Native Americans, Europeans and Africans 500 years ago. Their youth, joint to the demographic events experienced by their founder populations, and the skewed genetic mixture between Native women, and European, and African men, have yielded an intricate background ancestral. Ancestry Informative Markers (AIMs) are single-nucleotide polymorphisms used to infer populations' geographic origins. These markers have been broadly studied in worldwide populations. Nonetheless, their use in Latino populations lacks the systematic basis for its selection, validation, and minimal number to infer the ancestry of populations with complex genetic architecture. The multi-ethnicity of the Mexican population provides a unique opportunity to explore the utility of these markers, both in Mestizo as in Native populations, which data are poorly represented in HapMap. In the present study, we selected and tested the effectiveness of a reduced panel of 28 AIMs to be used in the Mexican population. These markers were selected after an exhaustive *in silico* data analyses from 200 AIMs previously reported. Different genetic parameters (i.e, genetic distance,  $F_{ST}$  among others) were used to evaluate its effectiveness. Genetic parameters such as genetic distance ( $F_{ST}$ ) was used. *In silico* effectiveness were explored using 2,504 genotypes from the 1000 genomes project with Genetix 4.05, Arlequin 3.5, and BAPS 6.0. These 28 markers were also tested in 150 samples from 90 Native Americans (Uto-Aztecan, Oto-Manguean, and Mayan linguistic families), and 60 Mexican Mestizos from Central Valley of Mexico. Our study provides additional evidence to unravelling, efficiently, the complex ancestral background of the Mexican population to low cost.

## PgmNr 2379: Ancestral origins and natural selection of Highland Tajik people in Xinjiang.

### Authors:

J. Liu<sup>1,2,3</sup>; Q. Feng<sup>1,2</sup>; X. Zhang<sup>1,2,3</sup>; Y. Lu<sup>1</sup>; C. Zhang<sup>1,2</sup>; Y. Gao<sup>1,2,3</sup>; X. Wang<sup>1,2</sup>; D. Mamatyusupu<sup>4</sup>; S. Xu<sup>1,2,3,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) CAS-MPG Partner Institute for Computational Biology, Shanghai, Shanghai, China; 2) University of Chinese Academy of Sciences, Beijing, China; 3) School of Life Science and Technology, ShanghaiTech University, Shanghai, China; 4) College of the Life Sciences and Technology, Xinjiang University, Urumqi, China; 5) Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China; 6) Collaborative Innovation Center of Genetics and Development, Shanghai, China

---

Recent studies revealed that Tibetan, Andeans and Ethiopian populations used different strategies and different genes to adapt high-altitude environments, although little is known about functional variants and genetic mechanisms. Studying those under-investigated highland groups might provide new insights into the genetic basis of high altitude adaptation. Here, we conducted a whole-genome deep-sequencing study of the Xinjiang Tajik (XJT) people who are residing on the Pamir Plateau with an average altitude > 3,000 meters. The majority of the ancestry of XJT was derived from west Eurasian (30.0% - 49.9%) and South Asian (42.6% - 57.3%) populations, with some minor (3.62% - 17.5%) ancestry derived from East Asian and Siberian populations. We identified a genomic region encompassing *CAPN3* and *GANC* showing signature of natural selection in XJT. The selected haplotypes[S.X.1] are enriched in XJT (carried by 26% of Tajik people) and likely originated from an unknown archaic group. Our data indicated that genetic admixture facilitated high-altitude adaption of XJK. A few candidate genes associated with XJT's adaptation were also reported previously in Ethiopian and Andeans, suggesting that some common genetic basis of high-altitude adaptation might be partly shared by indigenous groups from different highlands.

## PgmNr 2380: A comparative analysis of DNA replication timing in humans and chimpanzees.

### Authors:

A.N. Bracci <sup>1</sup>; Y. Gilad <sup>2</sup>; A. Koren <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY.; 2) Department of Human Genetics, University of Chicago, Chicago, IL.

---

DNA replication is accomplished by firing of replication origins across the genome at different times during S phase, resulting in a spatial and temporal pattern known as DNA replication timing. Replication timing shapes mutation rates across the genome, thus impacting sequence evolution and genetic variation. Replication timing variation between humans and other great apes can thus affect human evolution and underlie human-specific traits. To test this, we sequenced the genomes of 90 chimpanzee lymphoblastoid cell lines (LCLs) and seven induced pluripotent cell lines (iPSCs) along with similar numbers of human samples, and inferred replication timing by analyzing DNA copy number fluctuations along chromosomes. While replication timing was largely conserved between species and the greatest replication timing variation was related to cell type, we also identified numerous species-specific differences within the same cell type. Overall, human and chimpanzee replication timing varied at 311 genomic regions covering on average 523 Kb and cumulatively spanning 163 Mb (5.4% of the genome). The majority (52%) of this variation impacted replication origins, manifesting as gain (n=49) or loss (n=45) of origins in humans relative to chimpanzees, or changes in origin firing time (n=68). Of the latter, 29 regions showed advanced replication in humans compared to chimpanzees, while the remaining 39 were delayed in humans. Human-chimpanzee replication timing variants spanned 1,060 genes, including some with prior indications of human-specific evolution. For example, *CASP12*, one of the first identified cases of adaptive gene loss in humans, replicated earlier in humans compared to chimpanzees and was associated with an earlier firing human replication origin. Taken together, replication timing has been subject to dynamic evolution since the divergence of humans and chimpanzees, resulting in extensive variation between the species with potentially important implications to human evolution and phenotypes.

## PgmNr 2381: The origin of clustered mutations in human germline.

### Authors:

V. Seplyarskiy <sup>1,2</sup>; A. Sepliarskaia <sup>3</sup>; R. Soldatov <sup>2</sup>; R. McGinty <sup>1,2</sup>; J. Goldman <sup>4</sup>; C. Gilissen <sup>5</sup>; W.S.W Wong <sup>6</sup>; S. Sunyaev <sup>1,2</sup>; TOPMed

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA; 3) University of Amsterdam, Amsterdam, The Netherlands; 4) Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen, the Netherlands.; 5) Department of Human Genetics, Donders Centre for Neuroscience, Radboud University Medical Center, Nijmegen, the Netherlands; 6) Inova Translational Medicine Institute (ITMI), Inova Health Systems, Falls Church, VA, USA.

---

Most human mutations lead to independent single nucleotide variants (SNV). A minority, however, result in complex events such as clustered mutations. Mutation clusters provide an insight into underlying biochemical mechanisms, because a cluster contains more statistical information than a SNV.

Previously, clusters have been characterized by distances between individual mutations within clusters and by spectra of these mutations. Here, we have developed a statistical model of clustered mutations formulated as a mixture of Hidden Markov Models (HMM). Our model assumes existence of several mechanisms producing clusters, each represented by a separate HMM. We designed and implemented an algorithm to fit parameters of the HMM mixture.

We applied this method to more than 1,500,000 mutation clusters inferred from rare variants in the TOPMed dataset. This analysis identified three mutation processes generating clusters. The first process is characterized by a single A>T, A>G or G>T mutation at the 3' end of the cluster and any number of subsequent C>G mutations at the 5' end of the cluster. The second process displays nearly the same spectrum at the 3' end and any number of C>T mutations at the 5' end. The third process is dominated by A>G mutations in all positions.

The complex structure of the first and second processes lends itself to a mechanistic hypothesis. The directionality of these clusters suggests a two-step mechanism. We hypothesize that the first (3') mutation is likely caused by the nucleotide mis-incorporation opposing a DNA lesion. The following (5') mutation(s) are introduced by translesion polymerases recruited due to the replication obstacle at the original lesion. This model is supported by the analysis of clusters crossing the boundaries of transcribed regions.

Using trio sequencing data, we show that the first process is much more common among clusters of maternal origin, while the second process is more prevalent among paternal clusters. Separately, we demonstrate that the localized bursts of C>G mutations in oocytes are driven by the first mutational process.

## PgmNr 2382: A theoretical framework for interpreting the effect of geographic structure on GWAS.

**Authors:**

C. Porras; D.P. Rice; J. Novembre

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Human Genetics, University of Chicago, Chicago, IL.

---

Genome-wide association studies (GWAS) have shown that much of the variation in disease risk is due to rare deleterious alleles. This is likely due to the action of natural selection eliminating disease-risk alleles from the population. We expect these alleles to be geographically localized because natural selection removes them before they can spread far beyond their original location. However, the geographic distributions of these alleles have not been characterized in detail. We aim to understand how natural selection, geographic population structure, and geographic sampling bias interact to determine the inferred local and global genetic architecture of a trait.

It is inevitable that association studies will feature significant geographic bias. For example, the UK BioBank contains samples of people who have migrated to the United Kingdom from around the world, but still represents a small fraction of global diversity that is highly biased toward a relatively small geographic region. How can we use GWAS results from geographically localized studies like those from the UK BioBank to better predict disease risk in other geographic regions?

Here we develop theoretical models for the geographic spread of rare deleterious alleles. We focus on calculating the ascertained genetic architecture of a trait. The essential output of our population genetics analysis is the expected allele frequency spectrum as a function of: the geographic sampling scheme, the selection coefficient for an allele, and the population's geographic structure. Our theoretical results highlight the dependence of the inferred genetic architecture of a trait on geography and the evolutionary process, with implications for the interpretation of geographically localized GWAS cohorts.

## PgmNr 2383: Impact of selection on linkage disequilibrium in humans and its consequences for genetic epidemiological methods.

### Authors:

C.B. Rauchman <sup>1,\*</sup>; A.J. Stern <sup>2,\*,+</sup>; L.H. Uricchio <sup>1</sup>; R. Nielsen <sup>1,3,+</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Integrative Biology, UC Berkeley, Berkeley, CA; 2) Center for Computational Biology, UC Berkeley, Berkeley, CA; 3) Department of Statistics, UC Berkeley, Berkeley, CA

---

Many commonly-used statistical genetics/genetic epidemiology methods, such as LD Score Regression (LDSC), attempt to control for population structure. Recent evidence has shown that background selection in tandem with population structure can substantially bias estimates of narrow-sense heritability  $h^2$ . This bias is due in part to the effect of selection on linkage disequilibrium (LD). However, little work has been done to investigate more generally how selection has impacted LD in humans, and the consequences for statistical genetics methods. We use model-based evolutionary simulations to demonstrate the effects of a wide range of selection scenarios on heritability estimation using LDSC. We also investigate the extent to which selection drives LD between causal variants (i.e., the Bulmer Effect), and the bias this incurs in genome-wide associations (GWAS). Finally, we analyze empirical patterns of LD between expression quantitative trait loci (eQTLs). Our results suggest that both LDSC and estimated effect sizes from GWAS may have been susceptible to biases induced by selection-driven linkage patterns.

## PgmNr 2384: Standardized statistics to detect balancing selection utilizing substitution data.

### Authors:

K.M. Siewert <sup>1,2</sup>; B.F. Voight <sup>3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Harvard T.H. Chan School of Public Health, Boston, MA; 2) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 5) Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Long-term balancing selection leads to a build-up of alleles at similar allele frequencies at a balanced locus. Recently, we developed the  $\beta$  statistic to detect balanced haplotypes using this signature without the need for genomic data from an outgroup species. Here we extend this method to incorporate between-species substitution data, enabling this approach to outperform existing summary statistics in simulations. We also present the variance of  $\beta$  with and without substitution data, allowing calculation of a standardized score. Besides providing a measure of significance, this enables a proper comparison of  $\beta$  values across varying underlying parameters - a feature lacking from some related methods. We apply the standardized  $\beta$  statistic to 1000 Genomes data, along with substitution data from chimpanzee, to identify balanced loci in humans. One of our strongest, and novel, signals is near a bitter taste receptor, a gene class previously suggested to be under balancing selection. Additional top loci are both associated with complex traits and expression levels of nearby genes, suggesting that balancing selection may be affecting complex trait architecture via gene regulation.

## PgmNr 2385: A likelihood method for estimating selection on polygenic traits while controlling for pleiotropy.

### Authors:

A.J. Stern <sup>1</sup>; N.A. Zaitlen <sup>2</sup>; R. Nielsen <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Computational Biology, UC Berkeley, Berkeley, California.; 2) Department of Human Genetics, David Geffen School of Medicine, UC Los Angeles, Los Angeles, CA; 3) Department of Integrative Biology, UC Berkeley, Berkeley, California.; 4) Department of Statistics, UC Berkeley, Berkeley, California.

---

There has been substantial interest in identifying selection in the human genome. While there are a handful of examples of strong selection acting primarily on one locus, there is an overall paucity of strong signals of selection at individual loci. Furthermore, genome wide association studies (GWASs) have shown that most human traits of interest are highly polygenic. Thus, it is commonly speculated that polygenic adaptation may be a dominant mode of human adaptation, as it would allow traits to differentiate rapidly while only leaving subtle signals at each locus. Furthermore, compared to inferring whether selection has acted on a locus, inference of selection on traits can be confounded by genetic correlations; these can induce indirect selection on correlated traits, making it unclear which trait actually affects fitness if one tests traits marginally rather than jointly. To this end, we present a full-likelihood method to pinpoint selection among a set of correlated traits. We calculate the likelihood by aggregating estimates of both the local coalescent tree and effect sizes across loci. We show using simulation studies that even at modest sample sizes, our method is able estimate the strength and timing of selection, and accurately disentangles selection among genetically correlated traits. We apply our method to pinpoint selection among a set of genetically correlated anthropometric traits (including height), re-analyzing previous claims that waist-to-hip ratio appears has been under selection in Europeans, even after correcting for correlated response. We also analyze eQTLs from the GTEx consortium, using our method to identify tissues enriched for selection on tissue-specific eQTLs.



## PgmNr 2386: Mortality, comorbidity and fertility effects of a per-individual burden score.

### Authors:

H. Vy <sup>1</sup>; A. Dobbyn <sup>1</sup>; D. Jordan <sup>1</sup>; D. Balick <sup>1,2</sup>; R. Do <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Personalized Medicine, Icahn School of Medicine at Mount Sinai, NY; 2) Division of Genetics, Brigham & Women's Hospital, Harvard Medical School, Boston

---

Human genetic variation harbors a large number of deleterious variants which collectively can decrease an individual's fitness. Despite complex disease genetic studies showing that polygenic risk scores, which measure the cumulative effect of disease-associated variants typically with common or low-frequencies (>0.5%), can have profound effects on complex disease, little is known about the cumulative effect of deleterious variants are acted upon by purifying selection at the individual level and its effect on disease. In the present study, analogous to polygenic risk score analyses, we developed a per-individual burden score by summing the effect of deleterious variants after applying a weight reflecting purifying selection. Using the site frequency spectrum, we evaluated four established methods for our weight: Gerp, phyloP, CADD, and fitcons to see how well each score approximates purifying selection by qualitatively tracking with allele frequency, and observed that phyloP was the best performing score. We next calculated a per-individual burden score, weighted by phyloP, using five sets of variants: LOF sites; LOF plus polyphen damaging sites; LOF plus polyphen damaging and polyphen possibly damaging sites; synonymous site; and all variants available. These scores were calculated for both imputed data (only sites with imputed INFO score  $\geq 0.9$ ) from the global screening genotyping array in 258,491 individuals and from a whole exome data set in 29,658 individuals from the UK biobank. We tested the associations of our burden scores with *a priori* phenotypes believed to be affected by natural selection, mainly: mortality (five year mortality), comorbidity (Elixhauser comorbidity index), and fertility (number of children, survival rate). We observed consistent nominal associations of the burden score derived from LOF+polyphen damaging with number of children in both the imputed (beta: -0.004/SD units, SE: 0.002,  $P=0.007$ ) and whole exome dataset (beta: -0.01/SD units, SE: 0.005,  $P=0.04$ ). We further observed significant nominal signals of our burden score with comorbidity (beta: 0.02/SD units, SE: 0.008,  $P=0.008$ ) in the exome dataset only. These results suggest that the accumulation of deleterious variants at the individual level can have a detectable impact on mortality, comorbidity and fertility.

## PgmNr 2387: Insights on hair, skin, and eye color of ancient and contemporary Native Americans.

### Authors:

C. Mendes-Junior<sup>1</sup>; T.M.T. Carrato<sup>1</sup>; L. Marcorin<sup>2</sup>; G. Debortoli<sup>3</sup>; T. Hünemeier<sup>4</sup>; H. Norton<sup>5</sup>; E.J. Parra<sup>3</sup>; E.C. Castelli<sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Departamento de Química, Universidade de São Paulo (FFCLRP-USP), Ribeirão Preto, SP, Brazil; 2) Departamento de Genética, Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brazil; 3) Department of Anthropology, University of Toronto at Mississauga, Mississauga, ON, Canada; 4) Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo (IB-USP) São Paulo, SP, Brazil; 5) Department of Anthropology, University of Cincinnati, Cincinnati, OH, United States of America; 6) Department of Pathology, São Paulo State University (FMB-UNESP), Botucatu, SP Brazil

---

Over the past few years, tools capable of predicting pigmentation phenotypes have been developed to contribute for criminal and anthropological investigations. Recently, some studies made public sequencing data from Native Americans from Late Pleistocene/Early Holocene. In this study, we used four genetic systems (Snipper10/12/23 and HirisPlex-S) to infer eye, hair, and skin color, and one (SNPforID 34-plex) to infer genomic ancestry, of 20 ancient and 22 contemporary Native Americans. To achieve this goal, we retrieved 92 SNPs from these 42 samples available in online repositories of DNA sequences. Pigmentation phenotypes were predicted using multinomial logistic regression (HirisPlex-S) and naïve Bayes (Snipper) methods. For ancestry analysis, we used the Structure software. This workflow made possible the prediction of eye, hair and skin pigmentation for 29 samples. HirisPlex-S predicted brown eyes for all individuals, while Snipper23 predicted brown eyes for 16 individuals, green/hazel eyes for 7 and blue eyes for 1. Regarding hair color, Snipper12 predicted black hair for all individuals, while HirisPlex predicted black hair for all individuals except for a single one that was predicted as having dark-brown/black hair. It should be emphasized that the individual that presented blue eyes and lighter hair is a contemporary Amerindian that presents more than 26.5% European admixture. For skin prediction with HirisPlex-S, 21 individuals presented dark skin with dark-black influence or dark-black skin with dark influence. Three individuals presented even darker pigmentation (dark-black skin), while Mexican Mixtec-2 (dark skin with intermediate influence) and Alaskan Saqqaq (intermediate skin) presented lighter skin color. Snipper10 predicted intermediate skin for all individuals except for S\_Mixtec-2 (26.2% European ancestry), which resulted in a white skin prediction. Since these prediction tools have been developed primarily based on markers identified in Europeans, there are genetic markers relevant to Native American/East Asian populations that are certainly not included in these panels. In spite of these limitations, we provide the first prediction of pigmentation phenotypes from Paleoamerican genomes. Overall, ancient Native Americans presented intermediate/brown eyes, black hair and intermediate/darker skin pigmentation, which is highly consistent with both predictions actual phenotypes observed among contemporary Native American populations.

## PgmNr 2388: Imputation in South Asian individuals using a high-coverage South Asian reference panel.

### Authors:

S. Belsare <sup>1</sup>; J. Robinson <sup>1</sup>; R. Gupta <sup>2</sup>; J. Tom <sup>3</sup>; T. Bhangale <sup>3</sup>; R.K. Rai <sup>4</sup>; A. Butterworth <sup>5</sup>; J. Danesh <sup>5</sup>; V. Mohan <sup>6</sup>; A. Ghosh <sup>7</sup>; A. Barik <sup>4</sup>; A. Chowdhury <sup>4</sup>; D. Saleheen <sup>8</sup>; S. Kathiresan <sup>9</sup>; E. Stawiski <sup>2</sup>; A. Peterson <sup>2</sup>; J. Wall <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA.; 2) Genomic Medicines Division, MedGenome, Foster City, CA; 3) Department of Bioinformatics, Genentech, South San Francisco, CA; 4) Society for Health and Demographic Surveillance, Birbhum, West Bengal, India; 5) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 6) Madras Diabetes Research Foundation, Chennai, Tamil Nadu, India; 7) GROW Research Laboratory, Narayana Nethralaya Foundation, Bengaluru, India; 8) Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA; 9) Department of Preventative Cardiology, Massachusetts General Hospital, Boston, MA

---

Imputation involves the prediction of genotypes not directly assayed in a sample of individuals using genome-wide sequenced reference data and known genotypes at a subset of positions in individuals for whom unknown genotypes need to be imputed. Use of a population specific reference panel can lead to an improvement in imputation accuracy over a non-specific reference panel. We present here the use of high-coverage sequencing of 6610 individuals, the majority of which are from South Asia, to generate reference panels for imputation in South Asian individuals.

This newly sequenced panel of South Asian individuals consists of 1812 individuals from Pakistan, 500 from Bangladesh, 1356 from South India, and 1810 from the Birbhum district of West Bengal, as well as some outgroup individuals from Africa, Europe, and East Asia. All of these samples were obtained through various medical and sequencing projects and sequenced to high coverage.

We use the new high-coverage sequences from the public 1000 Genomes project as an alternative reference panel for comparison, since it includes both, South Asian individuals as well as individuals from other populations. Both the reference panels are phased using *eagle*, and imputation using both reference panels is performed using *impute2*. Both reference panels are filtered for genotype quality and coverage to minimize errors arising from sequencing and variant calling.

We observe an improved imputation accuracy in South Asian individuals by using the South Asian reference panel as compared to using the 1000 Genomes panel, even though both panels have been sequenced at high coverage. Our final goal is to generate a reference panel which can be used for high quality imputation in South Asian individuals.

## PgmNr 2389: The genetic architecture of complex disease at mutation-selection-drift balance.

**Authors:**

J. Berg

[View Session](#) [Add to Schedule](#)**Affiliation:** Human Genetics, University of Chicago, Chicago, IL.

---

While many human diseases are known to be highly polygenic, we have little in the way of formal population genetic theory to guide our intuition about they evolve. In this talk, I will present a theoretical investigation of a complex, highly polygenic disease evolving at mutation-selection-drift balance. Under this model, we show that for all but the largest effect size alleles, the genetic architecture is insensitive to factors such as the fitness cost of disease or the environmental variance in disease liability, and instead depends only on the strength of mutational bias toward the disease state, and the distribution of effect sizes of newly arising mutations. In contrast, the disease prevalence is expected to be extremely sensitive to such external factors, and can evolve in response to changes in such factors on relatively rapid timescales.

Finally, I will show that polygenic scores for disease traits evolving at equilibrium can exhibit confusing and misleading patterns of evolution across populations and over time. These patterns may lead to the erroneous conclusion that natural selection has driven divergence among populations in the mean genetic liability burden, even when all populations are evolving under the same steady state conditions. These results suggest the need for caution in interpreting the distribution of polygenic scores across populations and in ancient DNA samples.

# PgmNr 2390: Promoter CpG density is under selection in humans and predicts genic intolerance to loss-of-function variation.

## Authors:

L. Boukas<sup>1</sup>; H.T. Bjornsson<sup>1,2,3</sup>; K.D. Hansen<sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland.; 2) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland.; 3) Faculty of Medicine, University of Iceland, 101 Reykjavík, Iceland; 4) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland.

## Background:

The current dogma states that CpG islands are not under selection (Cohen, Kenigsberg, and Tanay, 2011). Here, using large-scale sequencing data, we show that CpG density in human promoters is in fact selected for, both at the individual CpG-level and the whole-promoter level. We then use this to predict constrained promoters of protein-coding genes and lncRNAs.

## Data:

We used allele counts from the whole-genome sequences of 62,784 individuals from TOPMed, estimates of genic intolerance to human loss-of-function variation (hLoF) from gnomAD, and whole-genome DNA methylation data in the human and chimp germlines (sperm).

## Results and Discussion:

We found that, conditional on germline methylation status, variants seen in TOPMed individuals in CpGs in hLoF-intolerant promoters (i.e. promoters of hLoF-intolerant genes) were more likely to be singletons compared to these in hLoF-tolerant promoters (unmeth  $p=1.3e-9$ , meth  $p=2.5e-14$ ). Comparing human/chimp ortholog promoters, a median of 5% more CpGs were retained between species at hLoF-intolerant promoters compared to hLoF-tolerant promoters ( $p=4.1e-14$ ). These results show that individual CpGs at hLoF-intolerant promoters are under purifying selection.

However, the most efficient way to preserve promoter CpG density is through an unmethylated germline state, as this reduces the mutation rate of all CpGs by ~15-fold. We found that hLoF-intolerant promoters tend to be unmethylated in the human germline (odds ratio = 38.3,  $p<2.2e-16$ ). Despite the overall similarity between the human/chimp germline methylomes, 74% of hLoF-intolerant promoters methylated in the chimp germline, are unmethylated in the human one, versus only 5% of hLoF-tolerant promoters (odds ratio = 585.3,  $p=0.0048$ ). This shows positive selection for an unmethylated germline state, only at hLoF-intolerant promoters.

Finally, we trained a model to predict hLoF-intolerant promoters using only their CpG density and conservation, achieving 85% precision in a hold-out test set. We predicted the promoters of 182 (of 3,897) short protein-coding genes (which are hard to ascertain for hLoF intolerance due to their short length), and 64 (of 6,545) lncRNAs, to be hLoF-intolerant. Supporting our predictions, these promoters

were depleted of deletions in gnomAD, compared to promoters of well-ascertained hLOF-tolerant genes (19%,  $p=0.013$ , and 16%,  $p=0.002$ ), with the same degree of depletion as well-ascertained hLoF-intolerant promoters (18%,  $p=9.2e-10$ ).

## PgmNr 2391: Re-examining the evidence for polygenic adaptation at height-associated loci in mainland Europe and Sardinia.

### Authors:

C. Chiang<sup>1,2,3</sup>; M. Chen<sup>2</sup>; M. Akiyama<sup>4</sup>; C. Sidore<sup>5</sup>; K. Ishigaki<sup>4</sup>; Y. Okada<sup>4,6</sup>; Y. Kamatani<sup>4,7</sup>; D. Schlessinger<sup>8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Preventive Medicine, Keck School of Medicine USC, Los Angeles, California.; 2) Center for Genetic Epidemiology, Keck School of Medicine USC, Los Angeles, California; 3) Quantitative Computational Biology Section, Department of Biological Sciences, USC, Los Angeles, California; 4) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy; 6) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka Japan; 7) Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 8) Laboratory of Genetics and Genomics, National Institute on Aging, US National Institutes of Health, Baltimore, Maryland

---

Adult height was one of the earliest putative examples of polygenic adaptation in humans. By constructing polygenic height scores using effect sizes and frequencies from hundreds of genomic loci robustly associated with height, it was demonstrated that Northern Europeans were genetically taller than Southern European beyond neutral expectation. However, this conclusion was recently challenged. Sohail et al. and Berg et al. showed the polygenic signature disappeared if GWAS summary statistics from UK Biobank (UKB) were used in the analysis, suggesting that residual uncorrected stratification from large-scale consortium studies like GIANT were responsible for the previously observed difference in genetic. While effect sizes in consortium GWAS were likely over-estimated, it remains an open question whether height loci exhibit signals of polygenic adaptation in any human population. In this study, we re-examined this question focusing on one of the shortest European populations, the Sardinians. We showed that previously reported signature of polygenic adaptation in Sardinians remained when using height loci and effect sizes ascertained from UKB (~ 0.39 s.d. shorter than CEU,  $P = 2.18e-5$ ). To further alleviate concerns of biased ascertainment of GWAS loci, we ascertained height loci from the Biobank of Japan (BBJ), and showed that the results were unchanged (~ 0.36 s.d. shorter than CEU,  $P = 8.85e-7$ ). As in Sohail et al. and Berg et al., when we focused on mainland European populations, we could not detect a polygenic selection signal using loci ascertained from BBJ data. However, we observed a nominally significant signal of allele frequency differences using precise frequency estimates from gnomAD ( $P = 0.02$ ), as well as a stronger adaptive signature using tSDS ( $P = 3e-4$ ). In summary, by examining height loci ascertained in a distant East Asian population, we further supported evidence of polygenic adaptation among the Sardinians. By contrast, an adaptive signature in mainland Europe is subtler, only becoming more evident in haplotype-based analysis.

## PgmNr 2392: The effect of consanguinity on between-individual identity by descent on the X chromosome.

### Authors:

D. Cotter <sup>1</sup>; A. Severson <sup>1</sup>; N. Rosenberg <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Biology, Stanford University, Stanford, CA

---

Consanguineous unions increase the frequency at which identical genomic segments are inherited along separate paths of descent, producing runs of homozygosity (ROH) within individuals. Consanguinity increases the lengths of ROH, and it has also recently been shown to increase autosomal identity by descent (IBD) *between* individuals in a population. This result was obtained by demonstrating that the mean time to the most recent common ancestor ( $T_{MRCA}$ ) for a pair of lineages in separate individuals decreases with increasing consanguinity, noting that genomic sharing increases with decreasing  $T_{MRCA}$ . Here, we extend this analysis to the X chromosome. We consider X-chromosomal IBD under a coalescent model with diploid, male-female mating pairs. We examine four possible types of first-cousin mating—matrilateral parallel, matrilateral cross, patrilateral parallel, and patrilateral cross—which are equivalent in their effects on autosomes, but which have different effects on the X chromosome. We calculate mean  $T_{MRCA}$  for X-chromosomal lineages sampled both within and between individuals. Our results illustrate an effect due to consanguinity on X-chromosomal  $T_{MRCA}$  that differs from the autosomal pattern under matrilateral first-cousin mating but not under patrilateral first-cousin mating. For matrilateral first cousins, the effect of consanguinity in reducing  $T_{MRCA}$  is stronger on the X chromosome than on the autosomes, with an increased effect of parallel-cousin mating versus cross-cousin mating. The differences produced by the various types of first-cousin mating potentially make it possible to distinguish between mating regimes using knowledge about the extent of within- and between-individual IBD on X chromosomes and autosomes in a population.



## PgmNr 2393: Mitochondrial DNA and Y chromosome variation among the South American Macushi in relation to geography and language.

### Authors:

K.L. DeRosa; C.A. Koch; D.A. Merriwether

[View Session](#) [Add to Schedule](#)

**Affiliation:** Anthropology, BINGHAMTON UNIVERSITY, BINGHAMTON, NY.

---

The Macushi are a Carib-speaking population indigenous to the Roraima state of northern Brazil that occupy the sparsely populated grasslands of the region. Their kinship system follows a bifurcate-merging pattern with preferential cross-cousin marriage. The Macushi share territories with the Wapishana and have established trade routes with the Yeukana. They also have had regular contact with non-indigenous cattle ranchers since the 19th century. Mitochondrial DNA (mtDNA) analysis of 94 maternally unrelated individuals was performed using the d-loop region (15,938-429) to assess the genetic diversity of the population. Y-chromosome analysis was performed on 50 paternally unrelated individuals using *DYS199*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, and the *YAP* indel. The Macushi population is comprised of haplogroups A (n=2), B (n=47), C (n=18), and D (n=27), as expected for indigenous South American (SA) groups, suggesting the absence of admixture despite contact with non-indigenous groups. These data were compared with previously published mtDNA data from SA, indicating female migration is greatest with Amazon and Andean populations. AMOVA was used to determine the relationship between geography, language (using the Loukotka, Greenberg, and Campbell classifications), and genetic distance, with both geography and language significantly positively correlated with genetic distance. The *YAP* insertion on the Y-chromosome is present in low frequencies in SA and was not identified in the Macushi. *DYS199T* was present in 79% of males, agreeing with the average ratio of this SNP observed in indigenous SA populations. Contrary to these findings, Y-STR analysis suggests there is significant admixture as approximately 30% and 26% of the population consists of haplogroups L and T respectively (which are primarily in SE Asia, Africa, and the Mediterranean). Additional Y-loci are being screened to better assess the extent of admixture observed. Y-chromosome data was further compared to previously published data, indicating widespread male migration throughout SA. Similar to our mtDNA analyses, AMOVA of the Y-chromosome data found a significant positive correlation between genetic distance, geography, and language.

The analyses for both mtDNA and Y-chromosome data indicate differing migration patterns for males and females of the Macushi population. However, further Y-chromosome markers are needed in order to confirm these findings and the presence of admixture within the Macushi.

## PgmNr 2394: Leveraging chromosomal linkage to infer natural selection from full-genome sequencing data.

**Authors:**

E. Friedlander; M. Steinrücken

[View Session](#) [Add to Schedule](#)

**Affiliation:** Ecology & Evolution, University of Chicago, Chicago, IL.

---

We present a method for inferring natural selection from full-genome sequencing data. Natural selection for (or against) a certain allele tends to increase (or decrease) its frequency within a population. As a result, selection leaves signatures in sequencing data that can be used to infer the mode and strength of selection. Due to chromosomal linkage, selection also impacts the variation at nearby neutral sites. Leveraging this signature, one can substantially increase the power of inference methods. However, changes in population size can yield patterns in the data that mimic the effects of selection. Therefore, the demographic history of the population has to be explicitly accounted for, which is not standard practice in most inference frameworks.

Our framework is built on the two-locus Wright-Fisher diffusion that describes the haplotype frequency dynamics of two linked loci separated by a certain recombination distance. General explicit solutions are not known for the transition density of this diffusion when selection and recombination act simultaneously, as it requires solving a multidimensional system of partial differential equations. Thus, we use the so-called 'Method of Moments', in which the moments of the transition density are expressed as solutions to ordinary differential equations. These moments can subsequently be used to compute the likelihood of the observed genetic variation in a sample from the population. A key technical challenge in this method is that the moments do not 'close' in models with selection. Namely, moments of order  $n$  depend on those of order  $n+$ ,  $n+2$ , etc. We surmount this challenge by developing a novel method to estimate higher order moments from those of lower order, which can also be applied to the general problem of estimating allele frequency spectra for large samples from smaller samples.

Using these efficient approximations to the dynamics of the diffusion and two-locus likelihood computations, we develop a composite likelihood framework for estimating the strength of selection from full-genome sequencing data in a population with arbitrary population size history. We demonstrate the accuracy and efficiency of the proposed methods on simulated data and show applications to the 1000 genomes dataset.

## PgmNr 2395: Adaptive archaic introgression of copy number variants introduces novel genes in humans.

### Authors:

P.H. Hsieh<sup>1</sup>; R.V. Vollger<sup>1</sup>; V. Dang<sup>1</sup>; D. Porubsky<sup>1</sup>; C. Baker<sup>1</sup>; S. Cantsilieris<sup>1</sup>; K. Hoekzema<sup>1</sup>; A.P. Lewis<sup>1</sup>; K.M. Munson<sup>1</sup>; M. Sorensen<sup>1</sup>; Z.N. Kronenberg<sup>1</sup>; S. Murali<sup>1,2</sup>; B.J. Nelson<sup>1</sup>; G. Chiatante<sup>3</sup>; F. Magiolini<sup>3</sup>; H. Blanche<sup>4</sup>; J.G. Underwood<sup>1,5</sup>; F. Antonacci<sup>3</sup>; J.F. Deleuze<sup>4</sup>; E.E. Eichler<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA; 2) Howard Hughes Medical Institute, University of Washington, Seattle, WA; 3) Department of Biology-Genetics, University of Bari, Bari, Italy; 4) Fondation Jean Dausset - Centre d'Etude du Polymorphisme Humain, Paris, France; 5) Pacific Biosciences (PacBio) of California, Incorporated, Menlo Park, CA

---

Copy number variants (CNVs), generated through deletions or duplications that affect large numbers of base pairs (bp) in the genome, are subject to stronger selective pressure than single-nucleotide variants (SNVs), but their roles in archaic introgression and adaptation have not been systematically investigated. In this study, we characterize Melanesian genomes and observe an enrichment of selective CNVs (n=37); of which 19 CNVs are likely introgressed from archaic hominins. We provide multiple lines of evidence for adaptive introgression of large CNVs at chromosomes 16p11.2 and 8p21.3 from Denisovan and Neanderthal, respectively, in Melanesians, and these variants are absent from most of other human populations. Using long-read sequencing data, we sequence resolve a large duplication of >383 thousand base pairs (kbp), which originated at chromosome 16p12.2, and show that the duplication was inserted at 16p11.2 and has a Denisovan ancestry. Our results show that this duplication is in high frequency (>79%) among Melanesian groups and has signatures of positive selection. At the locus 8p21.3, we identify a haplotype that carries two CNVs, a ~6 kbp deletion and a ~38 kbp duplication, with a Neanderthal origin in Melanesians and reaches to high frequency (44%) with signals of selection. We estimated the archaic haplotypes at the 16p11.2 and 8p21.3 loci introgressed into modern human gene pool 0.06-0.17 and 0.04-0.12 million years ago from Denisovan and Neanderthal, respectively. Using long-read sequencing genomic and transcriptomic data, we reconstruct the structure and complex evolutionary history for these polymorphisms and find novel protein-coding genes that have positively selected amino acid substitutions in both loci. Combined, our results suggest that large CNVs originating in archaic hominins and introgressed into modern humans have played an important role in local population adaptation and represent an under-ascertained source of large-scale genetic variation.

## PgmNr 2396: Mapping annual fine-scale genetic structure of Finland through the 20<sup>th</sup> century.

### Authors:

S. Kerminen <sup>1</sup>; A.S. Havulinna <sup>1,2</sup>; M. Perola <sup>2</sup>; V. Salomaa <sup>2</sup>; M.J. Daly <sup>1,3</sup>; S. Ripatti <sup>1,3,4</sup>; M. Pirinen <sup>1,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland, Helsinki Institute of Life Sciences, University of Helsinki, Helsinki, Finland; 2) National Institute of Health and Welfare, Helsinki, Finland; 3) Broad Institute of MIT and Harvard University, Cambridge, MA, USA; 4) Department of Public Health, University of Helsinki, Helsinki, Finland; 5) Helsinki Institute for Information Technology HIIT and Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

---

We illustrate the current possibilities for individual-level ancestry estimation by characterizing spatial and temporal changes in the genetic history of Finland in the 20<sup>th</sup> century.

Starting from 18,715 unrelated individuals from the National FINRISK Study, we identify a set of 1,600 reference samples who define 15 genetically homogeneous reference groups of the Finnish population. These samples are both geographically well-defined (parents born within 80 km) and strongly representative of their genetic group (ancestry component >0.5 in FineSTRUCTURE assignment). We estimate the genetic ancestry profiles of the remaining 17,115 samples with respect to the reference groups using SOURCEFINDv2.

We use the birth places (covering whole of mainland Finland) and birth years (ranging 1923-1987) of the 18,715 individuals to estimate how the contribution from each of the reference groups develops in time and across the country. As expected, more rural regions of Eastern Finland tend to be dominated by a single reference group while more urban regions of Southern Finland show diverse genetic profiles. While the individual genetic profiles on average diversify towards the present day, we observe regional differences in rate of change with the highest increase in diversity in southwestern Finland but no increase, for example, in Ostrobothnia.

We assess our estimates by monitoring how the genetic component of the ceded Karelia develops in time. Over 400,000 Karelians were relocated from the ceded areas (southeast of Finland) to other parts of Finland during and after the war 1939-1945. We detect this event as a quick change in yearly genetic profiles of other parts of Finland and can accurately assign this change to the particular component from the ceded Karelia as opposed to other reference groups from Eastern Finland. We also map the yearly patterns of further movements of the Karelian evacuees across the country. For example, we see that in Ostrobothnia the component of Karelian evacuees shrinks away already by 1950 whereas in Southwest Finland the component stabilizes to a constant level by that year.

To our knowledge, this is the first study where the annual contributions of genetic ancestry from closely related, neighboring subpopulations have been successfully separated from each other across a country. The work demonstrates the power of genetic ancestry estimation to reveal local population history and provides new avenues for individual-level ancestry estimation.

## PgmNr 2397: Fast estimation of effective migration surfaces.

### Authors:

J.H. Marcus <sup>1</sup>; W. Ha <sup>2</sup>; R. Foygel Barber <sup>3</sup>; J. Novembre <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University Chicago, Chicago, Illinois.; 2) Department of Statistics, University of California Berkeley, Berkeley, California.; 3) Department of Statistics, University Chicago, Chicago, Illinois.; 4) Department of Ecology and Evolution, University Chicago, Chicago, Illinois.

---

Understanding the relationship between geography and genetic differentiation is a long-standing problem in population genetics. A widespread feature in population genetic data is 'isolation by distance,' where genetic differentiation tends to increase as populations become more geographically distant. While spatial models of 'isolation by distance' have provided powerful predictions of population structure, they do not capture heterogeneous migration that occurs when geographic features, such as a mountain range, can slow down or enhance gene-flow in localized regions. Recently, Petkova et al. 2016 developed an elegant statistical method called Estimating Effective Migration Surfaces (EEMS) for visualizing heterogeneous effective migration rates on a geographic map using a coalescent inspired model for gene-flow on a dense spatial graph. EEMS has been applied to many systems and has allowed for the discovery and interpretation of geographic features that might have restricted or enhanced gene-flow. Most importantly, EEMS is a flexible tool for users to visualize geographic structure in their data. EEMS uses Markov Chain Monte Carlo (MCMC) to obtain posterior samples of these migration rates, employing a spatial prior using a Voronoi tessellation of the focal geographic region, which favors piecewise constant structure in the migration process. Here we draw inspiration from EEMS by utilizing a similar graph-based likelihood, allowing for heterogeneous gene-flow, but take an optimization-based approach. We impose spatial structure in the migration rates, allowing for fast model exploration and iteration. We take advantage of recent developments in the generalized lasso literature to adaptively fit a piecewise polynomial function defined on a graph, without having to pre-specify change-point locations. Our method accurately recovers migration surfaces for within-model and coalescent-based simulations. Applications to Human population genetic datasets from around the globe perform comparably to EEMS but are fit orders of magnitude faster. We see our method as being a useful complementary tool alongside EEMS, expanding the ability for users to quickly visualize and interpret spatial structure in their data.

## PgmNr 2398: Implementation of personalized ancestry information in the Colorado Biobank.

**Authors:**

J. Shortt; N. Rafaels; K. Crooks; K. Barnes; C. Gignoux

[View Session](#) [Add to Schedule](#)

**Affiliation:** Colorado Center for Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO.

---

Knowledge of genetic ancestry is a critical component in modern genetic research, but is also of increasing interest among consumers and patients within health systems. With over 100,000 consented participants as of May 2019, the Colorado Biobank ([cobiobank.org](http://cobiobank.org)) is a clinical and research biorepository offering implementation and return of results to patients within the UCHHealth system. As an additional offering, we are implementing a genetic ancestry inference pipeline, with features relevant to both patients interested in ancestry testing and the broader research community. Here we present key features of the Colorado Biobank's Personalized Ancestry Information Resource (PAIR), including global ancestry, recent relatedness, archaic admixture, and Y-chromosome and mitochondrial haplogroup features, from the pilot data comprised of 3,967 participants typed on a custom Illumina MEGA array. We compare the distribution of participants' self-reported race and ethnicity in the context of well-characterized global reference panels, and highlight several key points to consider in returning ancestry results to biobank participants. Although most (77%) participants are of majority European descent, there are substantial proportions (2%-15%) of participants with Indigenous American, sub-Saharan African, East Asian, South Asian, Near Eastern and Oceanian ancestry. Our pipeline can efficiently identify closely related individuals including 16 first- and second-degree relatives. Recent shared ancestry via identity-by-descent estimation and UMAP dimensional reduction also identifies 2 clusters of cryptically-related participants sharing on average 23 cM, and comprising over 10% of our pilot dataset. Beyond analysis, PAIR will provide educational tools and walkthroughs to help educate the Colorado Biobank community. As the size of the Colorado Biobank grows, we anticipate this resource providing additional ancestry information relevant to researchers and participants alike.

## PgmNr 2399: Constructing recombination maps for populations with complex ancestry.

**Authors:**

G.E. Van Eeden; C. Uren; M. Möller

[View Session](#) [Add to Schedule](#)

**Affiliation:** DST-NRF Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch Univer

---

It has been shown that the majority of human genetic studies are not representative of the diversity that exists globally, since they are based on populations of European descent. The lack of inclusion of under represented ethnicities in human genetic studies also means that when these studies are translated into clinical practice, the resulting policy or protocol might be incomplete or even mistaken. Studies that investigate the genetic factors that affect disease often rely on a recombination map. There are no publicly available population specific recombination maps for southern African populations at the moment. Southern African populations have complex ancestry as a result of many migration events that have occurred over the last few centuries. Software used to develop population specific recombination maps are well-established, however, most software have either never been tested on populations with complex ancestry or have only been tested on, at most, a three-way admixed population. Any research on southern African populations with complex ancestry that require a recombination map would thus benefit from a protocol that directs the creation of population specific recombination maps using established software. However, if such a protocol cannot be established, a new approach to create high resolution population specific recombination maps for populations with complex ancestry needs to be developed. We are currently working on determining what adjustments can be made to pedigree and linkage disequilibrium based methods in order to accommodate multi-way admixed populations and comparing the results to publicly available recombination maps. The next step would be to determine whether these population specific maps yield better results than the available maps when used in an ancestry association study.

## PgmNr 2400: Study genetic ancestry and homozygosity disequilibrium in global populations.

**Authors:**

H.-C. Yang; C.-W. Chen; Y.-T. Lin; S.-K. Chu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

---

Studying genetic ancestry of global population has been long recognized as an important topic in population genetics, evolution, and pharmacogenomics. This study analyzed more than 77 million of single nucleotide variation in The 1000 Genomes Project – Final Phase, which provided the whole-genome sequencing data of 2,504 individuals in 26 global populations from four major continents (the Continent of Africa, Americas, Asia, and Europe). An ultrahigh-dimensional whole-genome principal component analysis and hierarchical clustering analysis demonstrated genetic similarity and differentiation of the 26 studied populations. In addition to an identification of ancestry informative markers and construction of prediction panels for global populations, a whole-genome homozygosity disequilibrium analysis revealed that the African-ancestry group carries a lower number and shorter regions of homozygosity disequilibrium. In contrast to the indigenous residents in Africa, the African-ancestry diaspora populations undergo a much more diverse genetic admixture. The American-ancestry group and populations exhibit the largest spread in the number and length of regions of homozygosity disequilibrium. The South Asian-ancestry group and populations exhibit a number of outlier individuals who carry much more regions of homozygosity disequilibrium. The East Asian-Ancestry group and populations carry the maximum number and widest regions of HD among the studied ancestry groups. This study developed an efficient algorithm for an ultrahigh-dimensional principal component analysis, identified ancestry informative loci/regions, constructed prediction panels for genetic ancestry, and mapped genes under homozygosity disequilibrium. The results reveal differential population structures, demographic history, and population evolution of the studied ancestry groups.



## PgmNr 2401: Genetic structure effects of ancestry-based and trait-based assortative mating.

**Authors:**

Y. Banda <sup>1</sup>; K.A. Rand <sup>1</sup>; J.M. Granka <sup>1</sup>; E.L. Hong <sup>1</sup>; K.G. Chahine <sup>2</sup>; C.A. Ball <sup>1</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Ancestry, 153 Townsend St, Suite 800, San Francisco, CA 94107.; 2) Ancestry, 1300 West Traverse Parkway, Lehi, UT 84043

---

The genetic structure of a population is determined by its mating patterns. This implies that the distribution of alleles and genotypes could be impacted by spouse correlations for genetic ancestry and traits. While trait-based assortative mating leads to both within-locus correlation and between-locus correlation for alleles associated with a trait, ancestry-based assortative mating leads to the same phenomena but this applies to all loci that exhibit differences in allele frequency between ancestral populations, whether they are associated with a trait or not. In this study, we evaluate spouse-pairs who are customers of AncestryDNA and have consented to participate in scientific research to assess the general impact of assortative mating on the genetic structure of a population, and also to compare the trait and genetic ancestry correlations. Phenotype data is obtained from survey responses collected from a subset of AncestryDNA customers and is used to calculate the spouse-pair trait correlations. The genetic ancestry correlations are calculated using principal components (PCs) scores obtained from the PC analysis of genome-wide data. We also explore the genetic effects of these two types of assortative matings by characterizing the genetic structure of the offspring. The results of these studies have particular impact on studies such as genome-wide association studies (GWAS), since ignoring genetic structure may lead to many false positive loci detected, and incorrect estimation of genetic relationships and heritability. These results also strongly highlight the need for careful design and interpretation of population-based genetic studies given the complex nature of structured mating.

## PgmNr 2402: The genetic structure and admixture history of Hui population.

### Authors:

Ma <sup>1</sup>; W. Yang <sup>2</sup>; X. Wang <sup>1</sup>; K. Yuan <sup>1</sup>; Y. Lu <sup>1</sup>; D. Lu <sup>1</sup>; S. Xu <sup>1,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) CAS-MPG Partner Institute for Computational Biology, Shanghai 200031, China; 2) Key Laboratory of Reproduction and Heredity of Ningxia Region, Ningxia Medical University, Yinchuan, China; 3) School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China; 4) Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China; 5) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China

---

With a population size of 10.5 million, Hui is the second largest of 55 minorities and the largest of the 10 official Muslim ethnic groups in China. However, the genetic study of Hui population has been underrepresented. In this study, we presented the first whole genome investigation by whole genome sequencing 234 individuals from 5 prefectures across Ningxia Hui Autonomous Region (NX), China. We comprehensively assessed the fine scale population structure and inferred the demographic history of Hui by leveraging newly developed powerful tools. Our results showed that there are four major ancestral components in Hui population derived from populations in East Asia (~82.8%), Siberia (~7.8%), West Eurasia (~7.0%), and South Asia (~2.4%). We observed a distinct South-North cline across Ningxia. From North to South, the East Asia ancestral component decreased from 87.0% to 78.5%, the West Eurasia and South Asia ancestral components increased from 5.8% to 9.6% and from 1.4% to 3.4%, respectively, while the Siberia ancestral component varied little among regional sub-populations. Hui could be modeled as an admixture of two already admixed ancestries: Eastern ancestry consisted of East Asia and Siberia ancestral components and West ancestry composed of West Eurasia, South Asia and Siberia ancestral components. Leveraging haplotype-based methods, NX Hui can be divided into two clusters: North and South. The Xinjiang (XJ) Hui showed closer relationship with South NX Hui than with North NX Hui, which could reflect recent migration into XJ. Further, we employed different methods to infer admixture history of Hui and two admixture events were detected. An ancient wave of admixture occurred about 40 generations (1,200 years) ago and a recent wave of admixture occurred nearly 20 generations (600 years) ago, which was largely consisted with historical records. This study is expected to help understand the genetic origin and admixture history of the Hui people, the spread of Islam in China and the gene flow among Eurasian populations.

## PgmNr 2403: The genetic structure of pastoralists in northern Kenya.

### Authors:

A.M. Taravella <sup>1</sup>; C. Handley <sup>2</sup>; E.K. Howell <sup>1</sup>; A.C. Stone <sup>2</sup>; S. Mathew <sup>2</sup>; M.A. Wilson <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) School of Life Sciences, Arizona State University, Tempe AZ; 2) School of Human Evolution and Social Change, Arizona State University, Tempe AZ

---

Genetic analyses across human populations in Africa have shed light onto the details of major population movements and migrations within Africa. However, the genetic history and how it relates to cultural practices among human populations in Africa on a more local geographic scale remain relatively unexplored. This project aims to understand the genetic relationships within and among four pastoral populations inhabiting northern Kenya - the Turkana, Borana, Samburu, and Rendille. The Turkana and Samburu speak eastern Nilotic languages that are part of the Nilo-Saharan language family, while the Rendille and Borana are Cushitic speakers, languages that are part of the Afro-Asiatic family. Today, these four populations reside among ethnic conspecifics, where ethnolinguistic territories are geographically adjacent to one another. Additionally, these populations practice patrilocal - a post marital residence system in which males stay and affiliate with their birth clan while the women move to their husband's clan after marriage. Marriages are exogamous at clan or moiety kinship levels, where individuals do not marry close relatives. Have these populations remained relatively isolated or is there evidence of migration among these groups? Also, does their genetic composition reflect their marital practices? To address these questions, we SNP genotyped 378 individuals across these four populations. For each population, we sampled individuals from at least 2 different clans. We performed principal components analysis, ADMIXTURE analysis, and  $F_{st}$  to visualize the genetic structure and differentiation among these groups and to identify any genetic relationships among these groups. Additionally, for each population, we performed sex-stratified clan based  $F_{st}$  across the autosomes and X chromosome to identify if genetic males have higher genetic differences among clans. Initial results suggest that, broadly, genetics mirrors geography, however preliminary analyses suggest shared genetic relationships among populations based on culture over geography. As Africa is the birthplace of modern humans, knowledge of the genetic composition and history of African populations can enrich our understanding of human evolution and inform studies of human health and disease.

## PgmNr 2404: Large-scale whole-genome sequencing of three diverse Asian populations in Singapore.

### Authors:

C. Wang<sup>1,2</sup>; D. Wu<sup>1,2</sup>; J. Dou<sup>1</sup>; X. Chai<sup>1</sup>; C. Bellis<sup>3,4</sup>; A. Wilm<sup>5</sup>; C.C. Shih<sup>5</sup>; W.W.J. Soon<sup>6</sup>; N. Bertin<sup>1</sup>; C.C. Khor<sup>3</sup>; M. DeGiorgio<sup>7,8</sup>; S. Cheng<sup>2</sup>; L. Bao<sup>2</sup>; N. Karnani<sup>9,10</sup>; S. Davila<sup>11,12</sup>; P. Tan<sup>11,13,14,15</sup>; A. Shabbir<sup>16</sup>; A. Moh<sup>17</sup>; E.-K. Tan<sup>18,19</sup>; J.N. Foo<sup>3,20</sup>; L.L. Goh<sup>21</sup>; K.P. Leong<sup>22</sup>; R.S.Y. Foo<sup>3,23,24</sup>; C.S.P. Lam<sup>19,23,25,26</sup>; A.M. Richards<sup>23,24,27</sup>; C.-Y. Cheng<sup>28,29,30</sup>; T. Aung<sup>28,29,30</sup>; T.Y. Wong<sup>28,29,30</sup>; J. Liu<sup>3,24</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computational and Systems Biology, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore.; 2) Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China; 3) Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore.; 4) Genomics Research Centre, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia.; 5) Bioinformatics Core, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore.; 6) Integrated Genome Analysis Platform (iGAP), Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore.; 7) Departments of Biology and Statistics, Pennsylvania State University, University Park, PA 16802, USA.; 8) Institute for CyberScience, Pennsylvania State University, University Park, PA 16802, USA.; 9) Singapore Institute for Clinical Sciences, Agency for Science, Technology and Research, Singapore.; 10) Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.; 11) SingHealth Duke-NUS Institute of Precision Medicine, Singapore.; 12) Cardiovascular and Metabolic Disorders Program, Duke-NUS Medical School, Singapore.; 13) Cancer and Stem Biology Program, Duke-NUS Medical School, Singapore.; 14) Cancer Science Institute of Singapore, National University of Singapore, Singapore.; 15) Biomedical Research Council, Agency for Science, Technology and Research, Singapore.; 16) Department of Surgery, National University Health System, Singapore.; 17) Clinical Research Unit, Khoo Teck Puat Hospital, Singapore.; 18) Department of Neurology, National Neuroscience Institute, Singapore General Hospital Campus, Singapore.; 19) Duke-NUS Graduate Medical School, Singapore.; 20) Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore.; 21) Personalised Medicine Service, Tan Tock Seng Hospital, Singapore.; 22) Clinical Research & Innovation Office, Tan Tock Seng Hospital, Singapore.; 23) Cardiovascular Research Institute, National University Heart Centre, Singapore.; 24) Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.; 25) National Heart Centre Singapore.; 26) Department of Cardiology, University Medical Center Groningen, Groningen, The Netherlands.; 27) Christchurch Heart Institute, University of Otago, Christchurch 8011, New Zealand.; 28) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore.; 29) Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore.; 30) Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore.

---

Underrepresentation of Asian genomes has hindered population and medical genetics research on Asians, leading to population disparities in precision medicine. By whole-genome sequencing of 4,810 Singaporeans with Chinese, Malay, and Indian ancestry, we found 98.3 million SNPs and small insertions/deletions, over half of which are novel. Population structure analysis demonstrated great

representation of Asian genetic diversity in Singapore and revealed a Malay-related novel ancestry component. Furthermore, demographic inference suggested that Malays split from Chinese ~24,800 years ago, and experienced significant admixture with mainland East Asians ~1,700 years ago, coinciding with the Austronesian expansion. Additionally, we identified 20 candidate loci for natural selection, among which 14 harbored robust associations with complex traits and diseases. Finally, we showed that our data can substantially improve genotype imputation in diverse Asian and Oceanian populations. These results highlight the value of our data as a resource to empower human genetics discovery across broad geographic regions.

## PgmNr 2405: Estimating the ages of rare variants using long-range identity-by-descent (IBD).

### Authors:

Y. Si <sup>1</sup>; H.M. Kang <sup>1</sup>; S. Zoellner <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biostatistics Dept, University of Michigan, Ann Arbor, Michigan.; 2) Psychiatry Dept, University of Michigan, Ann Arbor, Michigan.

---

Large sequencing studies identify an abundance of rare variants, only a small proportion of which are functional. Identifying this functional subset enables designing powerful downstream analyses. As many functional variants are subject to selection, we aim to use the evidence for selection, namely their ages, to prioritize these variants.

To estimate a variant's age, we propose a Bayesian model to leverage the haplotype sharing pattern surrounding a variant in addition to the sample allele frequency. Intuitively, individuals sharing a recent mutation are also likely to share a large identity-by-descent (IBD) segment around that position. We model the length of this IBD segment conditional on the observed mutation as shaped by recent demographics, such as population expansion and structure. Our new method combines analytical and numerical solutions from coalescence theory to handle large sample size under any specified parameter settings of effective population size or history model used to infer the most recent demographics.

We use simulation to demonstrate that our new approach of jointly modeling IBD lengths, allele frequency and population history provides more accurate age estimates than previous methods, and that the tool we develop scales to samples with more than 100,000 individuals. Applying our method to the TOPMed data of >100,000 individuals to estimate the ages of all rare variants with sample allele counts below ten, we observe a two-fold enrichment of all protein-altering variants among the youngest 10% of doubletons, and a four-fold difference in the number of high impact (frameshift, stop/start lost/gain, etc.) variants when comparing the youngest to the oldest 10% doubletons. This indicates a clear signal that our age estimates serve as a functional annotation, provide new insights in interpreting GWAS findings and help grouping variants for association analysis.

## PgmNr 2406: Male control individuals show higher burden of rare coding SNVs across diverse ancestries.

### Authors:

M. Traglia; L. Weiss

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Psychiatry and Institute for Human Genetics, University of California, San Francisco, 401 Parnassus Ave, LangPorter, San Francisco, CA 94143, USA

A liability threshold model has been utilized to explain an excess of *rare* copy number variants (CNVs) and *de novo* single nucleotide variants (SNVs) in affected females compared to males observed for male-biased autism (ASD) [Jacquemont et al., 2014]. However, we surprisingly identified a significant excess of *rare* CNV mutational burden in female controls compared to male controls, particularly for deleterious CNVs [Desachy et al., 2015], later replicated independently across studies [Männik, et al., 2015; Han et al., 2016; Ruderfer et al., 2016]. We therefore hypothesized that the mutational burden of *rare* autosomal SNVs predicted to be damaging in humans, such as loss-of-function (LoF) , or missense mutations, may differ across male and female healthy carriers from diverse populations.

Assuming a random distribution of silent SNVs ( $MAF < 0.001$ ), we used 770 female-male pairs matched by synonymous SNV content and by subpopulation from 1000G phase3 data and we assessed the female and male autosomal burden of rare LoF and missense SNVs by population. We primarily observed excess burden of LoF (OR range 1.03-1.08) and missense SNVs (OR range 1.00-1.02) in male controls compared to female controls across populations. Meta-analysis of female and male burden of missense SNVs across populations confirmed significant male excess ( $OR_{Meta} = 1.01$ ;  $P < 0.008$ ). Although the LoF male excess is slightly greater ( $OR_{Meta} = 1.03$ ;  $P = 0.15$ ), it did not reach significance, likely due to the smaller number of variants or heterogeneity ( $I^2 = 45\%$ ;  $P_{het} = 0.12$ ). Conditioning for silent variant content allowed us to control for the possibility of subtle stratification and to assess population-specific rare mutation numbers across the genome. Indeed, we observed moderate correlation between LoF and silent SNV burden across sexes and populations ( $\rho = 0.29$ ;  $P < 0.01$ ).

This statistical approach will be applied to additional individual-data cohorts to further assess sexual dimorphism at human constrained loci and disease-associated loci, such as ASD. This project will help elucidate autosomal genetic sex differences, improve our understanding of sex as a biological variable, and enable more informed interpretation of rare SNVs in clinical testing.

## PgmNr 2407: Comparison study of lung function and high altitude adaptation of native Tibetan adults living at the different altitudes.

### Authors:

B. Bianba <sup>1</sup>; J. Li <sup>2</sup>; Q. Qula <sup>1</sup>; X.B. Qi <sup>3</sup>; C.Y. Cui <sup>1</sup>; O.U. Ouzhuuobu <sup>1</sup>; Y.Q. Gao <sup>4</sup>; T.Y. Wu <sup>5</sup>; Y. Yuzhen <sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) High Altitude Medical Research Center of Tibet University, Lhasa-850000, China; 2) Department of basic medicine, Tibetan University Medic College, Lhasa-850000, China; 3) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming-650223, China; 4) Department of plateau military medicine, army medical university, Chongqing- 400038, China; 5) National Key Laboratory of High Altitude Medicine, High Altitude Medical Research Institute, Xining-810012, China; 6) Department of pharmacy, Tibetan University Medic College, Lhasa-850000, China

---

**Objective** The aim of the present study was to compare the difference of lung function between native healthy Tibetan adults living at 2,200 and 5,300 m and its possible associated factors, as well as its role in altitude adaptation. **Methods** Lung function tests were performed on 287 and 229 Tibetan residents living in Yigong at an altitude of 2,200 m and Pumajiangtang at an altitude of 5,300 m. Meanwhile, arterial oxygen saturation and hemoglobin concentration were also measured. Anthropometric measurements and physiological indicators were described as mean  $\pm$ SD. Independent sample t test was used for the comparison of mean between groups, the prevalence of polycythemia between groups was tested by Chi-square test, and multiple linear regression was applied to test the correlation between the outcome variable and groups. **Results** Mean value of height, weight, arterial oxygen saturation and diastolic pressure of Yigong males were higher than those of Pumajiangtang males whereas the hemoglobin concentration was lower than those of Pumajiangtang males ( $P < 0.05$ ). The mean value of height, weight, arterial oxygen saturation, heart rate and blood pressure of Yigong women were higher than those of Pumajiangtang women, whereas the mean values of chest circumference and hemoglobin concentration were lower than those of women in Pumajiangtang ( $P < 0.05$ ). The mean value of lung function indicators of Tibetans living at the lower altitude was significantly higher than that of the Tibetans living at the higher altitude ( $P < 0.05$ ). After adjusting for gender, age and height, there was no significant difference in lung function between the two groups. **Conclusion** The elevated hemoglobin concentration and decreased arterial oxygen saturation were detected among Tibetan adults living at the higher altitude although no difference of lung function was found between the two groups. This may indicate that lung function could no longer increase when altitude reaches the Tibetan's limitation of physiological regulation, although the enhancement of lung function is an important mechanism for Tibetans to adapt to high altitude. The mechanism of adaptation to high altitude among native Tibetan highlanders needs to be further studied.



## PgmNr 2408: Transcriptional response variation of monocytes to influenza virus infection: From population differences to the single-cell.

### Authors:

M.B. O'Neill<sup>1,2,3</sup>; H. Quach<sup>1,2,3</sup>; M. Rotival<sup>1,2,3</sup>; L. Quintana-Murci<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Evolutionary Genetics Unit, Department of Genomes and Genetics, Institut Pasteur, Paris, France; 2) UMR2000, CNRS, Paris, France; 3) Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris, France

---

There is marked differences between individuals and populations in both the intensity of immune responses and susceptibility to infectious diseases. In characterizing the intra- and inter-population heterogeneity that exists in the transcriptional response of primary human monocytes following challenge with a live influenza A virus (IAV), we uncovered high inter-individual variability in RNA sequence reads of viral origin (1-13.5%). This suggests differences in the ability of IAV to enter and/or replicate among host cells. Interestingly, the average percent of IAV reads differs between African- (4.9%) and European-descent (6.8%) samples ( $p$ -value =  $1.9 \times 10^{-8}$ ), a finding that is replicated at the transcriptional (IAV mRNAs) and translational (IAV protein intensities) levels. Flow cytometry analysis suggests that in our experimental setting monocytes show varying degrees of susceptibility to infection after 6 hours of IAV challenge, and that the amount of viral mRNAs is driven by the percentage of cells that become infected. To explore whether the differences we observed in overall viral mRNA levels can be attributable to inter-individual differences in the proportions of specific monocytes sub-populations (i.e. do some individuals harbor more IAV-permissive cells), we have sequenced the transcriptomes of >100,000 individual monocytes following infection with IAV or no-virus (control) at five time points in 8 individuals who were previously found to present high or low viral mRNA levels following IAV challenge. Infected and non-infected cells are reliably distinguishable based on their transcriptomes. Within the infected condition, we find that resistant cells are characterized by increased expression of several antiviral genes, including those in type-1 interferon signaling and response to virus pathways (e.g., *IFITM3*, *APOBEC3*, *ISG20*, and *ISG15*). Using transcriptional regulatory networks inferred from single-cell profiles, we find that African- and European-descent individuals differ in terms of their activity of key antiviral transcription factors such as IRFs and STATs, these factors correlating negatively with viral mRNA levels. Collectively, these results highlight the extreme heterogeneity in response to infection observed at the cellular level, and suggest that inter-population variability in response to IAV infection is driven by differences in the proportion of cells that are able to rapidly trigger a protective response.

## PgmNr 2409: Analyzing the dynamics of short tandem repeats in large, multigenerational pedigrees.

### Authors:

C.J. Steely; W.S. Watkins; L. Baird; L.B. Jorde

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, University of Utah School of Medicine, Salt Lake City, Utah.

---

Short tandem repeats (STRs) are tandemly repeated sequences of 1-6 bp motifs. These sequences compose only a small fraction of the genome, but length polymorphisms at specific loci have been linked to several human diseases including amyotrophic lateral sclerosis, Friedreich ataxia, Huntington disease, and Fragile X syndrome. The rapid mutation rate of STRs has made them valuable for several biological applications. Improving our understanding of these length changes would increase our knowledge of the mutational dynamics of the genome and uncover additional loci that may have a role in causing disease. Here, to better estimate the genome-wide patterns of length polymorphisms at STR loci, we utilize blood-derived whole-genome sequencing data from the three-generation CEPH pedigrees, consisting of 33 families and 603 individuals. These pedigrees contain both sets of grandparents (generation 1), the parents (generation 2), and an average of 9 grandchildren (generation 3) per family. We use existing STR analysis software (HipSTR) to identify STR lengths in each generation. The lengths are compared between parents and their offspring, and *de novo* STR mutations are identified. To test for accuracy, *de novo* STR mutations identified in the 2<sup>nd</sup> generation were examined in the 3<sup>rd</sup> generation to ensure their transmission. Analyzing 1.6 million STR loci in these pedigrees, we have established an average *de novo* STR mutation rate as  $1.9 \times 10^{-4}$  mutations per locus per generation. Furthermore, we have found that this mutation rate varies with motif length, ranging from about  $3.2 \times 10^{-4}$  for monomeric repeats to  $2.1 \times 10^{-5}$  mutations per locus per generation for hexameric repeats. Phasing our *de novo* insertions back to the parent of origin, where possible, has improved our understanding of the inheritance patterns of STR loci. Additionally, the three-generation structure of the CEPH pedigrees allows for the differentiation of germline inheritance from mosaicism. Finally, for each family we examined the correlation between the number of *de novo* STR length changes and parental age at the time of birth. The large CEPH pedigrees have afforded us an opportunity to better understand the dynamics of STR length changes through multiple generations.

## PgmNr 2410: Human-specific tandem repeat expansion and differential gene expression during primate evolution.

### Authors:

A. Sulovari<sup>1</sup>; R. Li<sup>1</sup>; P. Audano<sup>1</sup>; D. Porubsky<sup>1</sup>; A. Pollen<sup>2</sup>; M. Chaisson<sup>1,3</sup>; E. Eichler<sup>1,4</sup>; The Human Genome Structural Variation Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA; 2) Department of Neurology, University of California, San Francisco (UCSF), San Francisco, CA 94143, USA; 3) Quantitative and Computational Biology, University of Southern California, Los Angeles, CA 90089, USA; 4) Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

---

Short and variable number of tandem repeats (STRs and VNTRs) are an important source of natural and disease-causing variation, but they have been problematic to resolve in reference genomes and genotype with short-read technology. We created a framework to model the evolution and instability of STRs and VNTRs in the ape lineage. We phased and assembled three nonhuman primate ape genomes (chimpanzee, gorilla, and orangutan) using long-read and 10X phasing data for 21,442 human tandem repeats discovered in six long-read haplotype-resolved assemblies of Yoruban, Chinese and Puerto Rican origin. The repeats are nonrandomly distributed with VNTRs more likely (OR = 1.39, p-value < 10<sup>-5</sup>) to overlap with genes than STRs even after controlling for subtelomeric distribution differences. We define a set of 1,584 STRs/VNTRs expanded specifically in the human lineage, including large tandem repeats such as a 6 kbp 30-meric repeat in the intron of *CACNA1C*, or protein-coding tandem repeats such as a 22 kbp protein-coding VNTR in mucin gene *MUC3A* and numerous disease-associated STRs. We systematically reconstruct sequence compositional changes for the largest expansions revealing that longer tracts of uninterrupted repeat purity associate with larger and more unstable alleles. We find that SVA retrotransposition is the main mechanism for distributing GC-rich human-specific tandem repeats expansions throughout the genome but with a bias against genes (OR = 0.57[0.38-0.85], p-value = 0.006). In contrast, we observe that VNTRs not originating from retrotransposons have a propensity to cluster near genes especially located in the subtelomeric region of chromosomes. Using single-cell expression from cerebral organoids, we identify a subset of genes where isoform usage differs significantly in humans when compared to chimpanzee cerebral organoids likely caused by cryptic splicing variation located within the human-expanded intronic VNTR. We also observe a significant overrepresentation of subtelomeric STRs/VNTRs in genes upregulated in human when compared to the chimpanzee. Interestingly, the strongest effect was observed for genes associated with transcription profiles analogous to intermediate progenitor cells (OR = 2.21 [1.59 - 3.08], p-value = 2.49×10<sup>-6</sup>). In addition to providing a valuable resource for future evolutionary studies, this resource will be useful for the identification of unstable alleles associated with human genetic disease.

## **PgmNr 2411: Intratumoral t(2;13) translocation-positive heterogeneity in pediatric alveolar rhabdomyosarcoma (ARMS) tumors correlates to patient survival prognosis.**

### **Authors:**

A. Hollenbach <sup>1</sup>; K. Gleditsch <sup>1,2</sup>; J. Peñas <sup>1</sup>; D. Mercer <sup>1</sup>; A. Umrigar <sup>1</sup>; M. Stark <sup>3</sup>; F. Tsien <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA.; 2) Division of Pediatric Hematology/Oncology, Children's Hospital of New Orleans, New Orleans, LA; 3) Department of Pathology, Children's Hospital of New Orleans, New Orleans, LA

---

Alveolar rhabdomyosarcoma (ARMS) is generally characterized by one of three translocation states: the presence of t(2;13)(p35;p14) leading to the production of the PAX3-FOXO1 fusion protein, the presence of t(1;13)(p35;q14) leading to the production of the PAX7-FOXO1 fusion protein, or a translocation-negative status. The presence of the t(2;13) is associated with greater disease severity and mortality than patients who are t(2;13) translocation negative or t(1;13) positive. To date, no clinical diagnostic assay can simultaneously differentiate between these three translocation states. The specific aim of this study is to investigate the disease severity and survival outcomes of tumor specimens obtained from ARMS patients with respect to translocation status using fluorescence *in situ* hybridization (FISH). We have designed and validated FISH probes in our laboratory that can simultaneously and distinctively identify t(2;13) and t(1;13) translocations. Our data indicates that heterogeneity of the (2;13) translocation is correlated to severity of the individual patients, in which samples with more than one t(2;13) per cell (aneuploidy) and greater amounts of t(2;13) positive cells correlate with a poorer prognosis with respect to failure-free survival. These results thereby provide a new clinical assay to differentiate between three translocation types and improve diagnostic and prognostic indicators for these tumors.

## PgmNr 2412: Optimized early cancer detection from whole genome sequencing of cell-free DNA.

### Authors:

C. Melton; P. Freese; S. Bagaria; E. Hubbell; D. Filippova; R. Calef; M.C. Maher; V. Nicula; S. Gross; O. Venn; A. Valouev; A. Aravanis; A. Jamshidi

[View Session](#) [Add to Schedule](#)

**Affiliation:** GRAIL, Menlo Park, CA

---

Cancer can be detected through whole genome sequencing (WGS) of cell-free DNA (cfDNA). The most commonly derived set of features from WGS data, fragment counts in bins tiled across the genome, enable detection of somatic copy number alterations (CNAs); however, other features can be leveraged to classify cancer status. In the first substudy of Circulating Cell-free Genome Atlas (CCGA; NCT02889978) we evaluated three prototype assays for discriminating cancer vs non-cancer; we determined that fragment methylation-based features from whole genome bisulfite sequencing had the best sensitivity (at high specificity), followed by small variants detected from targeted sequencing and somatic CNAs from WGS read depth. An exploratory post hoc analysis of the first substudy evaluated the impact of alternative features derived from WGS data on classification performance. We assessed the utility of counting fragments originating at cancer-enriched fragment endpoint positions and explored alternative means of detecting CNAs by observing (a) allelic imbalance at heterozygous SNPs and (b) changes in fragment length distributions in tiled bins. Optimization of features and learning algorithms were performed in 10-fold cross-validation on a training set (561 controls, 863 cancer participants [pts], 20 solid tumor types) and final assessment performed on an independent test set (362 controls, 464 cancer pts, 20 solid tumor types). At 98% specificity, fragment endpoints had 18.1% (95% CI, 14.7-21.9) sensitivity vs 34.1% (29.7-38.6) sensitivity for methylation and 29.3% (25.2-33.7) for binned counts. 29.1% sensitivity (25.0-33.5) for fragment lengths and 21.8% (18.1-25.8) for allelic imbalance was also observed. All classifiers showed strong stage and tumor fraction dependence. An ensemble WGS classifier combining endpoints, allelic ratios, fragment lengths, and binned counts outperformed binned counts alone (+2.2% sensitivity,  $p=0.044$ ) but had lower sensitivity than the methylation-based classifier (-2.6% sensitivity,  $p=0.045$ ). WGS and methylation features combined failed to improve upon the methylation classifier alone (-0.6% sensitivity,  $p=0.546$ ). Overall, combining WGS features improved sensitivity at high specificity in the absence of methylation features but was inferior to methylation-based classification; these results, in part, motivated development of an improved methylation assay assessed in the second substudy of CCGA.<sup>1</sup>

1. J Clin Oncol 37, 2019 (suppl; abstr 3049).

## PgmNr 2413: Loss of Y chromosome: Non-invasive biomarkers for early detection of colorectal and prostate cancers in males.

### Authors:

A. Asim <sup>1</sup>; S. Agarwal <sup>1</sup>; R. Kapoor <sup>2</sup>; N. Rastogi <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India, Lucknow, Uttar Pradesh, India; 2) Department of Urology, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India, Lucknow, Uttar Pradesh, India; 3) Department of Radiotherapy, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India, Lucknow, Uttar Pradesh, India

---

### Background

Loss of Y chromosome (LOY) is a well known established phenomenon associated with cancers and ageing. Recently, LOY in peripheral blood cells was suggested as a possible biomarker for different cancers in males

### Aims and objectives:

On the basis of previous findings, the present case-control study was conducted to evaluate the association of LOY in peripheral blood cells in prostate (PC) and colorectal cancers (CRC) in males.

### Methodology:

30 CRC patients (mean age =  $44.03 \pm 10.8$ ), 36 PC patients (mean age =  $60.8 \pm 15.8$  yrs) and 36 healthy control male cases (mean age =  $54.6 \pm 15.1$  years) were recruited. DNA was extracted using a standard phenol-chloroform method. Multiplex quantitative fluorescent (QF) PCR was used to co-amplify the homologous sequences present on the Y chromosome and other chromosome followed by their analysis on the genetic analyzer (ABI 3500) and finally the Y/X ratio was calculated on the basis of the peak height obtained from the electropherogram.

### Results:

The mean Y/X ratio was significantly lower in the whole group of cancer patients ( $0.709 \pm 0.02$ ;  $p < 0.0001$ ) when compared to the controls ( $0.92 \pm 0.044$ ). Also, the Y/X ratio when calculated separately was found to be lower in CRC ( $0.701 \pm 0.078$ ;  $p < 0.0001$ ) and PC ( $0.717 \pm 0.044$ ;  $p < 0.0001$ ) cases, when compared to controls ( $0.92 \pm 0.044$ ). Multivariate logistic regression was performed by matching cancer and control subjects with age and the results suggest that LOY is not influenced by their age.

### Conclusion

The results support the significant association of LOY in peripheral blood cells carcinogenesis in males. LOY can also serve as a non-invasive cancer biomarker to improve the early diagnosis and management of cancer patients in males

## PgmNr 2414: Urinary *TERT* promoter mutations as non-invasive biomarkers for the comprehensive detection of urothelial cancer.

### Authors:

P.H. Avogbe <sup>1</sup>; A. Manel <sup>2</sup>; E. Vian <sup>2</sup>; G. Durand <sup>1</sup>; N. Forey <sup>1</sup>; V. Voegelé <sup>1</sup>; M. Zvereva <sup>1,3</sup>; Md.I. Hosen <sup>1</sup>; S. Mezzani <sup>1</sup>; B. De Tilly <sup>2</sup>; G. Polo <sup>2</sup>; O. Lole <sup>1</sup>; P. Francois <sup>1</sup>; T.M. Delhomme <sup>1</sup>; C. Carreira <sup>1</sup>; S. Monteiro-Reis <sup>4</sup>; R. Henrique <sup>4,5,6</sup>; B. Abedi-Ardekani <sup>1</sup>; G. Byrnes <sup>1</sup>; M. Foll <sup>1</sup>; E. Weiderpass <sup>1</sup>; J. McKay <sup>1</sup>; C. Jeronimo <sup>4,5,6</sup>; G. Scelo <sup>1</sup>; F. Le Calvez-Kelm <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) International Agency for Research on Cancer (IARC), Lyon, France; 2) Protestant Clinic of Lyon, Urology department, Lyon, France; 3) Faculty of Chemistry, Lomonosov Moscow State University, Moscow, Russia; 4) Portuguese Oncology Institute of Porto, Research Center (CI-IPOP), Porto, Portugal; 5) Portuguese Oncology Institute of Porto (IPOP), Department of Pathology, Porto, Portugal; 6) Institute of Biomedical Sciences Abel Salazar, University of Porto (ICBAS-UP), Porto, Portugal

**Background:** Recurrent mutations in the promoter of the telomerase reverse transcriptase (*TERT*) gene (C228T and C250T) detected in tumours and cells shed into urine of urothelial cancer (UC) patients are putative biomarkers for UC detection and monitoring. However, the possibility of detecting these mutations in cell-free circulating DNA (cfDNA) in blood and urine, or DNA from urinary exfoliated cells (cellDNA) with a single-gene sensitive assay has never been tested in a case-control setting.

**Methods:** We developed a single-plex assay (UroMuTERT) for the detection of low-abundance *TERT* promoter mutations. We tested 93 primary and recurrent UC cases and 94 controls recruited in France (blood, urine samples and tumours for the cases), and 50 primary UC cases and 50 controls recruited in Portugal (urinary exfoliated cell samples). We compared our assay with urine cytology.

**Results:** In the French series, C228T or C250T were detected in urinary cfDNA or cellDNA in 81 cases (87.1%; 95% CI 78.6–93.2), and five controls (Specificity 94.7%; 95% CI 88.0–98.3), with 98.6% (95% CI 92.5–99.96) concordance in matched tumours. Detection rate in plasma cfDNA among cases was 7.1%. The UroMuTERT sensitivity was (i) highest for urinary cfDNA and cellDNA combined, (ii) consistent across primary and recurrent cases, tumour stages and grades, (iii) higher for low-risk non-muscle invasive UC (86.1%) than urine cytology (23.0%) ( $P < 0.0001$ ) and (iv) 93.9% when combined with cytology. In the Portuguese series – the sensitivity and specificity for detection of UC with urinary cellDNA was 68.0% (95% CI 53.3–80.5) and 98.0% (95% CI 89.3–100.0).

**Conclusion:** *TERT* promoter mutations detected by the UroMuTERT assay in urinary DNA (cfDNA or cellDNA) show excellent sensitivity and specificity for the detection of UC, significantly outperforming that of urine cytology notably for detection of low-grade early stages UC.



## PgmNr 2415: Distinctive expression of specific microRNAs in breast cancer patients and change during chemotherapy.

### Authors:

Y. Balcha <sup>1</sup>; D. Seifu <sup>3</sup>; A. Bekele <sup>2</sup>; W. Tighehe <sup>3</sup>; R. Kittles <sup>4</sup>; U. Menon <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Sciences, Arsi University, Arsi, Ethiopia; 2) Dean of Health Sciences University of Global Health Equity, Kigali, Rwanda; 3) Department of Biochemistry, Addis Ababa University, Ethiopia; 4) Associate Director of Health Equities, Comprehensive Cancer Center City of Hope, California, U.S.A; 5) Vice Dean of Research University of South Florida U.S.A

---

Breast cancer is the second most common cause of death from cancer in women. There is now increased prevalence of breast cancer among African women. In Ethiopia it is the leading cancer type constituting 33% of all cancer cases in women.

Tests and procedures for diagnosis and prognosis of breast cancer are still limited to invasive procedures and imaging techniques. Based on their differential expression in disease and with their exceptional stability in biological fluids, microRNAs are noticeable candidates to be used as non-invasive diagnostic and prognostic biomarker. The association of change in expression level of microRNAs with clinicopathological parameters may demonstrate their potential in sorting the heterogeneous disease to specific subgroups for effective treatment options.

The purpose of this study was to investigate the differential expression of specific circulatory microRNAs in breast cancer patients and study their association with clinicopathological parameters with samples from healthy volunteers serving as control. Changes in the level of microRNA after completing chemotherapy was also assessed and checked for association with serum CA 15-3 values to see the prognostic potential of specific microRNAs.

The levels of specific circulatory microRNAs in patient and control samples were analyzed using qRT-PCR. Serum CA 15-3 values from patient samples collected after completion of chemotherapy were measured using solid phase enzyme linked immunosorbent assay.

In this study miRNA 21-5p was significantly over expressed in the serum of breast cancer patients when compared with controls ( $P < 0.05$ ). Receiver operating characteristic curve analysis shows its potential in differentiating breast cancer patients from controls with AUC value of 0.6 ( $P < 0.05$ ). A significant decrease in serum miR-326 was observed after chemotherapy ( $P < 0.05$ ) as compared to serum samples collected before chemotherapy. No correlations was found between serum CA 15-3 values and levels of circulatory microRNAs. Mean serum CA 15-3 value of patients with stage IV breast cancer was found to be high when compared with patients with other stages ( $P < 0.001$ ). Since deregulated miRNAs in the circulation are shared by several cancer types and subtypes, further studies are necessary to identify a well characterized cluster of miRNAs with discriminative ability.



## PgmNr 2416: Performance of a novel GeneReader NGS System through hotspot panels: An analysis across specimen type and tumor type over 2000 patients.

### Authors:

A. Bisgin <sup>1,2</sup>; I. Boga <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cukurova University AGENTEM (Adana Genetic Diseases Diagnosis and Treatment Center), Adana, Turkey; 2) Department of Medical Genetics, Cukurova University Faculty of Medicine, Adana, Turkey

---

Novel NGS systems for somatic mutation identification need to be evaluated by assessing ease of use, reproducibility, implementation and validation in clinical samples. To gain insight into the issues, we investigated the performance of GeneReader via the reasons for and frequency of repeats in hotspot multigene panel tests for genomic DNA derived from formalin fixed paraffin embedded (FFPE) tissue and liquid biopsy (LB) across various cancer types of more than 2000 patients.

Total of 2044 patients were next generation sequenced which was modified, optimized and validated in our center and repeat rates were analyzed. The two hotspot panels; Qiagen ATP including *ALK, KIT, BRAF, NRAS, KRAS, EGFR, ERBB2, ERBB3, PDGFRA, RAF1, PIK3CA, ESR1* genes and 19 gene panel including *AKT1, DDR2, MAP2K1, NTRK1, PTEN, ROS1, RICTOR, FGFR1, MET* genes in addition to ATP were used. Samples with low amount of extracted DNA were not triaged. Samples run for validation purposes were excluded from the analysis. The numbers of repeated specimens were compared across specimen, cancer and test types. Quality controls and data processing were performed via QCI-Analyze and QCI-Interpret by an experienced user.

Among 2044 samples, 1051 (51.5%) of them were FFPE tissue samples and 993 (48.5%) were LB. 50.6% of all samples were sequenced with ATP panel while 1009 (49.4%) samples were with 19 gene panel. Of all the samples that were analyzed, 152 were repeated (7.5%). The repeat rates were 6.1% for LB and 8.8% for FFPE. The distribution of specimen types among test panels and repeat rates were: a) 27.9% of FFPE samples were sequenced with 19 gene panel with 11.3% repeat rate, b) 72.1% of FFPE tissue were sequenced with ATP panel and 7.8% was re-sequenced, c) 72.1% of LB samples were sequenced with 19 gene panel with 6% repeat rate and d) 27.9% of LB were sequenced with ATP panel and 6.5% was re-sequenced.

The focus of this study was to assess the suitability of GeneReader NGS System for routine use. Thus, we documented an overall 7.5% test repeat rate which showed as low as 6.1% for LB while there are no significant differences in repeat rate across cancer and test panel types. As a conclusion, while the repeat rates suggest room for improvement, more than 90% sequencing success rate with an easy-to-operate system is deemed acceptable. Investigation of other variables in starting DNA quality and sequencing library preparation chemistry may facilitate further improvements of repeat rate.

## PgmNr 2417: Sole trisomy 6 in childhood acute myeloblastic leukemia, uncommon finding associated to bad prognosis.

### Authors:

S.A. Brukman <sup>1</sup>; R. Corona-Rivera <sup>1,2</sup>; L. Bobadilla-Morales <sup>1,2</sup>; L. Mendoza-Maldonado <sup>1</sup>; C. Ortega de-la-Torre <sup>2</sup>; E. Torres-Anguiano <sup>1</sup>; U.F. Santana-Bejarano <sup>1</sup>; A. Corona-Rivera <sup>1</sup>; P.A. Chávez-Panduro <sup>2</sup>; F.A. Sánchez-Zubieta <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genética Humana, Universidad de Guadalajara , Guadalajara , Jalisco, Mexico; 2) Onco-hematología, Hospital Civil Nuevo #34;Juan I. Menchaca", Guadalajara, México

---

Sole trisomy is a frequent cytogenetic aberration in malignant neoplasms, but rare in acute myeloblastic leukemia, mainly chromosome 8 trisomy, followed by chromosomes 4, 9, 11, 13 and 21. Here, we report a 14-year-old male patient, product from the third pregnancy, dizygotic, obtained by abdominal delivery, Weight: 3350gr. Stature 52cm is admitted for an 8 day evolution with cervical lymphadenopathy, along with intermittent fever, polydipsia, emesis and hyporexia. He started chemotherapy, receiving Cytarabine (100mg/m<sup>2</sup>/dose q12h), Daunorubicin (50mg/m<sup>2</sup>/dose q48h) and Etoposide (100mg/m<sup>2</sup>/dose q24h). At day 2, an increase in the leukocyte count is reported: 500,000 /  $\mu$ L with a monocytes predominance of 75%; leukophoresis is indicated and induction is initiated with Cytarabine. He died on day 3, with hyperleukocytosis of 728,000  $\mu$  / L, acute renal failure and pulmonary leukostasis. The bone marrow aspirate showed hypercellularity, red and megakaryocytic series absent, increased white series with a predominance of monocytic series and myeloblasts; Immunophenotype reported a myeloblastic population of 72% % expressing CD13 +, CD33 +, HLA-DR +, CD38 +, CD117 +, and a second monocytic population in 20% expressing CD14 +, CD33 +, CD36 +, CD64 +, CD38 and CD45 +. All of the above were compatible with myelomonocytic AML M4 by FAB. Bone marrow aspirate was: 47,XX,+6[12]/46,XX[3]; By FISH: nuc ish(D6Z1)x3[143]/(D6Z1)x2[57], and the absence of 16q22 chromosome rearrangements. qPCR: negative for inv16(p13.1q22). The FLT3/ITD mutation was positive. Nested multiplex RT-PCR (HemaVision®- HV01-28 N, DNA Technology A/S) to the search of 28 common AML and ALL fusion transcripts was negative. To search the prognosis, we performed an OS analysis of 29 previously cases reported with trisomy 6. Our statistical evaluation using all the available OS previously reported values from the literature of AML patients with sole trisomy 6 against a comparative group of patients from our hospital with different types of AML but normal karyotype, supports the hypothesis that trisomy 6 in AML could be related to bad prognosis in patients with this unusual cytogenetic finding.

## PgmNr 2419: Detection of fusion genes for minimal residual disease in acute leukemia using next-generation sequencing.

### Authors:

E.N. de Boer <sup>1,2</sup>; K. de Lange <sup>1,2</sup>; L.F. Johansson <sup>1,2</sup>; A.G. Bosga-Bouwer <sup>1,2</sup>; E. van de Berg <sup>1,2</sup>; B. Sikkema-Raddatz <sup>1,2</sup>; C.C. van Diemen <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, University Medical Centre Groningen, Groningen, The Netherlands; 2) University of Groningen, The Netherlands

Acute leukemia can be caused by chromosomal rearrangements and a considerable number of different fusion genes have been identified. The type of fusion gene determines the type of leukemia and treatment choices. Measuring minimal residual disease (MRD), defined as persistence of leukemic cells after treatment, is important to monitor disease recurrence and is currently performed using flow cytometry and digital droplet PCR (ddPCR). Next generation sequencing (NGS) may have advantages, like the possibility of multiplexing different fusion genes and patients in one experiment, but has scarcely been investigated.

We developed an NGS method to detect low levels of plasmid DNA against a background of genomic DNA [1]. Here, we tested if an RNA-based version of this method has the required sensitivity (i.e. 0.001% cells [2]) to detect fusion transcripts for MRD.

We designed a multiplex assay targeting 13 fusion genes containing: a) labeled probes targeting the breakpoints of each fusion gene; b) non-labeled probes blocking wild-type transcripts and c) labeled probes fully tiling housekeeping genes for quantification.

RNA from 12 bone marrow patient samples and two cell lines were diluted in different percentages using wild-type control blood RNA. The RNA dilutions were rRNA and globin mRNA depleted and converted to cDNA. Sequence-adaptors containing unique molecular identifiers were added to the cDNA fragments. After capturing, libraries were sequenced on an Illumina MiSeq. Unique reads were directly aligned to fusion gene-specific reference transcripts.

First results show a sensitivity of at least 0.1% RNA copies. In further experiments, the sensitivity will be further improved by increasing the amount of input material resulting in more unique reads per sample. Currently, ddPCR is performed on the same samples for comparison of the sensitivity and to convert % RNA copies to % cells.

Concluding, we have designed an RNA based method for MRD detection of fusion genes with promising results. Expanding the panel with other fusion genes will not influence the performance of the probes already in the panel and allows the analysis of multiple samples in one experiment without the need for optimization.

[1]: E.N. de Boer, et al. A next-generation sequencing method for gene doping detection that distinguishes low levels of plasmid DNA against a background of genomic DNA. Accepted. Gene Therapy.

[2]: O. Landgren, et al. 2019 Advances in Cell and Gene Therapy:2(1).

## PgmNr 2420: Circulating tumor DNA methylation haplotypes in plasma can accurately detect early-stage colorectal cancer.

### Authors:

A. Gore <sup>1</sup>; J. Gole <sup>1</sup>; J. Min <sup>1</sup>; J. Dang <sup>1</sup>; H. Shi <sup>1</sup>; Y. Gao <sup>1</sup>; K. Zhang <sup>2</sup>; R. Liu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Singlera Genomics, La Jolla, CA; 2) University of California - San Diego, La Jolla, CA

---

**Background:** Colorectal cancer (CRC) is one of the most common forms of cancer and is responsible for approximately 700,000 deaths per year worldwide. Because of this, colorectal cancer screening is recommended by the USPTF, but patient compliance can be low due to the invasiveness of screening colonoscopy. We had previously described ColonES, a highly accurate, non-invasive plasma-based screening assay to identify colorectal cancer at an early stage using ctDNA methylation haplotypes. Here we present an additional study utilizing the ColonES assay in an independent cohort of samples, further demonstrating highly accurate detection of early-stage colorectal cancer. **Methods:** Blood samples from 64 healthy individuals and 66 individuals diagnosed with colorectal cancer by colonoscopy were collected in EDTA tubes and immediately separated into plasma. Plasma samples were then processed using the Singlera Genomics ColonES assay, a targeted bisulfite sequencing method which identifies methylation haplotype patterns specific to early- and late-stage colorectal cancer. **Results:** In this independent test set, the ColonES assay was able to show a sensitivity of 89% in colorectal cancer patients with a specificity of 97% in healthy patients. **Conclusions:** We have shown that ctDNA methylation can be utilized to non-invasively screen for early-stage colorectal cancer with high sensitivity in two independent sample sets, paving the way for a primary blood-based colorectal cancer screening assay.

## **PgmNr 2421: NSCLC-UHS: Detection of hotspot mutations of non-small cell lung cancer patient with plasma samples. Ultra-high-sensitive mutation detection system.**

### **Authors:**

J. Lee <sup>3</sup>; W. Park <sup>1</sup>; S. Jang <sup>2</sup>; S. Chun <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Asan Institute for Life Sciences, Asan Medical Center, Seoul, Select a Country; 2) Department of Pathology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea.; 3) Department of Medical Science, Asan Medical Institute of Convergence Science and Technology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea.

---

Circulating cell-free DNA (cfDNA) has great potential in clinical filed including care of cancer patients. Previously, prognostic and predictive value of cfDNA in non-small cell lung cancer (NSCLC) has been reported by several studies. Genetic mutations of EGFR, KRAS and BRAF genes in tumor derived cfDNA in patient with lung cancer could be applied as biomarkers for monitoring the early steps of tumor progression or a recurrence. However, low amount and extremely low tumor-derived DNA rate are recognized as hurdles to be overcome of cfDNA application.

In this study, we developed NSCLC-UHS (Ultra High-Sensitive) panel, which targets hot-spot mutations of BRAF, KRAS and EGFR in multiplex fashion using low amount cfDNA. Using the serially diluted positive samples, detection sensitivity of this panel was evaluated to be 0.05%. We performed targeted NGS analysis for detection of genetic mutations using FFPE tumor tissue DNA and matched cfDNA extracted from 0.4 ml plasma of 30 NSCLC patients, then, NSCLC-UHS panel test was done using cfDNA to compare the mutation patterns of EGFR, KRAS and BRAF genes. Focusing EGFR mutations, 20 of 30 cfDNA from NSCLC patients (66%) showed same mutation pattern with those of tumor tissue DNA by targeted NGS analysis. When considering the amount of applied cfDNA was very low from 0.1 ng to 1 ng, 66% detection sensitivity of NSCLC-UHS assay is very impressive. We expect that the sensitivity will be dramatically improved by increasing the input amount of cfDNA with higher starting volume of plasma.

In conclusion, Onco-UHS method is fast and cost-effective methods for a targeted set of mutations and can be used for selection of chemotherapy and for validation of NGS.

## PgmNr 2422: Digital PCR partitioning with near zero dead-volume for liquid biopsy applications.

**Authors:**

R... Lin; M. Dueck; A. Zayak; S. Gallagher; L. Jiang; P. Hung

[View Session](#) [Add to Schedule](#)

**Affiliation:** Combinati, Burlingame, CA

---

Digital PCR (dPCR) has emerged as a promising technique to both detect and precisely quantify rare targets. This has made dPCR particularly attractive for monitoring rare mutations in liquid biopsy. Partitioning of a sample into large number of independent reactions is fundamental to dPCR. Although microfluidics has greatly aided the partition process, currently available partitioning technologies suffer from limitations in 3 critical aspects – partition volume consistency, partition number consistency and sample waste.

Variable volume results in variable distribution of targets inside the partitions and undermines the mathematical basis of dPCR analysis. Fluctuating partition numbers underscores the instability of the partitioning process and introduces unnecessary variables into an assay. Wasted sample or “dead volume” introduces subsampling error and reduces accuracy of the result. Subsampling error is a function of both the expected target concentration and degree of subsampling. The combination of rare targets with large dead volume greatly diminishes dPCR accuracy.

Here we present a novel Microfluidic Array Partitioning (MAP) technique that utilizes injection molded arrays and fluidic control to overcome these limitations. Injection molding ensures all partitions have a consistent and defined volume regardless of inputted sample. Partition size is defined not by a stochastic process but by the physical limitation of the microwells. The partition array is of fixed geometry and ensures consistent number of partitions for all samples across experiments. The number of partitions is defined by the array geometry and is minimally susceptible to variation in the partitioning process. Enabled by a novel loading process, the MAP consumable utilizes >95% of inputted sample and significantly reduces subsampling error compared to other technologies that often partition less than three quarters of loaded sample.

To demonstrate the utility of the technique, we adapted a commercially available assay and performed dPCR quantitation using KRAS G12D mutation containing plasmids in a background of WT human genomic DNA. Using MAP devices, we were able to consistently quantify 10%, 1% and 0.1% mutant to wild-type DNA with consistent partition volume, minimal partition number variation and greater than 95% utilization of inputted sample volume. Taken together, the new technique provides a promising path towards bringing dPCR into the clinic for liquid biopsy applications.

## PgmNr 2423: Molecular cytogenetic characterization of a complex karyotype of a pediatric male patient with B-acute lymphoblastic leukemia.

### Authors:

A. Nguyen <sup>1</sup>; V. Tse <sup>2</sup>; K. Lapp <sup>3</sup>; G. Yang <sup>2</sup>; E. Peng <sup>2</sup>; W. Wyatt <sup>2</sup>; K. Cunnien <sup>3</sup>; D. Serk <sup>3</sup>; K. Stieglbauer <sup>3</sup>; C.A. Tirado <sup>1,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ecology and Evolutionary Biology, University of California, Los Angeles, Villa Park, California.; 2) The International Circle of Genetic Studies, Chapter Los Angeles, USA; 3) Allina Health, Inc., Minneapolis, MN; 4) HPA, Minneapolis, MN; 5) The University of Minnesota, Department of Laboratory Medicine and Pathology, Minneapolis, MN

An eight year old male patient was diagnosed with B-ALL. Chromosome studies of twenty G-banded metaphases of the bone marrow detected two abnormal clones with loss of 9p [i(9)(q10)] and loss of 17p [**der(17)(?:17q11.2->17p11.2::17p11.2->17qter)**] within the context of a complex karyotype in 8 metaphase cells. Four of these abnormal metaphases detected in addition additional material of unknown origin on 12 at p11.2 [add(12)(p11.2)]. The remaining 12 metaphases are cytogenetically normal.

FISH analysis of metaphases from a destained G-banded slide probed with KMT2A reveal an intact copy of KMT2A on the der(?). Metaphases were detected on the CDKN2A/CEP9, BCR/ABL1 and TP53/CEP17 interphase studies and revealed loss of CDKN2A on the add(9) and the i(9), 2 copies of ABL1 on the i(9), and loss of TP53 with gain of CEP 17 on the der(17). This complex karyotype was described as: 45,XY,+1,add(1)(q11),dic(1;8)(p11;p21),add(2)(p11.2),del(6)(p23p25),add(9)(q13),i(9)(q10),-11,-14,-16,der(17)(?:17q11.2->17p11.2::17p11.2->17qter),+der(?)(?:11q13->11qter),+mar[4].ish add(9)(CDKN2A-,CEP 9+),i(9)(q10)(CDKN2A-,CEP 9+,ABL1++),der(?)(KMT2A+),der(17)(TP53-,D17Z1+,D17Z1 dim+).nuc ish(CEP4x2,CEP10x2,D17Z1x2,D17Z1 dimx1)[35/200],(CDKN2Ax0,CEP9x2)[94/200],(ABL1,BCR)x2[500],(5'KMT2A,3'KMT2A)x2(5'KMT2A con 3'KMT2Ax2)[200],(ETV6,RUNX1)x2[200],(TP53x1,D17Z1x2,D17Z1 dimx1)[143/200]/45,idem,add(12)(p11.2)[4]/46,XY[12]

Deletion of 9p involving the CDKN2A locus (9p21) is seen in 7-11% of pediatric B-ALL. Although del(9p) does not appear to have an impact on prognosis in pediatric B-ALL, there is an overall poorer outcome compared with those lacking this abnormality. Abnormalities of 9p identify a subgroup of NCI standard-risk patients with increased risk of treatment failure.

The presence of the derivative chromosome (17), leading to loss of TP53, exerts an adverse influence on treatment outcome in B-ALL. Rearrangements of 12p [add(12)(p11.2)] usually occur as part of a complex karyotype (Heerema 2000). Complex karyotypes reflect genomic instability and are usually associated with poor prognosis (Heim and Mitelman 2015). Molecular cytogenetic characterization of complex karyotypes in B-ALL is an excellent tool to detect chromosomal abnormalities of prognostic

significance in B-ALL.

Clinical pathologic correlation of these results is recommended.



## PgmNr 2424: Common variants in PMS2CL that can present in *PMS2* as pathogenic variants with extremely low frequencies.

### Authors:

S. Pan; A. Brown; B. Leclair; M. Elias; D. Chen; J. Kidd; K. Bowles; B. Coffee; B. Roa; D. Mancini-DiNardo

[View Session](#) [Add to Schedule](#)

**Affiliation:** Myriad Genetic Laboratories, Inc., Salt Lake City, UT

---

**Background:** *PMS2*, a Lynch syndrome-associated DNA mismatch-repair gene, often is included in NGS hereditary pan-cancer panels. Molecular testing of *PMS2* is complicated by the interference of highly homologous pseudogenes. The most homologous pseudogene, *PMS2CL*, is >98% identical to *PMS2* exons 11-15. Therefore, additional analysis is required for variants identified in exons 11-15 to determine whether they are located in *PMS2* or *PMS2CL*. Correct allocation of variants identified by NGS in this region is critical for proper clinical management.

**Objective:** The purpose of this analysis was to evaluate the frequency with which common *PMS2CL* variants occur in *PMS2*, with a focus on those that are considered pathogenic when they occur in *PMS2*.

**Methods:** Pathogenic/likely pathogenic sequence variants (PV/LPVs) detected from July 2016-April 2019 and large rearrangements (LRs) detected from May 2017-April 2019 in the *PMS2* pseudogene region (exons 11-15) were evaluated in individuals tested using an NGS hereditary pan-cancer panel. These variants were initially identified by NGS, then confirmed by orthogonal assays as being present in either *PMS2* or *PMS2CL*. PV/LPVs were selected for this analysis if they had >100 observations upon NGS testing but were confirmed to be in *PMS2* in <1% of cases.

**Results:** Two sequence variants and three LRs were assessed (four variants in exons 13 and 14; one in exon 15). Collectively, these variants were detected in 12,217 individuals. In 99.91% (12,206/12,217) of cases, the variants were confirmed orthogonally to be present in *PMS2CL*. In 11 (0.09%) individuals, the variant was located in *PMS2*. The rarest PV in *PMS2* was c.2186\_2187del (p.Leu729Glnfs\*6); only one (<0.01%) patient was confirmed to carry this PV in *PMS2*.

**Conclusion:** Comprehensive testing and a large testing population enabled identification of PV/LPVs that are predominantly present in *PMS2CL* but can occur in *PMS2* with extremely low frequency. These data highlight the need to disambiguate PVs in *PMS2* versus *PMS2CL*. It is tempting to assume that when certain sequencing variants and LRs are known to occur in *PMS2CL* >99% of the time, additional analysis is unnecessary. We have demonstrated that, though rare, these variants can occur in *PMS2*. Failure to confirm the PV/LPV location can produce a false negative result with significant implications for clinical management.

## PgmNr 2425: Germ line mutation analysis of the *RB1* gene in retinoblastoma patients in Sri Lanka.

### Authors:

N. Panchanathan <sup>1</sup>; D.R. Lohmann <sup>2</sup>; D.C.de Silva <sup>3</sup>; H. Abeysekara <sup>4</sup>; D.P.S Nanayakara <sup>5</sup>; T.L.S Tirimanne <sup>6</sup>; D. Irugalbandara <sup>4</sup>; N.V. Chandrasekharan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Chemistry, University of Colombo, Colombo 3, Sri Lanka; 2) Institut für Humangenetik, Universitätsklinikum Essen, Germany; 3) Department of Physiology, University of Kelaniya, Sri Lanka; 4) Lady Ridgway Hospital, Colombo, Sri Lanka; 5) National Eye Hospital, Colombo, Sri Lanka; 6) Department of Plant Sciences, University of Colombo, Sri Lanka

---

**Introduction:** Retinoblastoma (RB) a tumour affecting those <5y has a prevalence of 1 in 20,000 with twenty cases/ year predicted in Sri Lanka. Unilateral RB (60%) present on average at 24 months, bilateral RB (40%) presents around 15 months. Family history reported in 10%. Many cases, diagnosed late in Sri Lanka, require enucleation. Genetic testing has been unavailable locally, but may enable better targeting of screening for patients and their siblings and reduce need for enucleation in affected cases.

**Materials and methods:** NGS sequencing of the *RB1* gene performed for 50 DNA samples from clinically diagnosed RB cases (bilateral (n=19(38%)), unilateral n=31) including two familial cases. NGS data were analysed using established bioinformatic tools (FastQC, BOWTIE 2, samtools, bcftools and vcftools software). Patients recruited following ethical clearance and informed consent.

**Results:** Pathogenic variants (2 missense, 7 stop gained, 1 splice donor, 8 frame shift variant) identified in 18/50 (34%) cases including 12/19(63%) bilateral cases and 6/31(19%) unilateral cases. Six were previously undescribed, likely pathogenic frameshift variants.

**Conclusion:** Germline mutations have been identified in 10-15% of unilateral, 80% bilateral and all familial cases worldwide. Detection rate of germline point mutation (36%) similar to worldwide data (40%). Germline mutation not identified in one familial case. Evaluation of NGS data for chromosomal rearrangements, which account for 10-15% of reported germline mutations, need to be identified in this sample.

**Acknowledgement:** We acknowledge the University of Colombo for financial support under the university of Colombo research grant AP/3/2/2016/CG/27

## PgmNr 2426: Breast-UHS: Metastasis tracking system of breast cancer using circulating cell-free DNA.

### Authors:

H. Park <sup>1</sup>; J. Lee <sup>3</sup>; H. Cho <sup>1</sup>; J. Kim <sup>4</sup>; S. Chun <sup>2</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Asan Institute for Life Sciences Asan Medical center, Seoul, Korea, Republic of; 2) Department of Pathology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea; 3) Department of Medical Science, Asan Medical Institute of Convergence Science and Technology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea; 4) Department of Breast Surgery, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

---

Despite progress in the management of early breast cancer, a substantial number of patients relapse with metastatic breast cancer. Tumor derived mutations can be detected in circulating cell free DNA (cfDNA) of patients with metastatic breast cancer. We assessed whether analysis of cfDNA can be used to predict which patients will relapse following treatment of primary breast cancer.

First of all, we found anchor mutations that are expected to play a role as cancer drivers through NGS analysis of breast cancer tissues and then tracked them every three months from patient-derived cfDNA using a technique called UHS, which we previously developed a novel approach for high-throughput, multiplexed, ultrasensitive mutation detection and used it for detection of mutant sequence mixtures as low as 0.1-0.5% minor allele frequency.

Through this platform, it was confirmed that anchor mutations such as *CDH1*, *TP53*, *NOTCH2*, *NOCH4*, and *MAPK3* found in tumor tissues were also found in cfDNA from 5 patients with metastatic breast cancer of 11 breast patients group. In our study, anchor mutations were mainly found in the pre-anticancer stage of the group with metastatic cancer, and these mutations were not found in the cfDNA obtained from the plasma through the subsequent treatment process. The same method has been applied to a group of 200 patients with primary cancer to track anchor mutation with cfDNA to construct a system to predict metastatic cancer.

So, we have conducted a long-term follow-up study to access the hypothesis that monitoring of tumor-specific mutations in cfDNA can detect metastatic disease following primary surgery and serve as a sensitive, specific, and thus potentially clinically useful noninvasive biomarker in the breast cancer.

## **PgmNr 2427: Harmonizing tumor sequencing with germline genetic testing: Identification of at-risk individuals for hereditary cancer disorders.**

### **Authors:**

D.E. Pineda-Alvarez; S.T. Michalski; M. Russell; S. Yang; I.Y. Song; E.D. Esplin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Invitae, 1400 16th Street, San Francisco, CA, 94103

---

### **Background:**

Germline and tumor genetic testing are both increasingly used in precision treatment of individuals with cancer. These tests are frequently ordered separately in clinical practice but have the potential to inform one another. Recent studies report that approximately 10% of patient's tumors have clinically significant variants in genes known to predispose to hereditary cancer, however, it has been unclear which patients or findings deserve follow-up with germline testing (GT). Here, we retrospectively reviewed a cohort of individuals who had tumor testing followed by GT. We report the concordance rate between germline and tumor variants, and the impact on the patient's medical management and implications for their relatives.

### **Methods:**

Our study used de-identified data from 1043 consecutive patients who underwent tumor genetic testing followed by GT with NGS-based hereditary cancer gene panels.

### **Results:**

Tumor genetic testing results included variants in BRCA2 (290), BRCA1 (174), TP53 (158), ATM (70), MLH1 (65), APC (65), PMS2 (61), MSH6 (58), PTEN (54) and CDH1 (42). In 364/1043 cases (35%) the variant was detected as likely pathogenic or pathogenic (LP/P) in the germline. Genes in which variants were confirmed to be germline in >60% of patients included: FANCA, AXIN2, RAD50, MUTYH, BLM, PALB2, CHEK2, FANCD2, MITF, SDHB. Variants in FH, BRCA2, RET, ATM, SDHA, BRIP1, MSH2, BRCA1, BAP1, EGFR, and RAD51D were confirmed in the germline in <60% of patients. Variants were rarely detected as germline for TP53 (3%), APC (3%), PTEN (2%) and none were detected for CDKN2A, NF1 and STK11. In 24 (2%) cases a LP/P germline variant was detected but not reported in the tumor.

### **Conclusions:**

One-third of the patients with clinically significant variants in tumors were identified to carry the same variant in the germline, uncovering a previously unknown risk of hereditary cancer. Notably, some genes had a high probability of variants occurring in the germline, while others were primarily seen in tumors. Interestingly, 6% of the germline variants were not included in the tumor report due to technical and gene content differences in either assay or due to differences of clinical classification between tumor and germline testing. Performing simultaneous germline and tumor testing may inform susceptibility to a hereditary cancer disorder, identify targets for precision therapy, early

detection of secondary malignancies, and guide genetic counseling for relatives.

## PgmNr 2428: Synergy between next-generation sequencing, chromosomal microarrays, and conventional cytogenetics in the diagnostic evaluation of pediatric leukemias.

### Authors:

G. Raca <sup>1</sup>; J. Ji <sup>1</sup>; M.J. Oberley <sup>1</sup>; D. Bhojwani <sup>2</sup>; A. Doan <sup>2</sup>; R.J. Schmidt <sup>1</sup>; J.A. Biegel <sup>1</sup>; M.C. Hiemenz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Los Angeles, CA; 2) Children's Center for Cancer and Blood Diseases, Children's Hospital Los Angeles, Los Angeles, CA

Chromosomal microarrays (CMA) and next generation sequencing (NGS) panels are increasingly used with standard cytogenetics (karyotyping) and fluorescent in situ hybridization (FISH) in the diagnostic evaluation of pediatric leukemias. To assess the clinical utility of NGS and CMA and develop optimal diagnostic algorithms, we analyzed 173 bone marrow samples by karyotyping, FISH, CMA and our OncoKids<sup>®</sup> NGS panel. OncoKids<sup>®</sup> is a DNA and RNA-based AmpliSeq pediatric cancer panel designed to detect sequence variants and RNA fusions. Pediatric B lymphoblastic leukemia (B-ALL; n=137) and acute myeloid leukemia (AML; n=36) samples were included in this retrospective analysis.

In B-ALL, FISH had the highest diagnostic yield as a single test (67%), and the combined use of CMA and OncoKids<sup>®</sup> had a similar overall detection of primary drivers (70% of B-ALL cases) as cytogenetics and FISH. With the exception of IGH rearrangements, CMA and OncoKids<sup>®</sup> detected all alterations identified by cytogenetics and FISH, but also revealed primary abnormalities in 11 cytogenetically negative cases. These alterations included 'Ph-like' fusions (n=4), ZNF384 fusions (n=3), PAX5 fusions (n=2), a MEF2D fusion (n=1) and a KMT2A fusion (n=1). Importantly, CMA and NGS also revealed secondary clinically significant variants including IKZF1 deletions (n=32) and JAK1/JAK2 mutations (n=15).

In AML, OncoKids<sup>®</sup> had the highest yield for detection of both primary and secondary DNA mutations and RNA fusions (81% of AML cases). All abnormal fusions revealed by cytogenetics/FISH were identified with OncoKids<sup>®</sup>. Additionally, in 12 cases, OncoKids<sup>®</sup> identified the key driver that was not observed with cytogenetics/FISH. These 12 cases included seven gene fusions (NPM-MLF1, CBFA2T3-GLIS2, three KMT2A fusions, and two NUP98-NSD1 fusions) as well as clinically significant sequence changes in NF1, GATA1, RUNX1 and FLT3.

Our data highlights that CMA/NGS and karyotyping/FISH are complementary in B-ALL testing. Due to rapid turnaround time (TAT), FISH may be useful as an initial screening test but may ultimately miss important primary driver alterations and secondary abnormalities that could be helpful in treatment decisions. Therefore, the optimal testing algorithm for B-ALL must integrate clinical utility with information regarding cost, TAT, and institutional resources. For pediatric AML, our data suggests that NGS testing may represent a superior alternative to karyotyping as a first line test.

## PgmNr 2429: A case of a lymphoplasmacytic lymphoma with trisomy 12 in the lymphoid population and deletion 13q of myeloid origin.

### Authors:

A. Reyes <sup>1,2</sup>; V. Tse <sup>1,2</sup>; G. Yang <sup>1,2</sup>; E. Peng <sup>1,2</sup>; M. Yun <sup>1,2</sup>; K. Cunnien <sup>3</sup>; K. Lapp <sup>3</sup>; B. Wyatt <sup>3</sup>; C.A. Tirado <sup>1,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The International Circle of Genetic Studies, Los Angeles, CA; 2) UCLA; 3) Allina Health Inc, Minneapolis, MN; 4) HPA, Minneapolis, MN; 5) The University of Minnesota, Department of Pathology, Minneapolis, MN

---

An 89 year old male patient who presents with lymphoplasmacytic lymphoma (LPL)(WHO 2016) involving 80% of the marrow cellularity with circulating lymphoma cells. Chromosome studies detected two unrelated abnormal clonal populations. One clone has trisomy 12 [9 metaphases] as the sole abnormality. The other clone has a 13q- [6 metaphases] as the sole abnormality. The remaining cells are cytogenetically normal. This karyotype was described as 47,XY,+12[9]/46,XY,del(13)(q12q21)[6]/46,XY[5]

Trisomy 12 (found in the B-cell mitogen stimulated cultured cells) is the most frequently reported chromosome abnormality in B-cell chronic lymphocytic leukemia (B-CLL) . However it has been seen also in other B-cell lymphoproliferative disorders including lymphoplasmacytic lymphoma. This finding was detected also by FISH using the FISH CLL panel analysis which is a statistical and morphology based study where only round, non-segmented cells are scored.

Deletions of 13q (found in the cells from the non-stimulated culture) have been identified in B-cell malignancies , non-Hodgkin's lymphomas [NHL] as well as myelodysplastic syndromes and chronic myeloproliferative neoplasms (Heim and Mitelman, 2015). However, due to this finding (13q- clone), the trisomy 12/13q- FISH slide was reviewed looking at the segmented cells. Fifty segmented cells were scored and a 13q- pattern was detected in 36% (18/50) of the cells suggesting that this finding (the 13q-clone) may be myeloid in origin. Clinicopathologic correlation of these results was recommended.

## PgmNr 2430: ADAM32 amplification in pediatric malignancies.

### Authors:

V.T. Tonk <sup>1</sup>; A.R. Rajan <sup>2</sup>; S.S. Singh <sup>2</sup>; K.B. Bhende <sup>1</sup>; S.T. Tonk <sup>2</sup>; S.A. Awasthi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, TTUHSC, Lubbock, Texas.; 2) Internal Medicine, TTUHSC, Lubbock, Texas.

---

DNA microarray techniques allow the simultaneous detection of vast number of copy number alterations (CNA) in cancers, including previously unknown ones. The study presented here was aimed at surveying the genes and types of CNA found in the pediatric cancer patient population in the west texas region and was done at TTUHSC, Lubbock. Towards this goal, we performed a retrospective analysis of data obtained from Cytosure Constitutional v3 Microarray that is used for studies of developmental disorders. 198 cases over the previous 5 years were selected for study. The majority of cases were of pediatric malignancies, though some adult cancers and benign hematological diseases were included. We found that the most frequent CNAs were at the following loci: 14q11.2 (127 cases), 8p11.21-8p11.22 (98 cases), 15q11.1-15q11.2 (82 cases), 14q32.33, 9p24 and 9p13. The most affected locus being 14q11.2 was expected as ALL was strongly represented. The 8p11.21-8p11.22 locus was of interest because in this pediatric population, we found that gains outnumbered losses by 4.8-fold (81 vs 17 cases) at this locus, greater than for any other locus. LOH of 8p is reported to be frequent in human breast cancer and also found in other endodermal origin adult solid tumors, but not previously reported for pediatric patients (PMID: 29682206). Genes at this locus previously shown to be involved with gains or amplifications include DUSP26 (gain 14%, loss 34%), ZNF703 (gain 26%, loss 20%) (REF). In this (primarily pediatric) population, CNAs were most frequent at ADAM32 and found in all types of cancer: T-cell Lymphoma, ALL, Enchondroma, AML, Wilm's Tumor, Hepatocellular Carcinoma, Breast Cancer, Neuroblastoma, Rhabdomyosarcoma, Fibromatosis, Chondroblastoma and Giant Cell Tumor of the Bone. Other affected genes at this locus were KAT6A, TACC1, GPAT4 and TM2D2. Chromosomal translocations with breakpoints at these loci have been reported in breast and other cancers. To investigate whether these genes could be involved in promotion of cancer growth, we conducted studies to determine whether siRNA knockdown would affect growth of adult cancer cell lines. Interestingly, we found relatively small and inconsistent effects of ADAM32 knockdown. In contrast, knockdown of TACC1 kills inhibited cancer growth by 34-68%. Consistent with its function in the mitotic spindle, TACC1 knockdown resulted in abnormal metaphases. Our studies suggest TACC1 is a potential therapeutic target.



## PgmNr 2431: Integrative copy number analysis of uveal melanoma reveals novel candidate genes involved in tumorigenesis including a tumor suppressor role for *PHF10/BAF45a*.

### Authors:

R. Verstraten<sup>1,2</sup>; H. Anbunathan<sup>2</sup>; A. Singh<sup>3</sup>; J.W. Harbour<sup>4</sup>; A.M. Bowcock<sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oncological sciences, Icahn school of medicine at mount Sinai, New York, New York.; 2) National Heart and Lung Institute, Imperial College, London, United Kingdom; 3) Dept. of Ophthalmic Oncology, Cole Eye Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, Ohio 44195, U.S.A.; 4) Bascom Palmer Eye Institute, Sylvester Comprehensive Cancer Center and Interdisciplinary Stem Cell Institute, University of Miami Miller School of Medicine, U.S.A.; 5) Departments Dermatology and Genetics & Genome Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, U.S.A.

---

Uveal melanoma (UM) is a primary malignancy of the eye with oncogenic mutations in *GNAQ*, *GNA11*, or *CYSLTR2*. Approximately 50% of UMs are at high risk of metastasis to the liver and most have undergone loss of one copy of chromosome 3 and developed mutations in *BAP1* (Chr 3p21). The remainder of UMs generally harbor *SF3B1* or *EIF1AX* mutations and have less chance of metastasis. There are other characteristic chromosomal alterations in UM but their significance is not clear. For example, ~20-30% of UMs undergo hemizygous loss of chromosome 1p, and ~70% undergo 8q gain. *SF3B1* mutant tumors with 6q loss are also more likely to metastasize. To investigate genes driving chromosomal alterations we integrated copy number, transcriptome and mutation data from three independent cohorts and followed up key findings. We observed significant enrichment of transcripts on Chromosomes 1p, 3, 6, 8 and 16q and seven shared focal copy number alterations on Chr 1p36, 2q37, 3, 6q25, 6q27 and 8q24. Integrated analyses revealed clusters of genes in focal copy number regions whose expression was associated more strongly with metastasis and worse overall survival (OS) than the broad chromosomal changes. For example 8q gain is associated with poor prognosis and upregulation of 13 transcripts from Chr 8q24.3 including *PTK2* and *PTP4A3* were correlated with poor OS. Downregulation of 11 transcripts from chr1p36 were associated with poor OS. At Chr 6q27 we identified two tumors with homozygous deletion of *PHF10/BAF45a* (encoding a PBAF complex member involved in chromatin remodeling). A third UM had a frameshift mutation with concomitant loss of the wild type allele. SiRNAs knockdown of *PHF10* in three UM cell lines followed by RNA sequencing revealed enrichment of pathways involved in development, regulation of angiogenesis, and focal adhesion. Functional analyses revealed less adhesion to most ECM proteins and reduced migration in a chemotaxis assay. Several chromatin remodeling factors altered by *PHF10* knockdown such as *PCGF5* exhibited the same trend in the *PHF10* mutant tumors. Mutations in *PHF10* have not been described in tumors before, although mutations in other PBAF components are found in other cancer types. *PHF10* lies in a region of Chr 6q which harbors an unidentified tumor suppressor for a variety of epithelial cancers so besides contributing to tumor development in some cases of UM it should be considered as a candidate for this elusive tumor suppressor gene as well.

## PgmNr 2432: Next generation sequencing-based gene panel tests for the diagnosis of hereditary cancers.

### Authors:

S. Zeybek <sup>1</sup>; A. Alemdar <sup>2</sup>; N. Kaya <sup>3</sup>; L. Aliyeva <sup>1</sup>; A. Kablan <sup>1</sup>; D. Pirim <sup>4</sup>; E. Yildirim <sup>4</sup>; S. Sag <sup>1</sup>; S. Temel <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bursa Uludag University, Faculty of Medicine, Department of Medical Genetics, Bursa, Turkey; 2) Bursa Uludag University, Institute of Health Science, Translational Medicine PhD Program, Bursa, Turkey; 3) Bursa Uludag University, Institute of Natural and Applied Science, Molecular Biology and Genetics MSc Program, Bursa, Turkey; 4) Bursa Uludag University, Faculty of Arts & Science, Department of Molecular Biology and Genetics, Bursa, Turkey

---

Next generation sequencing-based gene panels are more efficient and cost effective in diagnosis of many cancer syndromes. Large number of variants can be identified by these panels due to the genetic heterogeneity of these cancers. Interpretation of the obtained genetic information becomes very substantial. In this study, 193 patients with BRCA1/2 negative Hereditary Breast/Ovarian Cancer (HBOC) or different hereditary cancer syndromes admitted to our clinic were screened for 27 genes related with Hereditary Cancers. All exons and exon-intron boundaries of 27 genes were amplified using Hereditary Cancer Solution (Sophia) (156 cases) and BRCA Hereditary Cancer Mstr plus (Multiplicon) (37 cases) kits and sequenced with MiSeq (Illumina) system. Sophia DDM platform and Sophia Genetics' MOKA software was used for variant analysis and annotations. Pathogenicity of variants was analyzed using online tools (HGMD Professional, ClinVAR, MutationTaster, SIFT, Polyphen) and ACMG criteria. We identified the total of 146 variant; of these variants pathogenic (11), likely pathogenic (8), variant of uncertain significance (VUS)(112) and conflicting interpretation of pathogenicity (15) seen in Hereditary Cancer patients. Additionally in 60 BRCA1/2 negative HBOC patient, we found 3 pathogenic, 3 likely pathogenic, 34 VUS and 5 conflicting variants. Implementations of targeted next generation sequencing based panels for hereditary cancers give us many variations that need to be carefully reviewed. They provide clinically critical informations in the diagnosis of BRCA1/2 negative HBOC patients in addition to other familial cancer syndromes.

## PgmNr 2433: Amplification of *BCR-ABL1* fusion gene in acute leukemia: Report of three new cases.

### Authors:

J. Liu <sup>1</sup>; S. Gargan <sup>1</sup>; N. Nwaoduah <sup>1</sup>; Z. Wang <sup>1</sup>; J. Gong <sup>1</sup>; G. Uppal <sup>1</sup>; D. Foster <sup>1</sup>; R. Bajaj <sup>2</sup>; S. Peiper <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pathology, Anatomy and Cell Biology, Thomas Jefferson Univ, Philadelphia, Pennsylvania.; 2) Clinical and Molecular Cytogenetics, Section of Pathology, St. Christopher's Hospital for Children, Drexel University College of Medicine, Philadelphia, Pennsylvania.

---

The Philadelphia chromosome (Ph) resulting from a reciprocal translocation t(9;22)(q34;q11) is a characteristic finding in chronic myeloid leukemia (CML) and in a subset of acute lymphoblastic leukemia (ALL) as well as rarely in acute myeloid leukemia (AML). The derivative chromosome contains the BCR-ABL1 fusion gene that encodes for a protein with constitutive protein kinase activity. Amplification of BCR-ABL1 and mutations in the ABL1 kinase domain have been described in Ph+ leukemia cases that were resistant to tyrosine kinase inhibitor (TKI). Here we report three new Ph+ acute leukemia cases with amplification of the BCR-ABL1 fusion. Case 1 is a 35 year-old male with Ph+ ALL who underwent two allogeneic stem cell transplantations and treatment with multiple TKIs, and upon his second relapse demonstrated complex karyotype (50-51,XY,add(4)(q35),+8,t(9;22)(q34;q11),del(10)(q24),del(12)(q24.1),+17,+18,+20,der(22)t(9;22)(q34;q11)x1-2,+1-4mar[cp20]), amplification of Ph+ by fluorescence in situ hybridization (FISH), and multiple ABL1 kinase mutations. This patient was TKI-refractory and succumbed to his disease 7 years after diagnosis. Case 2 is a 77 year-old female with newly diagnosed AML had a complex karyotype (44~46,XX,t(1;3)(p36;p21),-4,-5,add(10)(p11.2),der(22)t(9;22)(q34;q11.2),+mar[20]), BCR/ABL1 translocation with both Ph and MLL amplification and TP53 deletion. This is the first documented case of AML with t(1;3) in conjunction with t(9;22) and Ph amplification. Case 3 is 60 years old male with newly diagnosed AML. Complex karyotype 51-69,XY,+Y,+1,+2,+3,+6,i(6)(p10),+8,t(9;22)(q34;q11.2),+10,+11,+12,+13,del(13)(q12q14),+14,+15,+16,+18,+19,+20,+der(22)t(9;22)[cp4]/46,XY[10] was seen by cytogenetics and BCR/ABL1 translocation with Ph amplification seen by FISH. The patient was given induction chemotherapy but deceased two month late due to severe sepsis and cardiogenic shock. These three cases illustrate the genetic heterogeneity of Ph-positive leukemias. Numeric and/or structural cytogenetic aberrations, as well as ABL1 kinase mutations, may be present at initial diagnosis or acquired with disease progression, and they are associated with an aggressive disease course and resistance to TKI therapy. Understanding the underlying molecular pathogenesis and appropriate of such cases would help to provide more effective clinical management.

## PgmNr 2434: Next-generation cytogenetics: High-resolution optical mapping to replace FISH, karyotyping and CNV-microarrays.

### Authors:

T. Mantere <sup>1</sup>; K. Neveling <sup>1</sup>; M. Stevens-Kroef <sup>1</sup>; D. Olde Weghuis <sup>1</sup>; M. Pauper <sup>1</sup>; G. van de Zande <sup>1</sup>; S. Vermeulen <sup>1</sup>; E. Kater-Baats <sup>1</sup>; M. Oorsprong <sup>1</sup>; R. van Beek <sup>1</sup>; D. Smeets <sup>1</sup>; A. Hastie <sup>2</sup>; A. Hoischen <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 2) Bionano Genomics, San Diego, California, USA; 3) Radboud Institute of Molecular Life Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; 4) Department of Internal Medicine and Radboud Center for Infectious Diseases (RCI), Radboud University Medical Center, Nijmegen, The Netherlands

---

Structural variants (SVs) are an important source of genetic variation in the human genome and they are involved in a multitude of human diseases. Somatic SVs are important for cancer development and progression while constitutional SVs are involved in many developmental disorders. In a diagnostic set-up, comprehensive analysis of all molecular cytogenetic aberrations in a given sample still requires a combination of techniques, such as CNV-microarrays, karyotyping and fluorescence *in situ* hybridization. We hypothesize that the combination of classical approaches could be largely replaced by high-resolution optical mapping.

Leukemia bone marrow aspirates and blood samples were analyzed using standard of care workflow and the residual samples were processed for Bionano Saphyr optical mapping to detect chromosomal aberrations. We generated up to 400-fold genome coverage by imaging long chromosomal fragments and detected chromosomal aberrations by comparing optical maps to a reference and control dataset.

In the first leukemia samples we compared the results with those obtained using standard methods. In each sample, all previously known clinically relevant aberrations were identified. This held true for deletions, insertions, inversions and translocations, including a 3-way Philadelphia chromosome (t(9;14;22)(q34;q11;q11.2)) and even chromothripsis structures were resolved. We were also able to identify aberrations in samples with cancer cell content less than 40% using Bionano's latest single-molecule SV detection tool. Importantly, optical mapping additionally identified multiple novel events, e.g. an inversion of chromosome 11 (chr11:24,875,044-26,299,641) and a translocation (t(5;14)(q35.2;q32.2)), both validated afterwards. These novel events are now followed-up to identify possible fusion genes as novel leukemia drivers.

We are now conducting a study to systematically compare the sensitivity and specificity of optical mapping in 100 leukemia samples and 50 samples with known germline cytogenetic aberrations against the standard of care workflow. Of these, 17 with constitutional aberrations and 19 leukemia samples have been processed successfully so far.

Optical mapping may have the potential to replace most classical cytogenetic tests. In addition this

holds the potential to map new fusion genes as novel leukemia drivers rapidly.

## PgmNr 2435: Acute myeloid leukemia with iAMP21.

**Authors:**

G. Tang; W. Xie

[View Session](#) [Add to Schedule](#)

**Affiliation:** Hematopathology, U.T. MD Anderson Cancer Ctr, Houston, Texas.

---

**Background:** iAMP21 is defined as amplification of a portion of chromosome 21, typically detected by FISH with  $\geq 5$  copies of *RUNX1* gene in a cell or  $\geq 4$  copies on a single abnormal chromosome 21. B-lymphoblastic leukemia with iAMP21 has been well studied and recognized as a provisional entity by the current WHO classification, it is often associated with poor prognosis and requires more intensive therapy. Acute myeloid leukemia (AML) with iAMP21 is rare and its clinicopathologic features and outcomes remain largely unknown.

**Methods:** We searched our database in the past 10 years for cases with iAMP21 confirmed by FISH, then narrowed down to the patients with AML. Clinicopathologic features, molecular and cytogenetic information were collected from the charts.

**Results:** we identified 9 patients with AML and iAMP21, four women and five men, with a median age of 64 years (range, 34 to 75 years). Five patients had iAMP21 detected at the initial diagnosis of de novo AML, 1 at relapsed AML, 2 at the AML progressed from myelodysplastic syndrome, and 1 therapy-related AML. All patients showed pancytopenia and various degrees of myelofibrosis; 8 patients had a complex karyotype; 4 patients harbored a *TP53* mutation. Eight patients received induction chemotherapy or hypomethylating agents, one patient also received allogeneic stem cell transplant (SCT). By the end of follow-up, 5 patients died, 3 alive with persistent disease and 1 alive in complete remission (who received SCT), with a median overall survival of 1 month.

**Conclusion:** We conclude that AML with iAMP21 is often associated with pancytopenia, myelofibrosis, *TP53* mutation, and a complex karyotype. Patients with AML-iAMP21 are often refractory to conventional AML therapy and have very poor outcomes. SCT at the first remission as well as novel therapies may improve the outcome.

## PgmNr 2436: Next generation sequencing of *BRCA1* and *BRCA2* genes among women with breast cancer in Burkina Faso.

### Authors:

M. Biancolella <sup>1</sup>; NLM. Ouedraogo <sup>2</sup>; B. Testa <sup>3</sup>; N. Zongo <sup>4</sup>; O.M. Lompo <sup>5</sup>; F. Djigma <sup>6</sup>; C. Conte <sup>7</sup>; I. Bagni <sup>7</sup>; F. Sangiuolo <sup>3</sup>; J. Simpoire <sup>2,6</sup>; G. Novelli <sup>3,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biology , University of Rome "Tor Vergata ," , Rome, Italy; 2) Université Saint Thomas d'Aquin/ Hôpital Saint Camille de Ouagadougou -06 BP 10212 Ouagadougou, Burkina Faso;; 3) Department of Biomedicine and Prevention, University of Rome "Tor Vergata ," Rome, Italy; 4) Service of General Surgery and Digestive, University Hospital Yalgado Ouédraogo, Burkina Faso;; 5) Yalgado Ouédraogo University Hospital (CHU/YO), 03 BP 7022, Burkina Faso; 6) Laboratory of Molecular Biology and Genetics/ Pietro Annigoni Biomolecular Research Centre (CERBA), (LABIOGENE/CERBA), Joseph Ky-Zerbo University, 03 BP 7021 Ouagadougou 03, Burkina Faso; 7) Laboratory of Medical Genetics, Tor Vergata Hospital , Rome, Italy; 8) Neuromed, I.R.C.C.S, Pozzilli, Italy

Breast cancer is the most commonly diagnosed cancer in Africa, including sub-Saharan Africa (SSA) and the leading cause of cancer mortality in women. To date, knowledge of the etiological factors of breast cancer in Africa has been poorly studied.

The early-onset and the aggressive clinical features of breast cancer in patients of African ancestry suggest that the hereditary predisposition may have an important function, but the absence of extensive studies, specially for Sub-Saharan Africa (SSA), greatly reduces the understanding of the impact of germline mutations in breast cancer patients and its effects in terms of prevention, diagnosis and patient management.

The aim of this study is to define the impact of germline mutations in *BRCA1* and *BRCA2* genes in breast cancer among young women in Burkina Faso and determine their implication in clinics perspectives.

Fifty-two women with early-onset breast cancer ( $\leq 40$  years), eleven of them with a family history were analyzed by Next Generation Sequencing (NGS).

Six different pathogenic mutations (3 in *BRCA1*, 3 in *BRCA2*), two of which recurrent in more than one patient, were identified in eight unrelated patients.

Moreover, three variants of uncertain clinical significance (VUS) and three novel unclassified variants were identified in six other patients with no family history of the disease.

The present study is the first one in which the entire coding sequence of both *BRCA* genes is analyzed in Burkinabe women with breast cancer. Our data suggest that hereditary predisposition is responsible for a significant proportion of breast cancer in Burkina Faso women that should undergo genetic cancer risk assessment. The identification of the most common mutations of *BRCA1* and *BRCA2* genes in the Burkina Faso population will allow the development of a cost-effective genetic test for the early detection of these mutations contributing to the medical treatment of women with breast cancer, with important consequences in society and economy of this country.

## PgmNr 2437: *BRCA1* and *BRCA2* germline mutational spectrum in Algerian population: What we know now.

### Authors:

F. Cherbal <sup>1</sup>; C. Mehemmai <sup>1</sup>; L.L. Saada <sup>1</sup>; M. Nehari <sup>1</sup>; K. Yatta <sup>1</sup>; H. Gaceb <sup>1</sup>; H. Mahfouf <sup>2</sup>; W. Benbrahim <sup>3</sup>; K. Boualga <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Unit of Genetics, LMCB, Faculty of Biological Sciences, USTHB, Algiers, Algeria; 2) Mohamed El Kolli Public Hospital, Academic Medical Oncology Services, School of Medicine, University of Algiers, Rouiba, Algeria; 3) Anti Cancer Center, Batna, Algeria; 4) Anti Cancer Center, Blida, Algeria

---

**Background:** BRCA genetic testing can reduce incidence and mortality of BRCA-linked cancers. There are limited data regarding BRCA genetic testing in Algerian population. Between 2008 and 2018, we screened 288 breast/ovarian cancer families for *BRCA1* and *BRCA2* germline mutations. Here, we present the experience of our research laboratory and some unpublished results.

**Methods:** 100 index cases and relatives were screened for complete *BRCA1* and *BRCA2* germline mutations using HRM-direct sequencing (86 individuals) or NGS using a cancer panel of 30 hereditary cancer genes or *BRCA1/2* genetic test in 14 probands (Color Genomics, Burlingame, California, USA). In addition, 215 index cases and relatives were tested for *BRCA1* and *BRCA2* recurrent mutations previously found in Algerian HBOC patients using PCR-direct sequencing. Screening for large genomic rearrangements (LGR) in *BRCA1* and *BRCA2* genes was performed in 79 patients tested negative for small point mutations using MLPA technique. To confirm the genomic rearrangement and its genomic breakpoints, long-range PCR of genomic DNA was performed.

**Results:** The analysis of DNA samples of 315 individuals revealed that 42 patients and relatives carried pathogenic germline mutations: 32 within *BRCA1* and 10 within *BRCA2*. 10 distinct pathogenic mutations have been identified in *BRCA1* gene: c.19\_47del, c.83\_84delTG, c.181T>G, c.505C>T, c.798\_799delTT, c.2125\_2126insA, c.5332 + 1G > A and 3 LGR (exon 2 del, exon 7 del and exon 15del). Six distinct pathogenic mutations have been identified in *BRCA2* gene: c.1310\_1313delAAGA, c.1813dupA, c.5722\_5723delCT, c.7654dupA, c.8485C>T and c.8940delA. The following recurrent *BRCA1* mutations have been identified in 12 unrelated families: c.83\_84delTG (4 families), c.181T>G (2 families), c.798\_799delTT (3 families) and c.2125\_2126insA (3 families). *BRCA1* exon 7 del and *BRCA1* exon 15 del have been identified within four different new genomic breakpoints :3,135bp and 4,200bp; 3,341bp and 4,112bp, respectively (each deletion/2 unrelated families). The *BRCA2* c.1310\_1313delAAGA mutation has been detected in two unrelated families. Interestingly, our study showed differences in the distribution of the mutation spectrum of *BRCA* genes between the Eastern region and the North central region of Algeria.

**Conclusions:** Our study will help to implement affordable genetic testing and to improve the clinical management and better risk assessment of hereditary breast and ovarian cancer in Algerian population.



## PgmNr 2438: Development of an expanded hereditary breast and ovarian cancer next-generation sequencing panel; experience of collaboration between the CHEO Genetics Diagnostic Laboratory and clinical Hereditary Cancer Program.

### Authors:

P.E. Harper; R. Potter; E. Smith; E. Tomiak; A. Rusnak; L. Huang; L. Bronicki; O. Jarinova; E. Sinclair-Bourque; J. McGowan-Jordan; A. Smith

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetics Department, CHEO, Ottawa, ON, Canada

---

Expanding next-generation sequencing (NGS) panels for hereditary breast and ovarian cancer syndrome (HBOC) beyond *BRCA1/2* testing increases the diagnostic yield of clinically actionable genes. Over the past years, NGS-based panels have been expanding to include many more genes, some with limited clinical utility that do not necessarily translate to higher quality clinical care. In some cases, genes with limited information or no clear guidelines for screening and management are included. Furthermore, expanding NGS-based panels comes with limitations of resource management for smaller laboratories operating in a publically funded health care setting. With the goal of expanding our HBOC panel, the CHEO Genetics Diagnostics Laboratory and clinical Hereditary Cancer Program collaborated to implement an expanded hereditary cancer panel, with a strong consideration to the clinical utility of the genes included and the limited resources available in the laboratory. The implementation of the expanded panel involved developing a new workflow, updating result report formats and requisitions to ensure timely and high quality results. An expedited surgical panel (*BRCA1/2*, *PALB2* and *TP53*), including sequence variant and copy number detection, is followed by one of two expanded panels, a 9 gene breast cancer panel (*ATM*, *CDH1*, *CHEK2*, *PTEN*, *STK11*) or a 17 gene breast and ovarian cancer panel (*ATM*, *BRIP1*, *CDH1*, *CHEK2*, *EPCAM*, *MLH1*, *MSH2*, *MSH6*, *PMS2*, *PTEN*, *RAD51C/D*, *STK11*). In addition, ongoing collaboration between clinical team and laboratory staff provides the opportunity for feedback including improved variant interpretations supported by clinical information. Developing multi-gene panel testing raises resource challenges for laboratories in publically funded health care settings. Our joint laboratory and clinical department approach implemented an expanded HBOC cancer panel under limited resources while maintaining a high quality of patient care.

## PgmNr 2439: DNA-depleted plasma-based extraction controls enable effective assessment of specificity and sensitivity of liquid biopsy assays.

### Authors:

Y. Lu; K. Ooi; K. Shi; T. Fu; E. Chan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Anchor Molecular Inc, Pleasanton, California.

The usage of PCR-based or NGS-based liquid biopsy assays for circulating tumor DNA (ctDNA) detection is growing rapidly in clinical labs. However, inconsistent results are often observed especially when assays from different developers are compared. One reason might be due to lack of thorough understanding of the role of plasma factors on cfDNA extraction since the extraction step contributes significantly to assay variation. To address this issue, Anchor Molecular has developed a proprietary technology that specifically removes DNA from plasma without affecting the composition of the plasma. The DNA content in the treated plasma was measured to be less than 0.01ng/mL. The composition of all other components such as proteins and lipids remain the same as in normal patient plasma. Extraction Quality control samples based on this DNA-depleted plasma were used to evaluate the effect of plasma factors on extraction and sensitivity and specificity of cfDNA assay.

These controls are made by either synthetic DNA or native DNA fragments derived from cell lines spiked into DNA-depleted plasma. The recovery and stability of these DNAs were studied by extractions followed by quantitative PCR (qPCR). Specificity of cfDNA extraction was studied by spiking both low- and high-molecular-weight DNA fragments into DNA-depleted plasma followed by extraction by five different cfDNA extraction kits. The purified DNAs were analyzed using Bioanalyzer 2100. The sensitivity of cfDNA extraction was examined by extractions of the spiked low concentration of target DNA fragments with different commercial kits followed by qPCR. The observed and expected concentrations of extracted DNA-depleted, plasma-spiked DNAs were shown to be linearly correlated, mimicking the spiked DNAs in normal plasma.

Both synthetic DNA and fragmented genomic DNA showed at least 125 days of real-time stability at 2-8°C. The specificity study showed that different kits exhibited varying preferences in extracting DNAs of different sizes. In the sensitivity study, differences in extraction efficiency and precision were observed among different kits.

The data demonstrated that plasma factors influence cfDNA extraction. Quality controls samples based on DNA-free plasma can be effectively used for monitoring the entire ctDNA assay process including the extraction step.

## **PgmNr 2440: Three way t(8;14;22)(q24;q24;q32;q11.2) characterized by molecular cytogenetics with involvement of *MYC/IGH/IGL* in a case of a diffuse large B-cell lymphoma (DLBCL).**

### **Authors:**

C. Tirado <sup>1,2,3,4</sup>; D. Chung <sup>4</sup>; A. Reyes <sup>4</sup>; B. Wyatt <sup>2</sup>; K. Cunniën <sup>2</sup>; K. Lapp <sup>2</sup>; D. Serk <sup>2</sup>; J. Rankin <sup>2</sup>; J. Ewing <sup>2</sup>; J. Han <sup>2</sup>; K. Stieglbauer <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Cytogenetics Laboratory, Allina Health Laboratory, Minneapolis, Minnesota.; 2) Department of Laboratory Medicine & Pathology University of Minnesota, Minneapolis, MN; 3) HPA, Minneapolis, MN; 4) The International Circle of Genetic Studies, Los Angeles, CA

A 48 year old male presented with palpable right inguinal lymphadenopathy. CT scan revealed a 3 cm enlarged lymph node and several adjacent smaller lymph nodes measuring up to 1.2 cm. No additional lymphadenopathy was detected. Past medical history was significant for right testicular seminoma treated by radical orchiectomy at 27 years of age, and remote history of Hodgkin lymphoma diagnosed at 15 years of age treated with right upper neck lymph node resection, radiation therapy and splenectomy. Excisional right inguinal lymph node biopsy revealed diffuse large B-cell lymphoma (DLBCL), NOS with a germinal center B-cell subtype (CD10+/BCL6+/MUM1-). The lymphoma cells were *BCL2*-negative and showed >95% proliferation index by MIB1 (Ki-67) immunohistochemical stain. The lymphoma cells lacked surface and cytoplasmic light chain expression by flow cytometry. Chromosome analysis revealed an abnormal clone with t(8;14;22)(q24;q32;q11.2) within the context of a complex karyotype. Metaphase FISH analysis detected a *MYC/IGH* rearrangement on the derivative chromosome 14 and also an *IGL* (22q11.21-q11.23) rearrangement. A signal of the distal portion of *IGL* (BCR) was present on the derivative chromosome 8. Metaphase FISH using the *IGH* break-apart probe showed intact fusion signal on the normal copy of chromosome 14 and in the derivative chromosome 14 as well as an 5' *IGH* signal on the derivative chromosome 22. FISH for 14q subtelomere showed 14q subtelomere signal on the derivative chromosome 22.

Herein we pinpoint the involvement of *MYC/IGH/IGL* in a three-way translocation in a DLBCL case characterized by FISH. The t(8;14)(q24;q32) is a recurrent chromosome abnormality described in non-Hodgkin lymphomas (NHL), especially in 60-70% of Burkitt lymphoma (BL) and diffuse large B-cell lymphomas. The variant t(8;22)(q24;q11) is also seen in these cases. *MYC* rearrangements have been observed in up to 10% of cases of Diffuse Large B Cell Lymphomas (DLBCL) and is usually associated with a complex pattern of genetic alterations. This particular pattern with *IGH-MYC* rearrangements within the context of complex karyotypes is seen in diffuse large B-cell lymphomas. Complex karyotypes are associated with genomic instability and a poor prognosis. Clinical correlation was recommended.

## PgmNr 2441: Cytogenetic findings in a case of blastic plasmacytoid dendritic cell neoplasm (BPDCN).

### Authors:

D. Zhao <sup>1,2</sup>; A. Nguyen <sup>1,2</sup>; K. Lapp <sup>3</sup>; K. Stieglbauer <sup>3,4</sup>; J. Reinartz <sup>3,4</sup>; C.A. Tirado <sup>2,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of California, Los Angeles, San Ramon, California.; 2) The International Circle of Genetics Studies Chapter Los Angeles; 3) Allina Health Inc, Minneapolis, MN; 4) HPA, Minneapolis, MN; 5) The University of Minnesota, Department of Pathology, Minneapolis, MN

---

The patient is a 67-year-old female with a recent peripheral blood morphology revealing a hematopoietic leukemia process. Flow cytometry revealed an atypical cell population without B-cell or T-cell lineage expression. It was positive for CD45 and CD123 and negative for CD34. She undergoes a bone marrow biopsy. Bone marrow and peripheral blood showed a blastic plasmacytoid dendritic cell neoplasm (BPDCN), Hypercellular marrow (estimated 95%) with 90.4% blasts (aspirate smear). The peripheral blood showed blastic plasmacytoid dendritic cell neoplasm, macrocytic anemia, and moderate thrombocytopenia.

Chromosome analysis of twenty G-banded metaphases revealed an abnormal karyotype with i(7)(q10) and monosomies of 13 and 15 in 10 metaphase cells examined . The karyotype was described as 46,XX,i(7)(q10),-13,-15[10]/46,XX[10].

Loss/deletions of 7p [i(7)(q10)], 13q and 15q have been reported as recurrent chromosome abnormalities in blastic plasmacytoid dendritic cell neoplasm (BPDCN). Clinicopathologic correlation of these results was recommended.

## PgmNr 2442: Gene panel based prediction of homologous recombination deficiency in adolescent and young adult breast cancers.

### Authors:

T. Watanabe <sup>1,2</sup>; T. Honda <sup>1,3</sup>; H. Totsuka <sup>4</sup>; E. Arai <sup>5</sup>; Y. Kanai <sup>5</sup>; K. Shiraishi <sup>1</sup>; T. Kohno <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genome Biology, National Cancer Center Research Institute, Tokyo, Japan; 2) Department of NCC Cancer Science, Tokyo Medical and Dental University, Tokyo, Japan; 3) Department of Respiratory Medicine, Tokyo Medical and Dental University, Tokyo, Japan; 4) StaGen Cooperation, Ltd., Tokyo, Japan; 5) Department of Pathology, Keio University School of Medicine, Tokyo, Japan

---

Breast cancer diagnosed at ages of adolescent and young adults (AYA; 15 to 39 years old) consists approximately 5% of all breast, waiting for developing therapeutic options. Homologous recombination deficiency (HRD) phenotype often observed in tumors with *BRCA1/2* deficiency is considered to be associated with efficacy of PARP inhibitor and platinum agent-based therapies. In this study, HRD score was calculated for 47 Japanese AYA breast cancers using their whole exome sequencing data. The HRD-high phenotype was defined as HRD scores  $\geq 42$ . Genetic and pathological factors in HRD-high cases were determined by analyzing the sequencing data and DNA methylation chip analysis data as well as clinicopathological information of the cohort. The HRD-high phenotype was observed in 13/47 (27.7%) cases and was preferentially observed in cases with germline *BRCA1/2* and somatic *TP53* mutations, triple negative subtype and higher nuclear grades. Two “pathogenic” *BRCA1/2* germline mutations were observed in the 13 HRD-high cases, while a case with a “likely pathogenic” *BRCA1* germline mutation was judged as HRD-low. Three *BRCA1* hypermethylated and a *RAD51C* hypermethylated cases were also included in the 13 HRD-high cases. A predictive model for the HRD-high phenotype was developed using a TCGA data set of all breast cancers ( $n = 744$ ; Area under the curve [AUC] = 0.86) based on *BRCA1/2* and *TP53* mutations, hypermethylation status, triple negative subtype and higher nuclear grades. This prediction model was generated from coefficients of a logistic regression analysis and these five factors were statistically significant in a TCGA data set ( $P < 0.01$ ). Its prediction power was validated in Japanese ( $n = 46$ ; AUC = 0.90) and European ( $n = 58$ ; AUC = 0.95) AYA cohorts. After excluding a factor of methylation status, the prediction model derived from four factors still showed high AUC (AUC = 0.85). Its prediction power was validated in Japanese ( $n = 46$ ; AUC = 0.90) and European ( $n = 58$ ; AUC = 0.96) AYA cohorts. These four factors can be assessed by gene panel tests and daily pathological analyses. Thus, this study clarified genomic and pathological factors associated with HRD phenotype of AYA breast cancers. The present predictive model would be a tool to identify AYA breast cancer patients who would benefit from PARP and/or platinum therapies in the clinical setting.

## PgmNr 2443: Role of calreticulin gene (*CALR*) in changing the WHO diagnostic criteria of myeloproliferative neoplasms.

### Authors:

R. Ahmed <sup>1,2</sup>; M. Rashid <sup>2</sup>; U. Zaidi <sup>3</sup>; T. Shamsi <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Genetics, University of Karachi, Karachi, Sindh, Pakistan; 2) Dept. of Postgraduate Studies and Research, NIBD, Karachi, Sindh, Pakistan; 3) Dept. of Hematology, NIBD, Karachi, Sindh, Pakistan

Calreticulin gene (*CALR*) encodes a calcium binding multifunctional protein localized primarily in the endoplasmic reticulum but is also found in the nucleus, cell membranes and extracellular matrix. *CALR* binds to mis-folded proteins and prevent their exported from the endoplasmic reticulum to the golgi apparatus. Outside the endoplasmic reticulum it has been implicated in diverse processes including proliferation, apoptosis, phagocytosis and immunogenic cell death. Until December 2013 *CALR* had never previously been reported mutated in cancer or to be associated with hematological disorders particularly Myeloproliferative neoplasms (MPN). The discovery of *CALR* mutations further assisted the diagnosis of *BCR-ABL1* negative MPNs and in response to this milestone and continuous research studies in favour of this finding, the World Health Organization (WHO) in 2016 eventually included *CALR* in the ET and PMF diagnostic criteria. All *CALR* mutations are majorly somatic heterozygous insertions or deletions resulting in a frame shift and cluster in exon 9 which is the last exon of the gene. Thus far, more than 50 different types of mutations in *CALR* have been reported. Two specific mutations, Type I (52bp deletion) and Type II (5bp insertion), are the most prevalent. Overall, these two mutation types are found in more than 80% of patients with mutant *CALR*.

**Methods:** Clinical and hematological features were obtained from 128 MPN patients. We studied 35 PV patients, 44 ET patients, 39 MF patients and 10 patients with Undifferentiated MPN (UMPN).

**Results:** In our cohort, *CALR* mutations were detected in 23 patients (17.9%; 23/128) of which one patients was positive for *JAK2* V617F as well. *CALR* mutations were found in 17 (13.2%; 17/128) ET patients and 6 (4.6%; 6/128) MF patients. Among ET patients *CALR* mutation frequency was 38.6% (17/44) and among MF patients it was 15.3% (6/39). *JAK2*V617F mutation was found in 74.2% (26/35) patients of PV, 21.9% (9/41) of ET, 13.7% of MF (4/29) and 40% (2/5) of UNPM. None of the UMPN patients exhibited any *CALR* mutation. In ET type I mutation was found in 6 patients (35.3%; 6/17) and among them 2 had two new scattered point mutations (c.1081C>G and c.1086C>G) as well, type II in 8 patients (47.0%; 8/17), and others in 3 patients (17.6%; 3/17). In MF patients type I was found in 4 patients (66.6%; 4/6), type II in 1 patient (16.6%; 1/6) and one novel homozygous *CALR* mutation in 1 patient (16.6%; 1/6).

## PgmNr 2444: Development of MethylArray for classifying tumors of the central nervous system.

### Authors:

M. Peracchio <sup>1</sup>; L. Holinka <sup>1</sup>; M. Hsaio <sup>1</sup>; K. Kelly <sup>1</sup>; D. Bergeron <sup>1</sup>; Q. Nie <sup>1</sup>; B. Meyers <sup>1</sup>; A. Hesse <sup>1</sup>; J. Davey <sup>1</sup>; Q. Wu <sup>2</sup>; R. Verhaak <sup>1</sup>; C. Lau <sup>1</sup>; H. Reddi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 2) Department of Pathology and Laboratory Medicine, UCONN Health, 263 Farmington Avenue, Farmington, CT

---

DNA methylation patterns have demonstrated clinical significance in classification, diagnosis, progression monitoring, and predicting treatment response in numerous cancer types. Recent studies have shown methylation patterns to be useful for classifying tumors of the central nervous system (CNS), many of which have poor prognosis and limited treatment success. Capper et al. (2018) developed a free online classifier tool (MNP v11b4, [www.molecularneuropathology.org](http://www.molecularneuropathology.org)) for classifying CNS tumors into one of 82 categories based on DNA methylation profiles. The classifier allows for a more robust, reproducible diagnosis of CNS tumors over traditional histopathology, which can be subject to pathologist-dependent variation. We have developed a diagnostic assay, MethylArray, which uses the Illumina Infinium HD Methylation EPIC array and the MNP classifier tool to categorize adult and pediatric CNS tumors from FFPE or fresh frozen specimens. We processed 76 CNS tumor samples, both adult and pediatric, through the Illumina Infinium HD Methylation EPIC protocol. Sample quality was assessed using the Infinium FFPE QC Kit, and bisulfite conversion was performed using the Zymo EZ DNA Methylation Kit. While the protocol suggested a minimum of 250 ng of input DNA, we were able to obtain reliable results at 50 ng of input DNA. Converted DNA was then processed through the Illumina Infinium FFPE Restore protocol to repair low-quality DNA from FFPE samples. Other workflow conditions, such as incubation temperatures and reagent volumes, were optimized for processing FFPE samples. The prepared BeadChips were scanned on the Illumina iScan, and the results were run through the MNP classifier tool. Of the 76 samples processed, 7 samples failed due to insufficient DNA or tissue. For the remaining samples, a specific CNS tumor classification was identified for 57 of 69 samples (83%), with 12 samples not matching to any of the 82 categories. The “No Match” classification indicates insufficient calibrated scores needed for a valid classification or that the case may represent a novel tumor entity. Based on our results, the Methylation EPIC array, in conjunction with the MNP classifier tool, has shown to be a reliable assay for classifying both adult and pediatric CNS tumors from FFPE or fresh frozen tissue. Our MethylArray assay will be a valuable tool for providing oncologists with an accurate and reliable diagnosis of CNS tumors leading to more informed treatment decisions for patients.



## PgmNr 2445: High accuracy NGS platform based on electrical impedance detection: Applications for oncology research.

### Authors:

S. Paliwal; M.R. Barmi; A. Nabi; X. Gomes; M. Fallahi; M. Jouzi; S. Stern; P. Kenney; K.B. Parizi; T.A. Clark; H. Rategh; S. Sankar; H. Esfandyarpour

[View Session](#) [Add to Schedule](#)

**Affiliation:** GenapSys Inc., Redwood City, California

---

Next Generation Sequencing (NGS) technologies have made rapid strides in the throughput and accuracy of DNA sequencing in recent years. These advances have revolutionized biomedical and clinical research, especially in oncology. NGS cancer panels are used to determine cancer predisposition, detect early cancer, identify tumor mutations, and develop personalized therapies. Here, Genapsys presents a novel, scalable, low cost, and high accuracy NGS platform, and demonstrates its applications to oncology research.

The GenapSys NGS platform is based on accurate detection of electrical impedance changes resulting from single base incorporations during sequencing-by-synthesis. We show that impedance changes measure a steady state dNTP incorporation signal, leading to higher accuracy. The core of the technology is a CMOS-based electronic chip that enables scalability and low instrument and consumable costs. Chips with 1M, 16M and 144M sensors can be run on the same GenapSys instrument, giving a lab flexibility in NGS assay design and sample multiplexing. We demonstrate that a single run with a 16M sensor chip generates 1.5 Gb of data, with greater than 99% raw accuracy and up to 175 bp read lengths.

We tested hybrid-capture and amplicon-based cancer panels on a range of DNA sources, including oncology reference standards derived from cell line DNA, as well as clinical FFPE and blood sample DNA. Reference DNA standards from Horizon Discovery included the Quantitative Multiplex (HD701), EGFR Gene-Specific Multiplex (HD802) and Oncospan (HD827). For hybrid-capture libraries, we tested the IDT xGen Pan Cancer Panel v1.5 (800 Kb target region, 127 genes) and the IDT xGen Exome Research panel (39 Mb target region, 19,396 genes). We detected low frequency mutations in the range of 1%-24.5% across multiple standards with the Cancer panel, with mean coverage of >600x in a single run. Whole exome sequencing of clinical FFPE and blood samples showed high concordance (F1 score > 95%) of SNV mutation calling with industry standard technology. For amplicon panels, we used the Ion AmpliSeq Cancer Hotspot Panel v2 (207 amplicon pairs, 50 genes) with Horizon reference standards. We demonstrate detection of low frequency mutations (>1%) and a high correlation (R squared > 0.99) with expected allele frequencies. Thus, we demonstrate that the GenapSys NGS platform is an accurate, scalable, and low cost solution for oncology research on a wide range of sample types and NGS assays.



# PgmNr 2446: Novel triadic fusions among *ZNF384*, *EWSR1* and *EHMT1* genes in pediatric B-cell precursor acute lymphoblastic leukemia with translocations resembling Philadelphia chromosome.

## Authors:

Chen <sup>1</sup>; F. Wang <sup>1</sup>; P. Cao <sup>1</sup>; Y. Zhang <sup>1</sup>; X. Ma <sup>1</sup>; M. Liu <sup>1</sup>; W. Teng <sup>1</sup>; J. Yang <sup>1</sup>; J. Zhang <sup>1</sup>; T. Wang <sup>1,2</sup>; H. Liu <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Hebei Yanda Lu Daopei Hospital, Langfang, Hebei, China; 2) Beijing Lu Daopei Hospital, Beijing, China; 3) Beijing Lu Daopei Institute of Hematology, Beijing, China

**Aims** We report triadic fusions among *ZNF384*, *EWSR1* and *EHMT1* genes in a 5-year-old girl who relapsed with Ph like chromosome after initial diagnosis of normal karyotype BCP-ALL.

**Methods** Conventional laboratory tests were performed for the diagnosis of BCP-ALL. Whole transcriptome sequencing (WTS) and whole genome sequencing (WGS) were performed to clarify the gene rearrangements in transcriptional and genomic levels.

**Results** The patient was diagnosed with BCP-ALL and the blast cells were partially positive for CD13 and CD33, but negative for CD10. The karyotype was normal and 36 fusion genes frequently reported in hematological malignancies were negative. She achieved a complete remission (CR) after induction chemotherapy but relapsed 17 months later. t(9;22)(q34;q11.2) translocation was identified by karyotyping with the ratio of 50% but *BCR-ABL1* fusion was not detected by RT-PCR. FISH testing using *BCR-ABL1* dual color, dual fusion probe showed fusion signals with a lower percentage (6%). WTS revealed triadic fusions among 3 genes: *EWSR1* exon 7-*ZNF384* exon 7 in-frame fusion; *ZNF384* exon 6-*EHMT1* exon 2 out-frame fusion; *EHMT1* exon1-*EWSR1* exon 8 out-frame fusion. *EWSR1-ZNF384* is deemed to be the pathogenic fusion. RT-PCR and Sanger sequencing confirmed the 3 fusion transcripts both in relapsed and diagnostic BM. WGS revealed breakpoints in *EWSR1* intron 7, *ZNF384* intron 6, and *EHMT1* intron 1 in both relapsed and diagnostic BM. Structural variations detected by WGS were further analyzed to decipher the *BCR-ABL1* FISH fusion signals. A breakpoint in 9q34.11 located upstream of *ABL1* and a breakpoint in 22q11.23 located between *BCR* and *EWSR1* were found in both diagnostic and relapsed BM. The secondary translocation caused by these two breakpoints resulted in a subclone with seemingly normal karyotype and FISH fusion signals due to ectopically proximity of *BCR* and *ABL1*. The patient underwent allo-HSCT after chemotherapy and CD19 CAR-T cell therapy and has survived for more than 13 months.

**Conclusion** we deciphered a novel pathogenetic *EWSR1-ZNF384* fusion isoform and the complicated chromosomal translocations resembling Ph chromosome in a pediatric BCP-ALL patient. Integrated genomic analysis is crucial in elucidating the pathogenic and rare fusion pattern. Further investigation is needed to elucidate the oncogenic properties of *EWSR1-ZNF384* and confirm the clinical and biological features of patients harboring this fusion.

# PgmNr 2447: Optical mapping for chromosomal abnormalities: A pilot feasibility study for clinical use.

## Authors:

G.A. Toruner; Z. Tang; G. Tang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Hematopathology, UT MD Anderson Cancer Center, Houston, TX.

**BACKGROUND:** Chromosome analysis, FISH and microarrays are three major methods used in clinical cytogenetics labs. Each method has its own limitations. Optical mapping (OM) is a technique which uses fluorescence- labeling (with a site-specific DNA binding protein) large genomic DNA fragments, then pass through nano channels for linearization and get imaged. The fluorescent tags provide a bar code that allows subsequent assembly of individual molecules into larger contiguous maps, which are compared to a reference genome to identify numerical and structural chromosomal abnormalities with high resolution.

**METHODS:** OM was performed with Saphyr™ system from Bionano Genomics on five bone marrow specimens with hematological malignancies. We compared the results from OM and the clinically reported results from karyotyping (4 cases), FISH (4 cases) and microarray analysis (1 case). (Table 1).

**RESULTS:** Overall, clinically reported results and OM results were concordant (Table 1). In addition, several novel structural and copy number changes were detected by OM. .

**CONCLUSION:** OM is a powerful technology that enables us to “scan” the whole genome for both numerical changes and balanced rearrangements with high resolution by one assay. Besides, it detects additional chromosomal abnormalities which could be of clinically significance. Confirmation of additional abnormalities detected by OM and further testing with larger cohorts for validation is warranted before it could be utilized in clinical service.

Table 1. Cytogenetics and OM results in five BM samples

Case	Dx	Cytogenetics Results	Aberrations confirmed by OM
1	AML	46,XY[20].nuc ish(RUNX1x2)[200].nuc.ish(CBFBx2)	N/A
2	AML	46,XY,del(1)(p32p35),-4,del(6)(q13q26),add(7)(p22),r(7)(p13q11.2),add(17)(p11.2),+21[20].arr[GRCh37]1p36.21p35.3(14290193-29085067)x1,(4)x1,(7)x1,(21)x3	Monosomy 4 & 7 Trisomy 21 Del 1p36.21-p35.3 Del 17p13.2-17p13.1
3	B-ALL	47,XX,+mar[6]/46,XX[14]. nuc ish(3'CRLF2x2,5'CRLF2x3)(3'CRLF2 con 5'CRLF2x1)[78/200]	t(X;14)
4	Myeloma	Extensive numerical and structural changes[3]/46,XX[17]. nuc.ish(CDKN2Cx1,CKS1Bx3)[116/200] /(CDKN2Cx2,CKS1Bx5)[39/200]. nuc ish(RB1x1,13q34x2)[93/100]/(RB1x2,13q34x4)[19/100]/(RB1x2,13q34x3)[11/100]	Dup 1q21.3-q44 Del 13q11-q31.1
5	CLL	nuc ish(ATM, TP53x2)[200](D12Z3x3,D13S319x2,LAMP1x2) [126/200]	Trisomy 12



## **PgmNr 2448: PAP technology: Ultrahigh sensitivity and specificity in detection of nucleotide variants using cell-free DNA from liquid biopsies.**

### **Authors:**

C. Boysen <sup>1</sup>; Q. Liu <sup>3</sup>; S. Sommer <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) SomicsMed, Azusa, CA; 2) MedGomics, Azusa, CA; 3) Genetic Tools, Upland, CA

---

PCR and NGS based technologies are being developed for mutation detection in cfDNA or ctDNA from liquid biopsies. However, they are often not sensitive or specific enough for early detection of cancer and other applications, where a minimum amount of template with specific nucleotide variants, such as single base substitutions, insertions, and deletions, are present in great excess of wildtype genome.

The Pyrophosphorolysis-activated polymerization (PAP) methodology is a more affordable, next-generation nucleic acid amplification technology beyond PCR that possesses ultrahigh sensitivity and selectivity in detection of such mutations without false positives in non-invasive liquid biopsies from patients. The high specificity of PAP derives from the serial coupling of activation of a 3' blocked pyrophosphorolysis-activable oligonucleotide with extension of the unblocked, activated oligonucleotide.

Here we demonstrate the detection of specific cancer mutations in KRAS, EGFR, and BRAF at very low frequencies and down to just one molecule of mutated DNA without detection of false positives.

## PgmNr 2449: RNA-seq identifies structural variants in hereditary cancer genes.

### Authors:

B. Conner <sup>1</sup>; M. Richardson <sup>2</sup>; F. Hernandez <sup>2</sup>; T. Landrith <sup>1</sup>; T. McBride <sup>2</sup>; B. Tippin-Davis <sup>1</sup>; R. Karam <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Research & Development, Ambry Genetics, Aliso Viejo, CA; 2) Clinical Diagnostics, Ambry Genetics, Aliso Viejo, CA

**Introduction:** Massively parallel RNA sequencing (RNA-seq) is able to identify structural variants (SV) in *MSH2*, a frequent site for this type of alteration due to its enrichment with intronic Alu elements<sup>1,2</sup>. We have expanded upon this previously reported *MSH2* cohort to investigate, by RNA-seq, other SV in additional Lynch Syndrome (LS) and hereditary breast and ovarian cancer (HBOC) genes.

**Methods:** RNA-seq was performed as described previously in the whole blood of patients identified with germline SV<sup>1</sup>. Reads supporting aberrant splicing of the involved exons were used as one line of evidence for variant classification using ACMG/AMP guidelines<sup>3</sup>. All study participants consented to RNA genetic testing on a research basis.

**Results:** A set of 15 SV were reclassified from variant of unknown significance (VUS) to clinically actionable pathogenic/likely pathogenic variants (LP) or likely benign (LB). RNA genetic testing provided evidence for reclassification of SV in the following HBOC genes: *BARD1* (n=1), *BRCA1* (n=3), *BRIP1* (n=1), *CHEK2* (n=2), *PTEN* (n=1), *PALB2* (n=1), *RAD50* (n=1). Nine exonic duplications had reads supporting aberrant splicing and were determined to be in-tandem, resulting in their reclassification from VUS to LP. *RAD50* EX8dup was found to have a mid exonic breakpoint and there was no evidence of aberrant splicing by RNA-seq. This evidence was used to reclassify the variant from VUS to LB.

The previously reported *MSH2* duplication cohort was expanded upon by adding *MSH2* EX3\_6dup. SV in the following LS genes were also analyzed: *MLH1* (n=3), *PMS2* (n=1). RNA genetic testing identified aberrant splicing in all five LS SV, leading to their reclassification from VUS to LP.

**Discussion:** RNA-seq has clinical utility as RNA evidence carries strong weight in the ACMG/AMP variant classification guidelines<sup>3</sup>. This proves to be a high-throughput approach by which variants can be reclassified from VUS to either benign or pathogenic/likely pathogenic. Such reclassifications empower clinicians to offer the appropriate clinical management to LS and HBOC patients.

### References:

1. Conner et al 2019
2. Mu et al 2019
3. Richards et al 2015

## PgmNr 2450: Structured narrative of functional assays to support the determination of damaging effect on protein function.

### Authors:

M.W. Wright <sup>1</sup>; M. Iacocca <sup>1</sup>; A. Lopez Pineda <sup>1</sup>; H. Wand <sup>1</sup>; B. Wulf <sup>1</sup>; H. Costa <sup>1</sup>; R.Y. Patel <sup>2</sup>; A. Milosavljevic <sup>2</sup>; D. Ritter <sup>2</sup>; C. Alves <sup>3</sup>; J. Chora <sup>3</sup>; M. Bourbon <sup>3</sup>; S.E. Plon <sup>2</sup>; C.D. Bustamante <sup>1</sup>; T.J. Montine <sup>1</sup>; ClinGen's Cardiovascular Familial Hypercholesterolemia Variant Curation Expert Panel

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanford University School of Medicine, Stanford, California; 2) Baylor College of Medicine, Houston, Texas, USA; 3) Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisbon, Portugal

---

There is an important need to understand the functional consequence of a specific genomic change to a gene or protein when interpreting a variant in the context of pathogenicity determination. The joint guidelines of the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) requires the assessment of well established in vivo or in vitro functional studies showing no damaging effect (BS3) or supportive of damaging effect (PS3) on protein function or splicing.

The Clinical Genome Resource (ClinGen)'s Familial Hypercholesterolemia (FH) Variant Curation Expert Panel manually curated a comprehensive list of 128 peer-reviewed publications describing functional studies of variants identified in FH patients. This body of literature identified 305 variants in three genes: APOB (19 variants), LDLR (263 variants), and PCSK9 (23 variants). We reviewed the diversity of biochemical experiments, and heterogeneity of results with the goal to find a pattern that could promote consistency across the aggregated results.

We propose the use of structured narratives with the structure of 1) method, 2) material, and 3) effect (with or without a quantifier), using standardized terminology from BioAssay Ontology (BAO), Methods and Materials Ontology (MMO), Cell Line Ontology (CLO), and Gene Ontology (GO). For example, a given publication could refer to the use of the luciferase reporter gene assay (method assigned to BAO:0002661) with Hep G2 cells (material assigned to CLO:0003704) finding 12% gene expression (effect assigned to GO:0010467, with a quantifier of 12%). For qualitative findings (i.e. abnormal transcript length), the quantifier would not be used.

We found that only 13 terms were used to describe methods, 17 for materials, and 22 for effects. For a panel of experts, like the ClinGen's FH-VCEP, deliberating about the correct assignment of a reduced number of terms is far less challenging than being familiar with all available literature.

Our structured narrative have the potential to be used by other expert panels and working groups in the ClinGen ecosystem. With the proper validation and expansion of existing ontologies, we believe that our framework can accelerate the curation of functional data in accordance with the BS3/PS3 criteria of the ACMG/AMP guidelines. In the future, with defined rules for numerical thresholds, this framework could pave the way for automated scoring of BS3/PS3.

## PgmNr 2451: Next generation sequencing to identify novel genetic variants of possible familial hypercholesterolemia patients in Taiwan.

### Authors:

T. Hsiao <sup>1</sup>; Y.M. Chen <sup>2</sup>; C.H. Lin <sup>1</sup>; W.W. Lin <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Research, Taichung Veterans General Hospital, Taichung City, Taiwan,; 2) Division of Allergy, Immunology and Rheumatology, Taichung Veterans General Hospital, Taiwan.; 3) Cardiovascular Center, Taichung Veterans General Hospital, Taichung, Taiwan

---

**Background:** Familial hypercholesterolemia (FH) is an autosomal dominant inheritable disease characterized with premature cardiovascular disease and high levels of low-density lipoprotein (LDL) despite intensive statin therapy. This study aimed to interrogate novel genetic variants of possible FH patients in Taiwan using a next-generation sequencing (NGS) platform.

**Methods:** We enrolled 89 possible FH participants (age  $50.9 \pm 13.8$  years, 45 female, 44 male) with Dutch Lipid Clinic Network score 3-5. A customized targeted amplification panel of the coding region of LDLR, PCSK-9 and APOB genes were performed with the Illumina iSeq 100 platform.

**Results:** In total, we identified 14 pathogenic (12 LDLR, 2 APOB) and 6 likely pathogenic variants (5 LDLR, 1 APOB & LDLR). Three variant of unknown significance (VUS) were also categorized. The mean LDL levels were  $137.6 \pm 69.2$  mg/dl after maximum dose of statin therapies. We found that LDL levels were significantly different among participants of pathogenic, likely pathogenic, benign or VUS groups ( $p = 0.045$  by Kruskal-Wallis test).

**Conclusions:** NGS-based methods can be used to identify heterozygous FH patients of Chinese Han ancestry in the possible FH patients by Dutch Lipid Clinic Network score. However, subjects with VUS deserve further investigations.

## PgmNr 2452: Dilated cardiomyopathy: Genomic and diagnostic approach.

### Authors:

L. Piherova <sup>1</sup>; H. Hartmannová <sup>1</sup>; K. Hodašová <sup>1</sup>; P. Melenovská <sup>1</sup>; D. Mušálková <sup>1</sup>; V. Stránecký <sup>1</sup>; M. Kubánek <sup>2</sup>; R. Dobrovolný <sup>1</sup>; S. Kmoch <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Research Unit for Rare diseases, First Faculty of Medicine, Charles University, Prague 2, Czech Republic; 2) Institute for Clinical and Experimental Medicine (IKEM), Department of Cardiology, Prague, Czech Republic

---

Cardiomyopathy is a disease of the heart muscle associated with a disorder of its function. This is a heterogeneous group of diseases with various clinical signs that can ultimately lead to heart failure. Subset of cardiomyopathies are genetically conditioned. In this group can be found congenital heart defects, muscular dystrophy and congenital myopathy, hereditary disorders of metabolism (e.g., diseases of the lysosomal and mitochondrial diseases,  $\beta$ -oxidation disorders of fatty acids, glycogenosis, etc.) and genetic syndromes, e.g., Rasopathies.

Paediatric cardiomyopathies are known for their high phenotypic and genetic variability and causal mutations are very rare in the population. The use of the whole exome sequencing is currently one of the most effective tools for elucidating the genotype of individual patients.

We have successfully used this method for studying of genetic architecture of dilated cardiomyopathy cohort where in about 80% of the 460 examined patients we have found probably causal variant for DCM. At present, we are also focusing on paediatrics forms of cardiomyopathy which are more complex and the clarification is below 50%.

Successful molecular biology diagnostic and helps to identify the risk of occurrence of the disease in the family and to ensure prenatal diagnosis. In selected cases, the phenotype can also be studied on cell models, which contributes to the understanding of the molecular mechanism of the disease and allows a more accurate interpretation of border clinical or laboratory findings.



## PgmNr 2453: Identifying causal genes from abnormal clinical microarray results in patients with heterotaxy-spectrum congenital heart defects.

### Authors:

A. Sridhar <sup>1</sup>; B.J. Landis <sup>1</sup>; L.R. Elmore <sup>1</sup>; V. Garg <sup>2</sup>; G.C. Geddes <sup>3</sup>; S.R. Lalani <sup>4</sup>; J.H. Lin <sup>5</sup>; C.W. Lo <sup>5</sup>; K.L. McBride <sup>2</sup>; S.B. Wechsler <sup>6</sup>; S.A. Yatsenko <sup>5</sup>; J.C. Hodge <sup>1</sup>; S.M. Ware <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Indiana University School of Medicine, Indianapolis, Indiana; 2) The Ohio State University College of Medicine and Nationwide Children's Hospital, Center for Cardiovascular Research, Columbus, Ohio; 3) Medical College of Wisconsin, Milwaukee, Wisconsin; 4) Baylor College of Medicine, Houston, Texas; 5) University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; 6) Emory University School of Medicine, Atlanta, Georgia

---

Heterotaxy is a rare congenital anomaly syndrome caused by abnormal left-right patterning in the developing embryo. Patients with heterotaxy exhibit a wide range of phenotypic features including abdominal situs anomalies and congenital heart defects (CHD). However, the genetic basis of heterotaxy is rarely identified clinically. We previously showed that 16% of heterotaxy patients carry damaging copy number variants (CNVs). To further identify candidate genetic loci, we used the Cytogenomics of Cardiovascular Malformation (CCVM) Registry, an ongoing multi-site study that prospectively enrolls participants with abnormal cytogenetic and echocardiogram findings. Of 1827 CHD cases in the registry, we identified 69 patients with heterotaxy-spectrum CHDs using a modified Botto cardiac classification system. A total 137 genetic alterations were detected in heterotaxy patients, including 44.9% gains, 40.6% losses, 2.9% complex losses+gains. Regions of homozygosity (ROH) are identified in 11.6% of patients (7.3% with multiple segments due to consanguinity; 4.3% with segmental uniparental disomy [UPD]). Ten well-known genomic disorders such as 22q11.2 deletion were found in 14.5% patients. CNVs spanning cilia or laterality genes, such as *DNAI2* and *NPHP1*, occurred in multiple unrelated patients, affecting 20.3% of participants; 8.7% of patients carry ROHs spanning known laterality genes including multiple occurrences of *DNAH5*, *DNAI1* and *PKHD1*. CNVs in 6 genetic regions suggested as novel candidate loci in previous studies were present in our cohort. In order to identify novel candidate genes in the unsolved cases, we utilized 32 clinically established heterotaxy genes as the training set for a ToppGene bioinformatics gene prioritization approach. We identified 18 candidate genes enriched for heterotaxy-related Gene-Ontology and Phenotype-Ontology terms including *LIFR*, *KIF26B* and *MSX1* ( $p < 0.01$ ). Taken together, these data demonstrate the spectrum of submicroscopic chromosome abnormalities causing heterotaxy: known genomic disorders, novel damaging CNVs, and ROH. We replicate previous associated novel loci and identify that CNVs encompassing genes known to cause laterality defects or cilia dysfunction may be important etiologic contributors. ROH resulting from UPD may be a previously unappreciated mechanism of heterotaxy. Finally, novel candidate genes for heterotaxy CHD can be identified using bioinformatics approaches to interrogate heterotaxy-associated CNVs.

## PgmNr 2454: Developing and validating a cost effective semi quantitative and quantitative polymerase chain reaction assay for the diagnosis of Williams syndrome in a developing country.

### Authors:

D.M. Ranaweera <sup>1</sup>; D.C. de Silva <sup>2</sup>; N. Panchananthan <sup>1</sup>; M. Kajan <sup>1</sup>; D. Samarasinghe <sup>3</sup>; S. Perera <sup>3</sup>; R. Morawakkorala <sup>3</sup>; S. Gunewardene <sup>4</sup>; N.V. Chandrasekharan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Chemistry, University of Colombo, Colombo 3, Sri Lanka; 2) Department of Physiology, University of Kelaniya, Sri Lanka; 3) Lady Ridgeway Children's Hospital, Colombo 8, Sri Lanka; 4) Department of Physics, University of Colombo 3, Sri Lanka

---

**Introduction:** Williams-Beuren syndrome (WBS), caused by the loss of 23 genes including elastin (*ELN*) on chromosome 7q11.23, is associated with heart malformations, developmental delay, hypercalcaemia and characteristic facies. Genetic testing enables confirmation of diagnosis, improves management and genetic counselling. Sri Lanka (average household income \$115/month) has a free health service but genetic testing costs are borne by parents, making it mostly unaffordable.

**Materials and Methods:** Suspected WBS cases (n=24) were recruited following ethical clearance and informed consent. Molecular diagnosis was performed using semi quantitative polymerase chain reaction (semi-qPCR) (target *ELN*; control *CFTR*), quantitative PCR (qPCR) (target *ELN*; control *TES*) using dyes SYBR green and EvaGreen; 10 results (n=10) validated by fluorescent *in-situ* hybridization.

**Results:** Among 19/24 (79%) clinically typical cases, an *ELN* deletion was identified by semi-qPCR and qPCR. Nine cases had their diagnosis confirmed by FISH. Five clinically atypical cases were not found to have a deletion (2 familial cases, one isolated SVAS, two unknown) on semi-qPCR and qPCR (EvaGreen dye). There was a discrepancy for two patients with qPCR (SYBR green dye): one duplication (FISH negative); one deletion (parents declined FISH). The cost of consumables for semi-qPCR, qPCR and FISH was around \$6, \$7 and \$225 respectively.

**Conclusion:** The semi-q and qPCR using EvaGreen, although not the gold standard, appear to offer a molecular diagnostic test that is more affordable for Sri Lankan families and when used with clinical data, is a useful diagnostic tool in a resource limited healthcare setting.

Acknowledgements- University of Colombo Research Grant AP/3/2/2014/RG/02

## PgmNr 2455: Characterisation of copy number variants from exomes of patients with inherited heart disease and sudden cardiac death.

### Authors:

E.S. Singer <sup>1</sup>; S.B. Ross <sup>1,2</sup>; J.R. Skinner <sup>3,4</sup>; J. Ingles <sup>1,2,5</sup>; C. Semsarian <sup>1,2,5</sup>; R.D. Bagnall <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Agnes Ginges Centre for Molecular Cardiology at Centenary Institute, The University of Sydney, Sydney, NSW, Australia; 2) Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia; 3) Cardiac Inherited Disease Group New Zealand, Green Lane Paediatric and Congenital Cardiac Services, Starship Children's Hospital, Auckland, New Zealand; 4) Department of Paediatrics Child and Youth Health, The University of Auckland, Auckland, New Zealand; 5) Department of Cardiology, Royal Prince Alfred Hospital, Sydney, NSW, Australia

---

A genetic diagnosis aids clinical management of families with inherited heart disease and sudden cardiac death. Although exome sequencing has revolutionised the genetic testing of these diseases, for many families the genetic cause is not found. Computational tools can detect copy number variants, such as large deletions and duplications, by examining exome sequencing read depth. We sought to look for clinically actionable copy number variants in a large cohort of patients with inherited heart disease and sudden cardiac death. We performed exome sequencing on DNA isolated from patients with inherited cardiomyopathy, arrhythmia syndromes, or sudden cardiac death. We used eXome-Hidden Markov Model (XHMM) to detect copy number variants from exomes that were batched according to the enrichment kit. We focused on copy number variants located within a panel of 369 cardiac genes. Breakpoints were mapped by 1) viewing exome read alignments at the predicted breakpoint junctions, 2) confirming and resolving the extent of the variants with quantitative-PCR, and 3) performing PCR and Sanger sequencing to define the precise breakpoint junction. Segregation analysis was performed on affected family members, where available. Analysis of exome data of 701 patients revealed 10 (1.4%) potentially relevant copy number variants in 4/56 with left ventricular noncompaction, 1/257 with hypertrophic cardiomyopathy, 1/48 with dilated cardiomyopathy, 1/20 with long QT syndrome, 1/14 with arrhythmogenic cardiomyopathy, 1/141 with sudden cardiac death, and 1/1 with unicuspid aortic valve. Of significance, a 9.3 kb deletion spanning exons three to six of *ACTN2* was found in a female with LVNC, and in an unrelated male with HCM. Sequencing revealed that the breakpoint junction was identical in both families and likely resulted from non-homologous end joining. A female with sudden unexplained death at age 17 years had a 1.1 Mb deletion spanning 24 genes, including the long QT syndrome gene *KCNH2*. Finally, a male with HCM has a 29.4 kb duplication involving part of *MYH6* and *MYH7*, which likely occurred by non-allelic homologous recombination between a duplicated segment within the two genes. Copy number variants are a rare cause of inherited heart disease. Characterising the precise breakpoints of copy number variants enables cascade genetic testing of family members and improves our understanding of the molecular mechanisms leading to large deletions and duplications.

## PgmNr 2457: Reanalysis of genomic data for sudden-death cases reveal new diagnoses and give greater context for variant of unknown pathogenicity.

### Authors:

E. Salfati <sup>1</sup>; E. Spencer <sup>1</sup>; S. Topol <sup>1</sup>; E. Muse <sup>1,2</sup>; J. Lucas <sup>3</sup>; G. Wagner <sup>4</sup>; S. Campman <sup>4</sup>; E. Topol <sup>1,2</sup>; A. Torkamani <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Scripps Research, La Jolla, California; 2) Division of Cardiology, Scripps Clinic, La Jolla, California; 3) Los Angeles County Department of Medical Examiner-Coroner, Los Angeles, California; 4) San Diego County Medical Examiner's Office, San Diego, California

---

Around 10% of natural death cases remain unexplained each year and are subsequently classified as sudden unexpected deaths (SUD). The use of exome sequencing-based (WES) molecular autopsy protocol has recently proven its potential to improve this yield. As genomic sequencing expands and knowledge of gene?disease and variant?disease associations increase, reanalysis of initial WES data may facilitate new discovery and maximize the accuracy of clinical sequencing.

We tested this hypothesis in the Scripps Molecular Autopsy study, where in 2017 we reported a diagnostic yield of 16% through WES of 50 sudden-death cases. We re-assessed these post-mortem genetic data ~2 years later using an improved variant classification methodology and updated variant annotation pipeline.

Reanalysis yielded one pathogenic variant that was not initially reported in one individual, comprising an 18% increase in P/LP yield. We also reclassified variants of uncertain significance (VUS) to benign due to updated population frequency data. Explanatory variants have been discovered in 9 of 50 previously SUD cases. Our re-annotation efforts also classified another 30 (60%) cases with uncertain clinical significance finding (VUS), an almost 10% increase from our original study. These new genetic diagnoses are mainly due to improved variant prioritization, newly available clinical information since our original publication and highly accurate annotation database.

Iterative reanalysis of negative WES data may prove useful, as knowledge of genetic etiologies grows. In addition, these findings underline the importance of systematic molecular autopsy to cases with an undetermined manner of death, with implications for screening of relatives.

## PgmNr 2458: Revisiting the genetic testing diagnostic yield in familial hypertrophic cardiomyopathy.

### Authors:

R.D. Bagnall<sup>1,2</sup>; J. Ingles<sup>1,2,3</sup>; N. Nowak<sup>1</sup>; C. Semsarian<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Agnes Ginges Centre for Molecular Cardiology at Centenary Institute, The University of Sydney, New South Wales, Australia; 2) Faculty of Medicine and Health, The University of Sydney, New South Wales, Australia; 3) Department of Cardiology, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia

---

Hypertrophic cardiomyopathy (HCM) is an inherited disease of the myocardium characterised by unexplained left ventricular hypertrophy. Recent expert curation has classified 23 HCM disease genes as having moderate, definitive, or strong evidence of causing isolated left ventricular hypertrophy. Consequently, previously reported pathogenic variants in genes with lower, or no, evidence of association with HCM are now considered unlikely to be responsible for disease causation. We sought to reassess the genetic testing diagnostic yield of familial HCM in expert curated genes using stringent variant classification criteria.

We re-evaluated genetic testing results in unrelated probands with a family history of HCM who had undergone cardiac gene panel, exome or genome sequencing. Variants were evaluated in genes with definitive or moderate association with HCM (*MYBPC3*, *MYH7*, *TNNT2*, *TNNI3*, *TPM1*, *ACTC1*, *MYL2*, *MYL3*, *CSRP3*, *TNNC1*, *JPH2*), and in syndromic genes where isolated left ventricular hypertrophy may be seen (*ALPK3*, *CACNA1C*, *DES*, *FHL1*, *FLNC*, *GLA*, *LAMP2*, *PRKAG2*, *PTPN11*, *RAF1*, *RIT1*, *TTR*). Variants with frequency < 0.004% in the Genome Aggregate Database of population controls were classified for pathogenicity using American College of Medical Genetics guidelines.

Amongst 229 familial HCM probands, 147 (64%) had a pathogenic or likely pathogenic variant, of which all but 5 variants were in definitive HCM genes. 33 (15%) probands harboured only a variant of uncertain significance, of which 24 were in definitive genes and 9 were in syndromic genes. 49 (21%) probands had no rare variants and were classified as having an indeterminate genetic test. When compared to familial HCM probands with a pathogenic or likely pathogenic variant, those with an indeterminate test were older at diagnosis (mean age 44 versus 33 years;  $p < 0.001$ ), had fewer affected family members (mean 2.4 versus 3.5;  $p < 0.001$ ) and a lower maximal left ventricular wall thickness (mean 18 mm versus 21 mm;  $p < 0.001$ ). There were no differences between the two groups with respect to a family history of sudden cardiac death or sex.

Patients with familial HCM and an indeterminate genetic test are older at diagnosis and have less hypertrophy than those with a clinically actionable variant. While a non-familial HCM sub-group is already known, further sub-classification of familial HCM may be warranted. Future research should focus on resolving the pathogenicity of variants of uncertain significance.

## PgmNr 2459: The genetic autopsy: A tool to provide insight and assess risks for surviving family members.

### Authors:

D. Macaya <sup>1</sup>; T.R. Hart <sup>1</sup>; P.L. Arscott <sup>2</sup>; L. Haskin-Leahy <sup>3</sup>; J. Slamon <sup>4</sup>; M.L. Murray <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) GeneDx Inc, Gaithersburg, MD.; 2) University of Michigan Departments of Internal Medicine and Pediatrics. Frankel Cardiovascular Center, Ann Arbor, MI; 3) University of Vermont Medical Center Clinical Genetics Program, Burlington, VT; 4) Division of Maternal Fetal Medicine, Vanderbilt University Medical Center, Nashville, TN

---

Sudden death is a traumatic event, often leaving surviving family members with unanswered questions regarding the cause of death and their personal risks for cardiac events. We performed multigene cardiac genetic testing in nearly 800 postmortem cases. Positive diagnostic results were obtained in 14%. To highlight the clinical utility of postmortem genetic testing (PMGT), we present three case examples, in which sudden death precipitated familial genetic testing and provided critical information for family members. **Case 1:** A teen female with a history of prolonged QTc interval and family history of syncope died suddenly in her sleep. A living sibling was later diagnosed with LQTS, and genetic testing identified a novel *KCNH2* missense variant, which segregated with LQTS in several affected relatives, including the deceased proband. **Case 2:** An adult male with DCM died suddenly in his early 40s, and his sister passed way in her 30s while sleeping. PMGT identified a pathogenic *LMNA* nonsense variant. Familial testing revealed that the proband's affected mother is mosaic for the variant. Multiple siblings of the proband are pre-symptomatic carriers and are at risk to develop and transmit an *LMNA*-related cardiomyopathy. **Case 3:** A male infant presenting with dyspnea, hypotonia, and metabolic acidosis shortly after birth was found to also have moderate biventricular dilatation with markedly reduced function, ultimately dying from cardiac arrest. Autopsy findings were consistent with DCM. PMGT revealed a maternally inherited pathogenic *TAZ* missense variant associated with X-linked Barth syndrome. In conclusion, PMGT may establish or confirm a clinical diagnosis in the proband, which allows for testing of at-risk family members and determine recurrence. Genetics professionals complement the medical examiner's services by facilitating genetic testing to obtain answers for family members. They also play a vital role in providing counseling and medical management options for surviving relatives. As state laws begin to mandate protocols for PMGT as part of the standard investigation when sudden death occurs under age 50 years, it will become increasingly important for genetics health care providers to stay abreast of both common and unique findings that may arise during these investigations.



## PgmNr 2460: Founder effects and genotype/phenotype correlations in homozygous cases of autosomal recessive congenital ichthyosis.

### Authors:

H.C. Hennies<sup>1,2,3</sup>; D. Lima Cunha<sup>1,2,3</sup>; K.M. Eckl<sup>1,3,4</sup>; R. Gruber<sup>5</sup>; N. Kakar<sup>6,7</sup>; J. Ahmad<sup>7</sup>; S. Alawbathani<sup>2</sup>; B. Krabichler<sup>3</sup>; J. Altmüller<sup>2</sup>; P. Nürnberg<sup>2</sup>; J. Zschocke<sup>3</sup>; G. Borck<sup>6</sup>; M. Schmuth<sup>5</sup>; A.S. Alabdulkareem<sup>8</sup>; K. Abdulaziz Alnutaifi<sup>9</sup>; O.M. Alakloby<sup>9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Biological and Geographical Sciences, University of Huddersfield, Huddersfield, UK; 2) Cologne Center for Genomics, University Hospital Cologne, Cologne, Germany; 3) Division of Human Genetics, Medical University of Innsbruck, Innsbruck, Austria; 4) Dept. of Biology, Edge Hill University, Ormskirk, UK; 5) Dept. of Dermatology, Medical University of Innsbruck, Innsbruck, Austria; 6) Institute of Human Genetics, University of Ulm, Ulm, Germany; 7) Dept. of Biotechnology, BUIITEMS, Quetta, Pakistan; 8) King Saud Medical City, Riyadh, Saudi Arabia; 9) Dept. of Dermatology, College of Medicine, Imam Abdulrahman Bin Faisal University Dammam, Saudi Arabia

---

Autosomal recessive congenital ichthyosis (ARCI) is a phenotypically and genetically heterogeneous skin disease. It is an ultra-rare disease with less than one patient in 50,000 people and can be caused by mutations in more than twelve different genes, with approximately 20% of the cases without mutations in a currently known candidate gene. The main pathophysiological feature is an impaired skin permeability barrier function, which leads to a disturbance of the cutaneous water homeostasis. A clear correlation between genetic causes and clinical picture has not been described to date, in part because of pronounced allelic heterogeneity and a vast majority of compound heterozygous cases. We have therefore collected cases from populations with a large percentage of consanguineous families, mainly from Saudi Arabia, Yemen, and Pakistan. We have assessed various approaches to identify candidate genes and mutations, including SNP-based homozygosity mapping, gene panel sequencing and whole exome sequencing. Interestingly, we found likely pathogenic variants in known candidate genes in all 19 families studied here, and these variants were identified in only five different genes, namely *TGM1*, *ABCA12*, *CYP4F22*, *NIPAL4*, and *ALOXE3*. Variants included both known and previously unknown mutations and nonsense, splice site, and missense variants. Patients from Saudi Arabia and Pakistan were assigned to potential founder mutations in *TGM1*, *ALOXE3*, and especially *NIPAL4*, respectively. Importantly, based on detailed clinical data including recent treatment information, we defined clear genotype/phenotype correlations. We attributed *TGM1* and *ABCA12* mutations to the most severe forms of lamellar and erythematous ichthyoses, respectively, almost regardless of treatment, and identified the phenotypic spectrum of ARCI associated with variants in each of the other genes. A detailed clinical analysis will be completed and specified. Our results contribute to the mutational spectrum of ARCI and revealed significant insights into genotype/phenotype correlations. The findings are instrumental for defining a fast and convenient procedure to make a diagnosis considering the patient background and the laboratory facilities available.

## PgmNr 2461: Somatic mutation spectrum of segmental overgrowth.

### Authors:

S.H. Heo <sup>1</sup>; Y.M. Kim <sup>2</sup>; A. Oh <sup>3</sup>; Y. Lee <sup>3</sup>; G.H. Kim <sup>4</sup>; T. Kim <sup>1</sup>; I.H. Choi <sup>4</sup>; J.H. Jang <sup>5</sup>; H.W. Yoo <sup>3</sup>; B.H. Lee <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Asan Institute for Life Sciences, Asan Medical Center, College of Medicine, University of Ulsan, Seoul, Korea; 2) Department of Pediatrics, Gangneung Asan Medical Center, Gangneung, Korea; 3) Department of Pediatrics, Asan Medical Center Children's Hospital, College of Medicine, University of Ulsan, Seoul, Korea; 4) Medical Genetics Center, Asan Medical Center, College of Medicine, University of Ulsan, Seoul, Korea; 5) Green Cross Genome, Yongin, Korea

Congenital overgrowth of body segment is noted in several genetic syndromes such as Klippel-Trenaunay (KT) syndrome, Cowden syndrome, Megalencephaly-capillary malformation (MCAP) syndrome, CLOVES (or CLOVE) syndrome and fibroadipose hyperplasia, and Parker-Weber syndrome. These syndromes share the common phenotypic findings, somatic segmental overgrowth (brain, limb or visera), fibro-adipose, lymphatic, vascular or muscular proliferation, and cutaneous nevi. Recently, PIK3CA-related signaling pathway has been suggested as the main molecular pathogenic pathway of these syndromes; The somatic mutations of the PIK3CA or its related genes have been identified and its target agent improved the segmental overgrowth in the affected patients.

In the current study, we designed the customized panel of 145 genes (372 kbps) related to PI3K/AKT/mTOR pathway, in order to identify the somatic mutations in the 12 patients suspicious of PIK3CA-related segmental overgrowth syndrome. DNA was identified both in the affected tissues (cutaneous vascular or lymphatic proliferative tissue) and normal blood leukocytes. The Mean depth of the 145 genes was 711X and 97.5% of the target genes were read >100X. PIK3CA mutations were found in 6 patients (MAF=2-18% of the affected tissues and 0% in the leukocytes). In addition, KRAS, MAP2K3, TBC1D4 and GNAQ somatic mutations were found in single cases (MAF=4%). One patient had a germline PTEN mutation. All of these variants were not found in normal population (MAF <0.002%).

With our customized panel, we identified somatic genetic alteration in a total 10 patients (83%). This genetic information would help to apply therapeutic target agent in select cases and understand the molecular pathology more in detail.



## PgmNr 2462: Exome-sequencing identifies a novel 3.6 Kb deletion from PYGL as the cause of glycogen storage disease type VI.

### Authors:

Y. Lu <sup>1</sup>; B. Liu <sup>1</sup>; K. Liu <sup>2</sup>; B. Wu <sup>1</sup>; H. Wang <sup>1</sup>; W. Zhou <sup>1</sup>; We wish Y. Lu and B. Liu to be cited as 'co-author'

[View Session](#) [Add to Schedule](#)

### Affiliations:

- 1) Center for Molecular Medicine, Children's Hospital of Fudan University, shanghai, Shanghai, China;
- 2) Digital China Health Technologies Co. Ltd., Beijing, Beijing, China

---

### Background:

Glycogen storage diseases type VI (GSD VI) is an autosomal recessive condition that presents in infancy/early childhood with hepatomegaly and hypoglycemia. Because of its wide spectrum of clinical manifestations, it is difficult to distinguish GSD VI from other liver GSDs. By now, several pathogenic single nucleotide variants (SNVs) and large size copy number variations (CNVs) on the glycogen phosphorylase L (PYGL), the only gene known to be associated with GSD VI, were identified. However, due to the limitations of detection methods and the rarity of the variations, the detection of some pathogenic point mutations can be missed, and pathogenic structure variants (SVs) have not been reported yet.

### Purpose:

In this study, we investigated whether there exist SVs on the PYGL gene that cause GSD6 by means of a new exome sequencing analysis.

### Methods:

Based on exome sequencing data, a rapid bioinformatics method of SAM file analysis searching for potential SVs was constructed. This method was applied in 16,247 unrelated samples for SV screening, including 9,708 samples performed clinical exome sequencing (CES) and 6,539 samples performed whole exome sequencing (WES). Detected potential SVs were validated by long-rang PCR.

### Results:

A novel 3.6 Kb deletion variant which involved the exons 14-17 of PYGL was identified. The deletion variant was detected in six samples, which suggested a potential mutation hotspot, with an estimated allele frequency of 0.018%. Three samples of them each had another different pathogenic SNV on the PYGL gene and clinical symptoms of hepatomegaly and high level of liver glycogen. The other three samples, due to no clinical phenotypes of GSD VI and undetected pathogenic mutations in the PYGL exon regions, were considered as carriers.

### Conclusions:

The identification of the novel PYGL exons deletion broaden the understanding of the genetic factors of GSD VI pathogenesis. In addition, this paper was also demonstrated the feasibility of exome sequencing data analysis in the search for SV. The three samples with clinical symptoms of GSD VI and identified SV suggested the limitations of SNV-alone analysis with exome sequencing detection.

## PgmNr 2463: Exonic duplication of the *OTC* gene by a complex rearrangement that likely occurred via a replication-based mechanism: A case report.

### Authors:

K. Yokoi <sup>1,2</sup>; Y. Nakajima <sup>1</sup>; H. Inagaki <sup>2</sup>; M. Tsutsumi <sup>2</sup>; T. Ito <sup>1</sup>; H. Kurahashi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) pediatrics, Fujita health university, Toyoake, Aichi, Japan; 2) Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Japan

---

**Background:** Ornithine transcarbamylase deficiency (OTCD) is an X-linked recessive disorder involving a defect in the urea cycle caused by *OTC* gene mutations. Although a total of 417 disease-causing mutations in *OTC* have been reported, structural abnormalities in this gene are rare. We here describe a female OTCD case caused by an exonic duplication of the *OTC* gene (exons 1-6).

**Case presentation:** A 23-year-old woman with late-onset OTCD diagnosed by biochemical testing was subjected to subsequent genetic testing. Sanger sequencing revealed no pathogenic mutation throughout the coding exons of the *OTC* gene, but multiplex ligation-dependent probe amplification (MLPA) revealed duplication of exons 1-6. Further genetic analyses revealed an inversion of duplicated exon 1 and a tandem duplication of exons 2-6. Each of the junctions of the inversion harbored a microhomology and non-templated microinsertion, respectively, suggesting a replication-based mechanism. The duplication was also of *de novo* origin but segregation analysis indicated that it took place in the paternal chromosome.

**Conclusion:** We report the first OTCD case harboring an exonic duplication in the *OTC* gene. The functional defects caused by this anomaly were determined via structural analysis of its complex rearrangements.

# PgmNr 2464: Molecular diagnosis of autoimmune polyendocrine syndrome type 1 (APS1) and systemic literature review.

## Authors:

Y.J. Lee <sup>1,3,4,5,6</sup>; W.F. Chen <sup>2</sup>; W.H. Ting <sup>1,7</sup>; C.Y. Huang <sup>1,7</sup>; C.L. Lin <sup>6</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Pediatric endocrinology, MacKay Children's Hospital, New Taipei, Taiwan; 2) Clinical Pathology and Medical Technology, MacKay Memorial Hospital Tamsui District; 3) Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University; 4) Department of Medicine, Mackay Medical College; 5) Institute of Biomedical Sciences, Mackay Medical College; 6) Department of Medical Research, MacKay Memorial Hospital Tamsui District; 7) MacKay Junior College of Medicine, Nursing, and Management, Taipei, Taiwan

## Background

Autoimmune polyendocrine syndrome, type 1 (APS1), also called APECED (autoimmune polyendocrinopathy, candidiasis, ectodermal dystrophy) syndrome, is rare. The clinical diagnosis of APS1 requires the presence of two of the three cardinal components: chronic mucocutaneous candidiasis, autoimmune hypoparathyroidism, and autoimmune adrenal insufficiency. We reported a girl with APS1 with molecular diagnosis and systemic review of the literature.

## Case Presentation

A 15-year-old girl presented to the pediatric endocrinology clinic because of hyperpigmentation of the oral mucosa, gingiva, lips, skin, and knuckles. Physical examination showed she also had oral candidiasis and onychomycosis of the fingernails and toenails. Her morning serum cortisol was 5.39 µg/dL and ACTH >1390 pg/mL. Pathological examination of the nail of the big toenail revealed fungal infections. Her calcium was 10 mg/dL, phosphorus 4.5 mg/dL, total cholesterol 204 mg/dL, triglyceride 57 mg/dL, uric acid 4.5 mg/dL, creatinine 0.9 mg/dL, ALT 19 U/L, and AST 24 U/L. She was treated with cortisone acetate. At the age of 20 she had cramps of extremities with a serum calcium of 7.9 mg/dL and phosphorus of 7.1 mg/dL and began to take calcitriol and calcium carbonate.

## Methods

Molecular genetic analysis of her genomic DNA was performed on the autoimmune regulator (*AIRE*) gene using PCR and sequencing.

## Literature review

We searched PubMed using MeSH terms 'Autoimmune polyendocrine syndrome, type 1' and 'polyendocrinopathies, autoimmune'. We also screened for relevant references from all original and review articles by hand, Online Mendelian Inheritance in Man (OMIM), and Human Genetic Variation Database.

## Results

We found that the patient was a compound heterozygote with mutations c.32T>G (p.Leu11Arg) and c.483\_484delC (p.Lys164Argfs\*214). The two mutations were not found in 50 healthy individuals. C.32T is conserved in *Mus musculus*, *Macaca nemestrina*, *Gorilla gorilla*, *Pan troglodytes*, and *Homo sapiens*. Mutation c.32T>G is pathogenic by PolyPhen-2, SIFT, and LoFTool. Mutation c.483\_484delC is probably damaging by LoFTool.

Systemic literature review showed mutations are specific in ethnicity and of genotype-phenotype correlation.

**Conclusion**

Although APS1 is rare, alertness for early diagnosis is beneficial to patients for appropriate treatment to various symptoms. Molecular genetic diagnosis provides proper genetic counseling. Mutations show ethnic specificity and genotype-phenotype correlation.

## PgmNr 2465: Diagnostic yield for multi-gene testing in patients with a clinical diagnosis of nephrotic syndrome/focal segmental glomerulosclerosis.

### Authors:

P. Celestino Soper; E. Williams; L. Goodell

[View Session](#) [Add to Schedule](#)

**Affiliation:** GeneDx, Gaithersburg, Maryland, USA

---

Nephrotic syndrome (NS) is a clinically heterogeneous group of disorders characterized by proteinuria, hypertension, hypercholesterolemia and edema. The most prevalent disorder is focal segmental glomerulosclerosis (FSGS), which is characterized by scarring in a limited number of glomeruli of the kidneys. Multiple genes have been linked to NS, which may be inherited in an autosomal dominant (AD), recessive (AR) or X-linked manner. Currently, there are no clear guidelines for genetic testing of patients with NS/FSGS. Published studies estimate the diagnostic yield of genetic testing in patients with well-defined steroid resistant NS to be about 15-30%, however the yield in a broader/less defined phenotype has not been evaluated thoroughly. The purpose of this study was to review the results from clinician-built multi-gene NGS tests based on exome capture (XomeDxSlice) to determine the yield and validity of panel testing for NS/FSGS. Probands were selected based on the presence of NS, FSGS, glomerular disease and/or proteinuria, whose test included at least one gene from a list of 25 known to be strongly associated with NS/FSGS (tests reviewed had an average of 84 genes, range of 2 to 340 genes). Approximately 21% (18/85) of individuals had a positive diagnostic result. Of these, the majority (55.6%; 10/18) had pathogenic or likely pathogenic variant(s) (PV) in a gene for Alport syndrome (*COL4A3*, *COL4A4*, *COL4A5*). Two affected individuals had biallelic PV in AR genes (*NPHS1*, *SMARCAL1*), while the remaining 6 cases had PV in AD genes (*WT1*, *INF2*, *TRPC6* and *PAX2*). Another 48% (41/85) of individuals had a result of possible clinical importance, such as a single PV in an AR gene or variant(s) of uncertain significance in an AD or AR gene, where additional follow-up was recommended. Of these, 54% (22/41) had a result in at least one of the 25 strongly associated genes. The high number of probands with a molecular diagnosis of Alport syndrome in this study may indicate a considerable clinical overlap with NS/FSGS, thus illustrating the benefits of multi-gene testing for accurate diagnosis, appropriate management and follow-up. Genetic diagnosis may guide decisions regarding steroid therapy and kidney transplant candidacy, and assist with prognostic predictions. Additionally, increased utilization of genetic testing may, over time, allow for an expanded understanding of the genetic factors that cause NS/FSGS.

## PgmNr 2466: First report of late-onset Peutz-Jegher Syndrome (PJS) with APC gene mutation, and no alteration of PJS-related STK11 gene.

### Authors:

M. Ibrahim <sup>1</sup>; A. Ibrahim <sup>2</sup>; M.J. Hajianpour <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Internal Medicine, East Tennessee State University, Johnson City, Tennessee.; 2) School of Medicine, University of Jordan, Amman, Jordan; 3) Division of Medical Genetics, Department of Pediatrics, East Tennessee State University

Peutz-Jegher syndrome (PJS) is a rare autosomal dominant syndrome usually due to a germline mutation in STK11 gene. It is characterized by multiple hamartomatous polyps in the gastrointestinal tract with mucocutaneous pigmentation and has an increased risk of cancer. Currently only mutations in STK11 (AKA LKB1) have been identified as causative of PJS. The STK11 gene sequencing detects 55% of cases when the family history is positive for PJS, and 70% of sporadic cases.

Deletion/duplication analysis detects 45% of familial cases and 21% of sporadic cases. High penetrance rate by the age of 30 years, but 10-20% of the cases result from de novo mutations<sup>1</sup>. To our knowledge this is the first report of association of APC gene alteration and Peutz-Jegher syndrome (PJS).

A 56-year-old who presented to the emergency department with abdominal pain and rectal bleeding. No family history of PJS, but breast and colon cancer found in first degree relatives. He had dysphagia, nausea, vomiting and fatigue at presentation with pallor, multiple nevi/brown spots over the arms and legs and right palm, superficial tenderness in the right side of the abdomen on physical examination. Rectal examination revealed no masses or active bleeding at the time. Patient has renal cyst, iron- deficiency anemia and a history of recurrent rectal bleeding over the past four years. Upper GI endoscopy showed duodenal hamartomatous polyps; histologically P-J polyps with a stricture at the distal esophagus.

Colonoscopies performed over the past 8-10 years showed 2-10 polyps which were histologically confirmed to be benign tumors with atypical hyperplasia. Molecular studies showed a variant of uncertain significance (VUS) of APC gene: c2297C>T (p. Ala766Val) (aka A766V (2297C>T)). The APC gene is associated with familial adenomatous polyposis (FAP). The sequencing and deletion/duplication analysis of STK11 gene was normal.

A clinical diagnosis with PJS requires the presence of any one of the following:

Two or more histologically confirmed Peutz-Jeghers (PJ) polyps, any number of PJ polyps detected in an individual who has a family history of PJS in a close relative, characteristic mucocutaneous pigmentation in an individual who has a family history of PJS in a close relative, any number of PJ polyps in an individual who also has characteristic mucocutaneous pigmentation.

## PgmNr 2467: Cost-effectiveness analysis of using *TBX6*-associated congenital scoliosis risk score in genetic diagnosis of congenital scoliosis.

### Authors:

Z. Chen<sup>1,2,3</sup>; J. Liu<sup>1,2,6</sup>; Y. Zhang<sup>1,2,3</sup>; S. Zhao<sup>1,2</sup>; S. Wang<sup>1,2,4</sup>; Z. Wu<sup>2,4,5</sup>; G. Qiu<sup>1,2,4</sup>; J. Zhang<sup>1,2,4</sup>; N. Wu<sup>1,2,4</sup>; DISCO (Deciphering disorders Involving Scoliosis and COMorbidities) study group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China.; 2) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Beijing 100730, China.; 3) Graduate School of Peking Union Medical College, Beijing 100005, China.; 4) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences, Beijing 100730, China.; 5) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China.; 6) Department of Breast Surgical Oncology, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

### Background

This study tested whether using *TBX6*-associated congenital scoliosis risk score (TACScore) as a screening method is more cost-effective compared to using Whole Exome Sequencing (WES) as a first-line genetic testing for congenital scoliosis.

### Method

Molecular diagnosis of 416 patients in the DISCO (Deciphering disorders Involving Scoliosis and COMorbidities) cohort were collected retrospectively. A decision tree was constructed to estimate the money and time costs of two alternative strategies (TACScore versus WES). Bootstrapping simulations and sensitivity analyses were performed to examine distributions and the robustness of estimates. The economic evaluation adopted both health care payer and personal budget perspective.

### Result

The strategy of using 'TACScore' for primary screen resulted in a cost of \$1074.2 and diagnostic time of 38.7d in average to gain a molecular diagnosis or an uninformative ES report for each patient compared with \$1169.6 and 41.4d for the 'ES as a first-line test' method (incremental cost, -\$95.4, 95%CI: -\$124.9 to -\$66.0,  $P < 0.001$ ). Patients predicted by TACScore received a result within \$715.1 and 30.4d, significantly lower than \$1193.4 and 44.0d when ES served as a first-line test (incremental cost, -\$478.3, 95%CI: \$-599.1 to \$-357.5,  $P < 0.001$ ). In 100% of bootstrapping simulations, the strategy using TACScore was less costly and time-saving. The sensitivity analyses revealed that the strategy using TACScore remained cost-effective even the cost of per ES decreased to \$8.8.

### Conclusion

From both health care payer and personal budget perspectives, using TACScore as a screening test is more cost-effective than using ES as a first-line test, for pursuing a molecular diagnosis in patients with congenital scoliosis.

## PgmNr 2468: Establishment of next-generation sequencing based polydactyly gene panel.

### Authors:

J. Hong; K. Park; H. Kim; W. Lee; S. Chun; W. Min

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Laboratory Medicine, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea, Republic of

---

### Introduction

Polydactyly is a congenital condition characterized by extra digits, though its phenotype varies from skin tag to complex duplication. Polydactyly is an anomaly observed right after birth, and it can occur either alone or as a sign of various syndromes. The main etiology of polydactyly is known to be genetic. However, genetic testing for polydactyly is not routinely performed. We established a gene panel for polydactyly using next-generation sequencing (NGS), a method that can analyze many genes simultaneously.

### Method

We reviewed literatures and searched databases such as Human Gene Mutation Database and Online Mendelian Inheritance in Man to list up and select genes associated with polydactyly. BED file including exons and part of introns of selected genes was created and probes were designed. The reagents used were DxSeq Library Preparation Reagent and DxSeq Target Capture Reagent. Illumina Miseq DX was used for sequencing. Bioinformatics software employed were Illumina VariantStudio and DxSeq Gene Analysis System. Reporting methods followed standards and guidelines of the American College of Medical Genetics and Genomics. Designed panel was applied to NA12878 reference genome to validate and the result was compared to databases such as Genetic Testing Reference Materials Coordination Program and The Genome in a Bottle Consortium hosted by NIST.

### Result

Overall, 167 genes and ZRS, a region in intron of *LMBR1* gene, were included in the polydactyly panel. 503,063 base pairs of target region were assigned. Several regions that repeatedly generate low coverage due to deletion and GC bias were excluded from target and left to manual review and Sanger sequencing if needed. Sequencing and analyzing of NA12878 were performed according to the established panel and protocol. 205 previously reported variants in the target region were detected in all duplicated samples of NA12878, yielding analytical sensitivity of 100%. The match rate was calculated for the detected sequence and the high-confidence region of NIST, and the analytical specificity and accuracy were over 99.9%.

### Conclusion

To test many genes related to polydactyly at once, NGS is a suitable method. Some regions showed low coverage and made QC workflow difficult. We excluded the regions from target and prepared manual review process and Sanger sequencing to solve the problem. In conclusion, polydactyly NGS panel can be used for detecting causal genes of polydactyly and benefit patient management.





## PgmNr 2469: Validation study of the integration of next-generation phenotyping in exome analysis.

### Authors:

K.W. Gripp<sup>1,2</sup>; H.W. Moore<sup>3</sup>; G. Nadav<sup>1</sup>; Y. Hanani<sup>1</sup>; Y. Gurovich<sup>1</sup>; T. Hsieh<sup>4</sup>; P.M. Krawitz<sup>1,4</sup>; S.A. Skinner<sup>3</sup>; M.J. Friez<sup>3</sup>; R.J. Louie<sup>3</sup>; J.R. Jones<sup>3</sup>; S. Savage<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) FDNA Inc., Boston, MA; 2) Dept Gen, DuPont Hosp, Wilmington, Delaware.; 3) Greenwood Genetic Center, Greenwood, SC; 4) Institute of Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany

---

**Background:** The expanding use of next-generation sequencing technologies will increase the challenges in interpretation and prioritization of resulting genomic variants. Variant prioritization and clinical correlation require the integration of rich phenotypic data, which can be accomplished through the use of next-generation phenotyping (NGP) technologies. The Face2Gene platform (F2G) leverages NGP analysis of various phenotypic signals, including DeepGestalt for facial analysis. The PEDIA algorithm is the first framework to integrate F2G-generated phenotypic scores along with molecular data to enable variant prioritization based on phenotypic relevance.

**Methods:** Using F2G and the PEDIA algorithm, we analyzed facial photos, HPO features, and VCFs from 126 real-world cases sent for exome sequencing at the Greenwood Genetic Center (GGC). Cases were segmented by diagnosis status and difficulty level. For each diagnosed case, we generated a PEDIA score for each variant and gene. Using these scores, we defined the PEDIA rank for the known causative gene in each case. This PEDIA rank was compared to the original GGC variant rank and a molecular rank based on the CADD score.

**Results:** Integration of NGP into the variant prioritization process dramatically increased diagnostic efficiency, as demonstrated by the increase in top-ranked causative genes and the overall improvement in gene ranking. PEDIA ranked the causative variant first in 27.5% of cases, compared to 2.5% based on GGC rank and 7.5% based on CADD rank. Compared to GGC and CADD ranks, PEDIA resulted in a significantly improved rank ( $\geq 15$  positions) in 17.5% and 33% of cases, respectively. For cases originally solved only with analysis of parental data, PEDIA showed improved ranking in 25% of cases (average improvement 27.7 positions) and successfully placed the causative gene in the top 5 in 33%. In cases diagnosed with a syndrome supported by a facial model in F2G, PEDIA ranks the causative gene in the top 5 for 67% of cases, compared to 11% based on either GGC or CADD ranks.

**Conclusion:** Integrating NGP into the molecular diagnostic process increases diagnostic efficiency and yield. Further work may demonstrate this method's effectiveness as an alternative to trio analysis. As more NGP technologies are developed, the impact of NGP on the molecular diagnostic process will increase substantially.

## PgmNr 2470: 3q26.1-q29 microduplication: A case report.

### Authors:

R. Quero <sup>1,2</sup>; M. Manotas <sup>1,2</sup>; J. Villalta <sup>1,2</sup>; M. Guerra <sup>1,2</sup>; F. Suarez <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Instituto de Genética Humana, Facultad de Medicina, Pontificia Universidad Javeriana, Bogotá, Colombia.; 2) Hospital Universitario San Ignacio, Bogotá, Colombia

---

Microdeletion and microduplication syndromes involving many different chromosomal segments are significant factors implicated in mental retardation and dysmorphism. Overall, the detection rate of chromosomal aberrations in patients with ID and/or congenital malformations using conventional karyotyping is approximately 9.5%. The use of array comparative genomic hybridization (array CGH) as a diagnostic tool in molecular genetics has facilitated the identification of many new microdeletion/microduplication syndromes (MMSs).

We report a case of an 9-month-old child who was referred to our genetic consultation by prenatal suspicion of malformation of the central nervous system and neurodevelopmental disorder. Physical examination revealed dolicocephaly, dysmorphic ears, wide nasal bridge, anteverted nostrils, nasal tip bifid, long philtrum, micrognathia, malformed auricles, short, webbed neck, no visual contact and generalized hyperreflexia. The child underwent genetic karyotype reported duplication interstitial of 3q, this duplication was confirmed with comparative genomic hybridization that it reported interstitial duplication of approximately 30 Mb at chromosome 3 bands q26.1-29.

The duplicated segment overlaps with Dup(3q) syndrome, 3q26 microduplication syndrome and recently described 3q29 microduplication syndrome (MIM 611936). Clinically, the patient's phenotype is congruent with some anomalies described in all syndromes that compromised of 3q. We suggest that 3q26.1-q29 microduplication identified is a novel genomic spectrum of Dup(3q) syndrome.

## PgmNr 2471: Copy number variation (CNV) identification, interpretation and database from a Brazilian cohort.

### Authors:

M.I. Melaragno <sup>1</sup>; V. Cabral <sup>1</sup>; F.T. Bellucco <sup>1</sup>; M. Colovati <sup>1</sup>; H.R. Oliveira-Junior <sup>1</sup>; M. Moysés-Oliveira <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics Division, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; 2) Center for Genomic Medicine, Massachusetts General Hospital, Boston, USA

---

Copy number variations (CNVs) constitute an important class of variation in the human genome and the interpretation of their pathogenicity considering different frequencies across populations is still a challenge for geneticists. Since the CNV databases are predominantly composed of Caucasian and/or non-mixed-race individuals, and Brazilian genetic constitution is admixed and ethnically diverse, diagnostic screenings on Brazilian variants are greatly confounded by the lack of populational references. We analyzed a clinical cohort of 268 Brazilian individuals, including patients with neurodevelopment disorders and/or congenital malformations. We used Affymetrix's arrays, considering a minimum size of 150 kb for losses and 200 kb for gains. The pathogenicity of CNVs was classified according to their gene content and overlap with known benign and pathogenic variants. A total of 1,504 autosomal CNVs (1,207 gains and 297 losses) were classified as benign (92.9%), likely benign (1.6%), VUS (2.6%), likely pathogenic (0.2%) and pathogenic (2.7%). The mean size of the CNVs was ~763 kb, being ~704 kb for gains and ~1.0 Mb for losses. The mean sizes of benign, likely benign, VUS, likely pathogenic and pathogenic CNVs were 586 kb, 522 kb, 621 kb, 1.0 Mb and 7.2 Mb, respectively. Some of the CNVs were recurrent and with frequency increased in our cohort, when compared to open resources of structural variants: 14q32.33 gain (96.6% patients), 22q11.22 gain (32.1%), 1q21.1 gain (4.5%), and 1p36.32 gain (7.5%). Then, CNV's pathogenicity interpretation was reevaluated according to their frequency. Highly recurrent CNVs classified as likely benign or VUS were considered non-pathogenic in our Brazilian sample. This study shows the relevance of introducing CNV data from diverse cohorts to improve on the interpretation of clinical impact of genomic variations. Financial support: FAPESP, Brazil

## PgmNr 2472: *EP300* related Rubinstein-Taybi syndrome: Expanding the phenotype and genotype spectrum.

### Authors:

R. Kosaki <sup>1</sup>; H. Okuno <sup>2</sup>; K. Kosaki <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Div Med Genet, Natl Ctr Child Hlth & Dev, Tokyo, Japan; 2) Department of Physiology, Keio University School of Medicine; 3) Center for Medical Genetics, Keio University School of Medicine

**Background?** Rubinstein-Taybi syndrome (RSTS) is a multiple malformation syndrome characterized by typical facial appearance, broad thumbs and toes, and developmental delay. RSTS is an extremely rare autosomal dominant genetic disease. An estimated prevalence of one case per 125,000 live births.

The diagnosis of RSTS is primarily based on clinical features with distinctive facial features and hand and foot abnormalities. RSTS is caused by heterozygous mutations in *CREBBP* and in *EP300* genes in fifty to sixty percent and eight to ten percent. *EP300* is responsible for a minority case of RSTS, the phenotype has been shown various spectrum. Some do not have characteristic facies nor hand and foot abnormalities. Intellectual disability also appears to be affected mildly.

**Purpose?**

The purpose is to describe three unrelated RSTS patients without classic manifestations who have *EP300* mutations.

**Method?**

We examined three intellectual disability /developmental delay with congenital multiple malformations.

Genomic DNA was extracted from a peripheral blood sample. Medical exome sequencing was performed using the TruSight One Sequencing Panel (Illumina San Diego, CA) running on a MiSeq platform (Illumina, San Diego, CA). The bioinformatics analysis was performed. The mutation was confirmed using Sanger sequencing.

**Result?**

Case1: A 5-months old boy with developmental delay, large VSD and broad thumbs. We identified with *EP300* nonsense mutation.

Case2: A 2-years old girl with intellectual disability, downslanting palpebral fissures, and slightly big toes. We identified with *EP300* missense mutation.

Case3: A 11-months old girl with DORV, TR, myelomeningocele, hydrocephalus, imperforate anus, coloboma and hearing impairment. We identified de novo *EP300* frameshift mutation and de novo *SOS1* mutation.

Case4: A 6-years old girl with PDA, PS, coloboma and hearing impairment. We identified de novo *EP300* frameshift mutation.

**Discussion?**

We showed the clinical variability of RSTS with *EP300* mutations which provided the expansion of RSTS spectrum. It will contribute the genotype-phenotype correlation in RSTS. We also showed the clinical utility of medical exome analysis as a diagnostic test in patients with congenital multiple malformations.

Informed consent from the parents and approval from the local institutional review board were obtained prior to the molecular studies.

There are no conflicts of interest to declare.

## PgmNr 2473: A 3.99 MB interstitial deletions of 8q12.1q12.2 in a girl with CHARGE syndrome.

### Authors:

H. Numabe <sup>1,2</sup>; K. Kiribayashi <sup>1</sup>; K. Shirai <sup>3</sup>; M. Nishibata <sup>2</sup>; S. Nara <sup>2</sup>; Y. Morishima <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Genetics Center, Tokyo Medical University Hospital, Shinjuku, Tokyo, Japan; 2) Department of Pediatrics and Adolescent Medicine, Tokyo Medical University, Shinjuku, Tokyo, Japan; 3) Department of Otorhinolaryngology, Head and Neck Surgery, Tokyo Medical University, Shinjuku, Tokyo, Japan

---

CHARGE syndrome is a genetic syndrome characterized by a specific and recognizable pattern of features. Mutations or deletions of the gene encoding chromodomain helicase DNA binding protein 7 (CHD7) cause this syndrome.

A girl was born after 40 weeks and 4 days of gestation. Apgar scores at 1 and 5 min were 8 and 8, respectively. Her birth weight was 2386 g (-2.19 SD), height 48.0 cm (-0.97 SD), and head circumference 31.0 cm (-1.96 SD). She was transferred to our hospital because of right preaxial polydactyly, bilateral ear deformations, and low oxygen saturation (70% in room air).

Echocardiography revealed patent ductus arteriosus (PDA), and patent foramen ovale (PFO). Hearing loss was also noted.

The initial G-banding karyotyping showed 46,XX. Whole-genome SNP microarray analysis revealed a 3.99 MB interstitial deletion of chromosome 8q. arr[hg19] 8q12.1q12.2(58,009,323-61,994,781)x1. The Online Mendelian Inheritance in Man database includes two morbid genes in this interval. *CA8* gene is an autosomal recessive causative gene of cerebellar ataxia and mental retardation with or without quadrupedal locomotion 3. *CHD7* is an autosomal dominant causative gene of CHARGE syndrome.

CHARGE syndrome caused by the microdeletion of 8q is rare (5 cases in ClinVar). In 2013, Palumbo O. et al. reported the 8q12.1q12.3 de novo microdeletion involving the *CHD7* gene in a patient without the major features of CHARGE syndrome.

## PgmNr 2474: Clinical and genetic variant spectrum of 23 Chinese Fanconi anemia patients.

### Authors:

J. Zhang <sup>1</sup>; D.J. Nie <sup>1</sup>; L.L. Liu <sup>1</sup>; W. Zhang <sup>1</sup>; X.X. Li <sup>1</sup>; F. Wang <sup>1</sup>; X. Chen <sup>1</sup>; Y. Zhang <sup>1</sup>; P.X. Cao <sup>1</sup>; H.X. Liu <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Pathology & Laboratory Medicine, Hebei Yanda Lu Daopei Hospital, Langfang, 065201, China; 2) Division of Pathology & Laboratory Medicine, Beijing Lu Daopei Hospital, Beijing, 100176, China; 3) Beijing Lu Daopei Hospital Institute of Hematology, Beijing, 100176, China

---

Fanconi anemia (FA) is the most common inherited bone marrow failure syndrome with 22 causal genes identified to date. Main clinical features of FA include congenital malformations, bone marrow failure onset at pre-school age, and predisposition to hematological and solid malignancies. *FANCA* gene variants contribute to over 60% of all cases followed by variants in *FANCC* and *FANCG*, which both account for 10% of cases. We identified 24 Chinese FA patients via clinical manifestations combined with Chromosome breakage tests and whole genomic sequencing (WGS) and described their phenotype and genetic aberrations here. There were 17 males and six females. The media age of the onset of bone marrow failure was 5 years old. two patients had family history of anemia. 20/23 (87%) patients manifested malformations, including skin pigmentation (10/23), polydactyly (8/23), growth retardation (6/23), café au lait spots (6/23), flat nose bridge (5/23), hypertelorism (5/23), microcephalus (5/23), adduction deformity thumbs (4/23), blepharoptosis (3/23), patent ductus arteriosus (3/23), absent of thumbs (2/23), renal ectopia (1/23), unilateral ovary absent (1/23), delayed puberty (1/23), thenar dysplasia (1/23), indirect inguinal hernia (1/23), and hydronephrosis (1/23). 19 patients had significantly increased chromosome breakage rates induced by mitomycin C. 40 variants were identified including 28 in *FANCA* gene, 4 in *FANCD2*, 3 in *FANCB*, 2 in *FANCE*, 2 in *SLX4*, and 1 in *FANCC*. Two homozygous mutations (*FANCC* c.545C>A and *FANCA* c.1867C>T) were discovered in two patients came from consanguineous families. Variants composed by 13 missense mutations, 10 large deletions, 7 nonsense mutations, 7 frameshift mutations, 2 splicing mutations, and one deep intron mutation. Most variants were novo and private, and no overlaps harbored by different patients except one CNV (*FANVA* chr16:89780001-89822000del) shared by two. The only efficient therapeutic strategy of FA is hematopoietic stem cell transplantations (HSCTs) up to now, which were performed on 19 patients in this cohort. In the patients who accepted HSCTs, 14 patients have achieved complete remission, while five patients died from thrombotic microangiopathy or sever infections. Although this cohort is relatively small, data obtained are still informative. Nationwide multicenter study on this disease entity is expected for its rarity.



## PgmNr 2475: Skewed X-chromosome inactivation in symptomatic female carriers of hemophilia A and B.

### Authors:

G.C. Ma <sup>1</sup>; S.P. Chang <sup>1</sup>; M.C. Shen <sup>2,3</sup>; M. Chen <sup>1,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genomic Medicine, Changhua Christian Hospital, Changhua, Taiwan; 2) Division of Hematology-Oncology, Department of Internal Medicine, Changhua Christian Hospital, Changhua 50006, Taiwan; 3) Hemophilia Treatment and Thrombosis Center, Changhua Christian Hospital, Changhua 50006, Taiwan; 4) Department of Obstetrics and Gynecology, Changhua Christian Hospital, Changhua 50006, Taiwan; 5) Department of Obstetrics and Gynecology, College of Medicine, National Taiwan University, Taipei 10041, Taiwan; 6) Department of Medical Genetics, National Taiwan University Hospital, Taipei 10041, Taiwan

---

Hemophilia A (HA) and B (HB) are X-linked recessive bleeding disorders caused by defects in the factor VIII and factor IX. Patients of HA and HB are typically hemizygous males and females homozygous or compound heterozygous for mutant *F8* or *F9*. Symptomatic female carriers are relatively rare. Here, we report three families afflicted with female hemophilia, one with severe HA, another with moderate HA and the other with severe HB. Cytogenetics, molecular analyses on *F8* and *F9*, and X-skewed inactivation assay were performed to determine the genetic defects of diseases. All the three female patients are demonstrated to carry a heterozygous mutation for the *F8* or *F9* (one mutation is inherited and the remaining two mutations are sporadic). In one HA and one HB patients, highly skewed inactivation of the wild-type X-chromosome (100%) was noted, which are thought to be responsible for the HA and HB. Our result highlight the importance of X-chromosome inactivation analysis to explore the underlying cause of affected females with X-linked recessive diseases (e.g. HA and HB) when only one mutant allele was determined.

## PgmNr 2476: A newborn with novel mutation in the hematopoietic-specific RAC2 GTPase identified by targeted exome analysis.

### Authors:

M. Sun <sup>1,4</sup>; J. Han <sup>1</sup>; D. McKeone <sup>2</sup>; W. Bu <sup>3</sup>; R. Marsh <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cincinnati Children's Hospital, Cincinnati, OH.; 2) Penn State Hershey Children's Hospital, Hershey, PA.; 3) University of Pennsylvania, Philadelphia, PA.; 4) University of Cincinnati, Cincinnati, OH.

---

The Rac subfamily belongs to the Rho family of GTPases. Unlike the widely-expressed RAC1 and RAC3, RAC2 is restricted to hematopoietic cell lineages. Germline alterations in RAC2 have recently been linked to immunodeficiency syndromes. To date, only 3 RAC2 germline mutations were reported in patients affected with primary immunodeficiency but distinct clinical features (W56\*, D57N, and E62K). Interestingly, all of them occur in the region between switch I and switch II, a region critical in defining major conformational differences between the GTP- and GDP-bound forms. Mouse models have recently revealed the essential role of Rac2 in neutrophil functions. However, these different mutations apparently cause distinct immunodeficiency, which warrants further systematic investigations.

The patient reported herein was a full-term newborn with no remarkable prenatal findings. At ~24 hours of life his CBC differential showed total WBC of 0.2K, and 0.07K at 3 days of age. He had normal IgG, but low IgM/A. Screening showed no detectable TRECs. His lymphocyte studies showed essentially absent B/NK cells, minimal T cells, and low response to mitogen stimulation. At 45 days of age, the patient underwent allogeneic bone marrow transplantation with an HLA-matched sibling. An assessment of post-transplantation chimerism showed full T-cell engraftment with normal proliferation to PHA, and normal diversity of the T-cell repertoire.

Targeted exome analysis was performed on the patient's DNA. A novel missense variant in RAC2 (p.Q61R) was identified. Parental studies confirmed this change is *de novo*. Interestingly, this variant also occurs at the highly conserved N-terminal region of switch II and closely approximates all 3 previously published germline mutations. Structural modelling predicted this mutant [R61] may disrupt the hydrogen bond between the residue D39 and Y64. Additionally, this mutant might also obstruct Mg<sup>2+</sup> cofactor to achieve high catalytic efficiency and specificity.

Although all patients with RAC2 mutations exhibited profound lymphopenia, these individuals presented with distinct phenotypic features, presumably due to varying pathogenic mechanisms. The *de novo* RAC2 mutation (p.Q61R) identified by our targeted exome analysis has not been observed previously. To fully understand the disease mechanism, we aim to further examine its biological pathways in hematopoiesis, which may help us better understand its pathogenic roles during T-cell development in humans.

## PgmNr 2477: Characterization of causal variants associated with hereditary thrombocytopenias.

### Authors:

Z. Vrzalova <sup>1,2</sup>; I. Blahakova <sup>1,2</sup>; L. Radova <sup>1</sup>; J. Trizuljak <sup>1,2</sup>; S. Pospisilova <sup>1,2</sup>; M. Doubek <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genomics, CEITEC Masaryk University, Brno, Czech Republic; 2) Department of Internal Medicine, Hematology and Oncology, University Hospital and Faculty of Medicine, Brno, Czech Republic

Inherited thrombocytopenias (IT) are a heterogeneous group of 33 different forms of monogenic disorders caused by molecular defects affecting at least 40 genes. Pathogenic variants in these genes usually lead to the disruption of megakaryopoietic and thrombopoietic processes and present as the thrombocytopenia phenotype (low platelet count, blood-examination). However, patients are occasionally misdiagnosed with the immune thrombocytopenia and unsuccessfully treated with steroid therapy and splenectomy. In some patients, accurate diagnosis of IT can only be established based on the results of molecular genetic testing.

In our patient cohort, we have identified novel variants in three IT families. First, a rare variant in a proband with Wiskott-Aldrich syndrome who was originally misdiagnosed as Bernard-Soulier syndrome. Second, the both family members carried a variant for *TRPM7*-related thrombocytopenia and one of them had hematological disease in their medical history. We detected in the third family rare variants of autosomal recessive Glanzmann thrombasthenia which was caused by mutations in the *ITGA2B* gene. Germline DNA analysis was performed on all available samples and somatic DNA analysis was done for the oncological patient. Sequencing libraries were prepared according to the SeqCap EZ Human Exome Probes v3 protocol and sequencing was performed on NextSeq500. Within each family, the obtained variants were compared between the individuals with thrombocytopenia phenotype and their disease-free relatives.

Using whole exome sequencing (WES), we characterized a unique variant segregating with thrombocytopenia phenotype for each of the three families: *WAS*: NM\_000377: exon 10:c.998G>C:p.Gly333Ala, *TRPM7*: NM\_017672: exon4:c.223A>G:p.I75V, and *ITGA2B*: NM\_000419: exon29:c.2965G>A:p.A989T; exon29:c.2944G>A:p.V982M. *In silico* analysis revealed a structural defect of amino acid changes found in the protein structures, indicating that all of the variants are pathogenic. We identified a causal variant for each of the families analysed. These results helped the clinicians determine the correct diagnosis to patients. Besides that, accurate diagnosis of IT allow clinicians to conduct further examinations if the identified variant poses additional risk to the carrier, i.e., a higher risk of oncological disorders.

The research was done according to the Declaration of Helsinki.

Supported by Ministry of Health of the Czech Republic, grant No 16-29447A.

## PgmNr 2478: New case of Fanconi anemia complementation group T caused by a pathogenic missense variant in *UBE2T*.

### Authors:

A. Ferrer<sup>1</sup>; L. Schultz-Rogers<sup>1</sup>; A. Mangaonkar<sup>2</sup>; N. Jacobi<sup>3</sup>; M. Litzow<sup>2</sup>; E. Klee<sup>1,4</sup>; M. Patnaik<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota.; 2) Department of Hematology, Mayo Clinic, Rochester, Minnesota; 3) Department of Clinical Genomics, Mayo Clinic, Rochester, Minnesota; 4) Hennepin Healthcare, Minneapolis, Minnesota

---

Fanconi Anemia (FA) is characterized by genomic instability, bone marrow failure and skeletal abnormalities. Several subtypes of FA have been described and associated with variants in different genes. The FA complementation group T (FA-T) is caused by variants in *UBE2T*. Only 3 patients with (FA-T) have been reported to date, all carrying at least one loss-of-function (LoF) variant. Here we report a fourth FA-T case and the first one caused by a homozygous missense variant in *UBE2T*.

A 21-year-old Ecuadorian female presented to the Bone Marrow Failure Precision Medicine Clinic (Mayo Clinic, MN) for assessment. She reported periodic fevers, persistent macrocytosis and intermittent cytopenias since 8 years of age. Her last complete blood count (October 2018) indicated anemia, thrombocytopenia and increased MCV. Her bone marrow was moderately hypocellular (40% to 50%) without evidence of dysplasia or lymphoproliferation. Other phenotypes included urticaria, discoloration of hands with cold and intermittent ulcers on lips. Previous genetic testing including cytogenetics and a periodic fever gene panel were negative, as well as testings for autoimmunity and infectious diseases. No related family history or consanguinity was described.

The patient underwent a custom-designed XomeDSlice panel through GeneDX that uncovered a homozygous variant in *UBE2T* (c.196C>T; p.P66T) not reported in gnomAD or associated with disease (ClinVar and HGMD). The encoded proline is highly conserved across species and located in the UBC fold domain, critical for protein function. *In silico* tools (SIFT, MutationTaster, PolyPhen2, MCAP and PredictSNP2) agreed on a deleterious effect. To confirm this, a chromosomal breakage assay was performed and resulted positive (58% and 36% breakage positive cells with MMC and DEB, respectively) in line with previous FA-T reported. Complementation testing is currently underway. With these results we classified the variant as likely pathogenic by ACMG criteria.

This is the first FA-T patient reported with a homozygous missense likely pathogenic variant in *UBE2T*. This variant may cause a milder effect in the *UBE2T* function compared to the other three reported patients harboring one loss of function variant, possibly explaining the lack of skeletal finding in the proband. However, more research is needed to confirm this hypothesis. This case adds to our knowledge of FA-T causal variants and highlights the relevance of including *UBE2T* in FA evaluations.

## PgmNr 2479: Survivors of childhood cancer are at high risk for clonal hematopoiesis.

### Authors:

Y. Han <sup>1</sup>; J.D. Fish <sup>2</sup>; R. Torene <sup>1</sup>; J. Scuffins <sup>1</sup>; L. Havens <sup>1</sup>; K. Retterer <sup>1</sup>; B. Solomon <sup>1</sup>; K. Onel <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) GeneDx, Gaithersburg, MD; 2) Northwell Health, New York City, NY; 3) Icahn School of Medicine at Mount Sinai, New York City, NY

With aging, hematopoietic stem cells acquire somatic mutations and undergo clonal expansion, a process called clonal hematopoiesis (CH). CH is associated with an increased risk of hematologic cancers, coronary artery disease, and all-cause mortality. As survivors of childhood cancer (SCC) experience these same complications as late-effects of cancer treatment at younger ages than the general population, we hypothesized that SCC have a higher prevalence of CH relative to sex and age matched peers. As such, CH may represent a genomic biomarker for the premature aging phenotype seen in survivors.

82 consecutive SCC (46 females and 36 males) who were  $\geq 18$  years old,  $\geq 5$  years from cancer treatment completion, and who had not undergone a bone marrow transplantation were recruited and consented from a structured long-term follow-up program. Age and sex matched controls were pulled in a 3:1 ratio from healthy parents of probands referred for clinical exome sequencing who had no personal or family history of cancer (N=246). Whole exome sequencing was performed using DNA extracted from whole blood (100X median coverage), with somatic variant detection for 309 previously published CH SNVs in 56 genes. A variant allele fraction cutoff of  $>5\%$  was used to call CH. Stratified analysis was conducted to test clinical factors for association with CH.

Survivors had a mean age of 26.45 years (range: 18.1-50), and 13.4 years off therapy (range: 6.6-34.8) at the time of sample collection. All survivors were treated with chemotherapy, and 48.8% were exposed to therapeutic radiation. 18.3% of survivors had at least 1 CH variant, compared with 6.5% of controls ( $p=0.0036$ , two-tailed Fisher's exact test). Among SCC, individuals with CH had multiple CH variants whereas the controls with CH tended to have 1 variant (mean CH count 0.256 for SCC and 0.0732 for controls;  $p=0.0014$ , Wilcoxon rank sum). Among genes analyzed, SCC had the most CH variants in DNMT3A (7 variants from 7 individuals), while controls had the most CH variants in RAD21 (7 variants from 7 individuals). We did not observe an association between CH and primary cancer type, treatment modality, time from therapy, or outcome, although this study was not powered to detect such associations.

In summary, SCC are at greater risk for CH, and have more CH variants than age and gender-matched controls. Future studies are needed to examine the association of CH with exposures, risk of chronic disease and mortality.

## PgmNr 2480: Lymphocyte abnormalities in patients with 22q11.2 deletion syndrome.

### Authors:

D. Soares<sup>1</sup>; M. Matta<sup>2</sup>; K. Jatahy<sup>2</sup>; A. Dantas<sup>3</sup>; N. Nunes<sup>3</sup>; M. Montenegro<sup>1</sup>; L. Kulikowski<sup>1</sup>; M. Melaragno<sup>3</sup>; M. Carneiro-Sampaio<sup>1</sup>; L. Torres<sup>2</sup>; C. Kim<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Faculdade de Medicina da Universidade de São Paulo, Brazil; 2) Translational Research Laboratory, Instituto de Medicina Integral Prof. Fernando Figueira, Brazil; 3) Department of Genetics, Universidade Federal de São Paulo, Brazil

---

**Background and aims:** 22q11.2 deletion syndrome (22q11.2DS) is the most common human microdeletion. It is known to have a heterogeneous presentation that includes multiple congenital anomalies and immunodeficiency one of the most striking features. Usually it is characterized by T cell lymphopenia, B cell dysfunction and autoimmunity.

**Methods:** Peripheral blood mononuclear cells were isolated from heparinized venous blood from 15 healthy individuals and 15 22q11.2DS patients. Patient history and medications were also collected. We performed immunophenotyping of T cells in patients and controls by flow-cytometry, through incubating peripheral blood mononuclear cells with a specific panel of antibodies.

**Results:** Patients presented low levels of TCD3<sup>+</sup> lymphocytes and elevated relative values of double-negative T cells (TCD3<sup>+</sup>/CD4<sup>-</sup>/CD8<sup>-</sup>) when compared to healthy controls (p=0.01 e p=0.03, respectively). The TCD4<sup>+</sup> and TCD8<sup>+</sup> lymphocyte values were slightly lower in the patients when compared to the healthy controls (without statistical significance).

**Conclusions:** 22q11.2DS patients present not only T-cell lymphopenia but also high levels of double-negative T cells that may be one of the signs of an immune dysregulation observed in these patients.

**Support:** FAPESP and CNPq.

## PgmNr 2481: Clinical significance of reinterpreting previously reported immunologic disease genomic tests.

### Authors:

J.A. SoRelle <sup>1</sup>; C.A. Wysocki <sup>2</sup>; J.Y. Park <sup>1,4</sup>; G. Gotway <sup>3,4</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Pathology, University of Texas Southwestern Medical Center, Dallas, Texas.; 2) Internal Medicine, Division of Allergy and Immunology, University of Texas Southwestern Medical Center, Dallas, Texas.; 3) Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas.; 4) Eugene McDermott Center for Human Growth & Development, University of Texas Southwestern Medical Center, Dallas.

---

**Introduction:** Next-generation sequencing (NGS) testing has become a common diagnostic tool for evaluating pediatric patients with immune disorders. NGS testing can comprehensively identify gene variants of interest; however, determining the clinical significance of novel variants is challenging. Advances in publicly available databases and standardized interpretation criteria have improved the clinical interpretation of variants. However, the utility of re-interpreting previously reported NGS immune tests has not been systematically evaluated. In this study, we identify the frequency and significance of variant reclassification from previously reported NGS immune gene panels.

**Methods:** All NGS immune gene panel reports sent out from a tertiary-care pediatric hospital were retrospectively reviewed (July 2012 to July 2018). Reports were from multiple reference laboratories, and included genes for severe combined immunodeficiency, primary ciliary dyskinesia, and autoimmune and lymphoproliferative syndrome. Previously reported variants were first screened using population (gnomAD, 1000 Genomes phase 3) and clinical (ClinVar) databases. Any variant marked for having a high population frequency (>1%) or a conflicting report in ClinVar was 1) flagged then 2) reinterpreted using ACMG variant classification criteria.

**Results:** 120 patients were previously tested yielding 110 variants in 62 patients. Of the 110 variants, 36% were pathogenic or likely-pathogenic (P-LP n=40), and 64% were variants of uncertain significance (VUS n=71). 15% (n=6) of the P-LP variants were reclassified; variants were downgraded to VUS (n=5, CYBB, DOCK8, FOXP1, MBD5), or B-LB (n=1, CFTR). Of the 71 VUS, 27% (n=19) were reclassified downward in significance (B-LB), and none were upgraded to clinically significant (P-LP). Overall, 23 of 120 patients (19%) were impacted by variant reclassification.

**Conclusions:** Based on our single institution experience, a large proportion of variant classifications were revised (21%; 25/120). These reclassifications impacted 19% of patients (23/120). A clinically significant reclassification (P-LP to VUS or benign) occurred in 15% of patients with a diagnostic variant. In addition, many reclassified VUS were downgraded. Overall, these findings indicate that all P-LP and VUS variants from previously reported NGS tests should be routinely reviewed and reanalyzed.



## PgmNr 2482: Companion genetic aberrations in CEBPA double mutated AML and their clinical significance.

### Authors:

Y. Zhang; F. Wang; X. Chen; Y. Zhang; M.Y. Wang; H. L; X.L. W; M. L; H.X. L

[View Session](#) [Add to Schedule](#)

**Affiliation:** Hebei Yanda Lu Daopei Hospital, Langfang, China

---

Background: Acute myeloid leukemia (AML) with double mutant CEBPA (CEBPAdm) is a new entity in the 2016 WHO classification with favorable prognostic implications, but still have some patients relapsed. This retrospective analysis aimed to thoroughly analyzed the molecular abnormalities of AML patients with CEBPAdm, explored the pathogenesis and prognostic factors, and provided guidance for clinical diagnosis and treatment. Methods: Amplicon-targeted, next-generation sequencing was performed on 609 patients who were diagnosed of de novo AML. Results: CEBPA mutation were detected in 15.9% of patients (97/609), of which 76 cases were CEBPAdm and 21 cases were CEBPA single mutation (CEBPAsm). A total of 88.2% of CEBPAdm patient concomitant with other gene mutations and 39 kinds of mutation combinations were detected. The frequency of GATA2 mutation in CEBPAdm group was significantly higher than that in CEBPAsm group (31.6% vs 0%,  $P=0.003$ ), and the GATA2, CSF3R, JAK3 and KIT mutations were exclusively occurred with CEBPAdm. The proportion of epigenetic modifiers mutations in CEBPAsm group was significantly higher than that in CEBPAdm group (47.6% vs 15.8%,  $P=0.005$ ). In patients with normal karyotype, CEBPAdm patients with activated signaling mutations had poorer OS and DFS than patients without activated signaling mutations ( $P=0.035$ ;  $P=0.023$ ), with FLT3-ITD mutations showed significantly poor OS than patients with FLT3-ITD wild-type ( $P=0.012$ ), and with KIT mutations showed significantly poor DFS than patients with KIT wild-type ( $P = 0.015$ ). The incidence of germline mutations in this cohort was 5.3% (4/76), including one C-terminal mutation. Conclusions: A comprehensive analysis of the mutation profile of CEBPAdm patients will help us to better understand the pathogenesis and identify prognostic factors. Therefore, it is helpful for early identification of patients with poor prognosis, thus better guiding treatment and improving prognosis.



## PgmNr 2483: Genotyping in regions of high homology.

### Authors:

C.S. Nelson; D. Black; K. Aull; C. Bruckner; R. Varma; J. Gollub; M. Shapero; Blood transfusion Genomics Consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** Thermo Fisher Scientific (Affymetrix), 3450 Central Expressway Santa Clara, California, 95051

---

Many SNP and indel variants of particular interest occur in genes with very high homology to pseudogenes, paralogs, and other gene family members. *CYP2D6*, a gene critical to the metabolism of 25% of drugs, has two highly homologous pseudogenes, *CYP2D7* and *CYP2D8P*. Genotyping many critical variants in *CYP2D6* is complicated by variation at or near the equivalent site in the pseudogenes, or by variable background signal arising from the frequent deletion of the pseudogenes. Similarly, variants in the *RHCE* gene determine critical antigens of the Rh blood group system, the most important blood group system after ABO for transfusion medicine, can be complicated by SNP or copy number variation in the paralogous *RHD* gene, the other gene determining Rh blood type.

Here we demonstrate accurate genotyping of *CYP2D6* and *RHCE* polymorphisms on the Applied Biosystems™ Axiom™ PMD Array, a microarray that includes pharmacogenomic and blood-typing research variants as well as genome-wide coverage for global populations. We demonstrate accuracy for 96 samples from the 1000 Genomes Project with two very different approaches. For *CYP2D6* and other drug metabolism genes, we employ a target enrichment approach in the assay, combining mPCR of the target variants with whole genome amplification for other genotyping. For *RHCE*, we use an algorithmic approach, producing accurate genotype calls with no enrichment step required. In particular we show accurate genotyping of the *RHCE* exon 2 T307C variant determining Rh C/c alleles, historically an especially challenging target due to the 100% homology of the C allele to *RHD* exon2.

The ability to genotype these variants, both specifically and in the context of genome-wide tests, supports both targeted and broad research in precision medicine. In contrast to assays for a relatively small number of SNPs, target enrichment in combination with whole genome amplification in the same microarray assay enables efficient data collection on *CYP2D6* and other variants of research interest in pharmacogenomics as well as genome wide markers for discovery of novel associations. The success of an algorithmic approach for the *RHCE* gene suggests that the simpler assay, without target enrichment, might suffice for other difficult target variants as well.

## PgmNr 2484: Functional analysis of a novel splice site variant in *GLI3* gene causing Greig cephalopolysyndactyly syndrome.

### Authors:

E. Siavriene<sup>1</sup>; V. Mikstiene<sup>1</sup>; D. Radzevicius<sup>2</sup>; Z. Maldziene<sup>1</sup>; T. Rancelis<sup>1</sup>; G. Petraityte<sup>1</sup>; I. Kavaliauskiene<sup>1</sup>; L. Sarkinas<sup>2</sup>; A. Utkus<sup>1</sup>; V. Kucinskas<sup>1</sup>; E. Preiksaitiene<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human and Medical Genetics, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University, Vilnius, Lithuania; 2) The Children's Hospital, affiliate of Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania

---

**BACKGROUND:** Preaxial polydactyly type IV (MIM 174700; ORPHA 93338), also referred as polysyndactyly, has been described in a few syndromes including Greig cephalopolysyndactyly syndrome (GCPS; MIM 175700; ORPHA 380). GCPS is caused by heterozygous mutation in the *GLI3* gene (MIM 165240), which is located in the cytoband 7p14. In this study, we present the functional analysis of a novel *GLI3* splice site variant, which cosegregates in three generations of a family with preaxial polydactyly type IV and other clinical features of GCPS.

**METHODS & RESULTS:** Sequencing analysis of the *GLI3* coding region identified a novel donor splice site variant NC\_000007.14(NM\_000168.6):c.473+3A>T in the proband and the same variant was subsequently identified in other affected family members. In order to elucidate the pathogenicity of the detected variant in *GLI3* gene, proband's RNA was isolated from fibroblast culture and template cDNA was synthesized. Further Sanger sequencing of the proband's cDNA sample revealed that the splice site variant disrupts the original donor splice site, thus leading to exon 4 skipping. Based on *in silico* analysis, the pathogenic splice site variant consequently results in a truncated protein NP\_000159.3:p.(His123Argfs\*57), which lacks functionally important domains: zinc finger domain, proteolytic cleavage site, transactivation domain 2, transactivation domain 1 as well as part of repressor domain.

**CONCLUSION:** The functional cDNA analysis revealed that NC\_000007.14(NM\_000168.6):c.473+3A>T led to the haploinsufficiency of *GLI3* that causes GCPS in the affected family members. This analysis provides a unique possibility to identify the molecular basis of GCPS as well as many other hereditary diseases and conditions. The work was funded by the Research Council of Lithuania (No. S-MIP-17-19/LSS-150000-1179, Ingenes project).

## PgmNr 2485: Chromosomal microarray analysis uncovers pathogenic copy number variations in neurodevelopmental disorders and congenital anomalies.

**Authors:**

S.K. Bhattacharya; L. Rawal; V. Lal

[View Session](#) [Add to Schedule](#)

**Affiliation:** Clinical Cytogenomics, National Reference Laboratory, Dr. Lal PathLabs Ltd., Block-E, Rohini Sector-18, New Delhi, Delhi, India

---

Neurodevelopmental disorders (NDDs) represent broad spectrum of cognitive, neurological, and/or psychiatric dysfunction caused by impairment of the brain during development. We used chromosomal microarray (CMA) to study NDDs in patients with unexplained developmental delay/intellectual disability (DD/ID) accompanying dysmorphism and/or congenital anomalies. Of all the 102 patients identified with DD/ID, 48 patients had a normal profile (46XX/XY), 53 showed pathogenic CNVs along with an exceptional case (case 199) encompassing high levels of homozygosity (approx. 17.5%). The size of the CNVs in affected patients ranged from 36 kb to 15.5 MB. The most common variant in cases with ASD and developmental delay was duplication 22q11.2 involving 400Kb region, which was validated using karyotyping and FISH. Five of 53 sporadic patients had known microdeletion syndromes. It is envisaged that the application of microarray will expand the spectrum of cytogenomic abnormalities by including complex and cryptic structural variants. Further, delineation of molecular mechanisms of these cytogenomic abnormalities coupled with development of novel therapeutic approaches will ultimately lead to disease-specific personalized management and precision treatment.

## PgmNr 2486: Revealing a familial atypical deletion in 5p using checklist directing and correct indication of molecular tests.

### Authors:

S.N. Chehimi <sup>1</sup>; E.A. Zanardo <sup>1</sup>; A.T. Dias <sup>1</sup>; M.M. Montenegro <sup>1</sup>; A.M. Nascimento <sup>1</sup>; Y. Gasparini <sup>1</sup>; A.B. Freitas <sup>1,2</sup>; G.M. Spolador <sup>3</sup>; C.A. Kim <sup>3</sup>; L.D. Kulikowski <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina da Universidade de São Paulo, São Paulo, SP, BR.; 2) Departamento de Obstetrícia e Ginecologia, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, BR.; 3) Unidade de Genética, Departamento de Pediatria, Instituto da Criança, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, BR.

---

Deletions in the short arm of chromosome 5 (5p) are usually diagnosed when the patients present cat-like cry and typical phenotype involving microcephaly and intellectual disability. The most used molecular test to detect this syndrome is karyotype but it is only useful for deletions greater than 10 Mb. In the cases of patients with smaller deletions, the diagnosis may be delayed due to the availability of the molecular testing and correct indication. With the assistance of specialized physicians and our research group, we developed a checklist for patients with 5p deletion suspicion and confirmation using cytogenomic tests (karyotype, MLPA and array). The checklist evaluated familiar background, physical and intellectual development of the patient. Two siblings, a 6-years-old boy and 8-years-old girl, were referred to genetic investigation due to the moderate to severe intellectual disability and speech delay. During the appointment, we discovered both presented the high-pitched cry at birth although did not present any of the typical physical features of 5p-syndrome. Karyotype did not show any alteration. P064 MLPA kit, with five probes for 5p region, showed that the patients and the mother presented an atypical deletion with four probes deleted (*TERT\_ex2*; *TERT\_ex13*; *CLPTM1L*; *IRX4*) but one was still present (*CTNND2*). Using array, we were able to determine that the siblings presented a 6,2 Mb deletion, inherited from their mother who presented similar features. *CTNND2*, the gene that was not deleted, is associated to cerebral development and neurons migration. The presence of this gene may explain why the patients were still able to communicate, learn and understand other people, some of the features that patients with larger 5p deletions are not able to do. Our results indicate that a complete evaluation and different investigations tools, as checklists, can lead to the correct cytogenomic investigation without delays to clarify the genetic cause for intellectual disability. Also, be engaged with research projects may be the only opportunity for patients to have access to molecular tests that are expensive and not available in the Brazilian public healthcare.

## **PgmNr 2487: An international collaboration between a high- and low-resource country leads to the identification of a novel chromosome 18 deletion, chromosome 5 duplication with translocation causing facial dysmorphism and global developmental delay.**

### **Authors:**

S. Upadhyaya<sup>1</sup>; G. Verdi<sup>1</sup>; J.L. Smith<sup>2</sup>; S.H. Elsea<sup>2</sup>; B. Nelson<sup>4,5</sup>; K. Yearwood<sup>5</sup>; H. Hakonarson<sup>3</sup>; D. Li<sup>3</sup>; E.J. Bhoj<sup>3</sup>; A.K. Sobering<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Biochemistry, St. George's University School of Medicine, Grenada; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 3) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 4) Grenada General Hospital, Pediatrics Ward, Grenada; 5) Clinical Teaching Unit, St. George's University School of Medicine, Grenada

The experience for families of children born with developmental delay in low- and middle-income nations is in stark contrast to high-income nations. In resource-limited countries, barriers to obtaining genetic diagnosis might involve economic disparities, geographical isolation, and inadequate access to genetic medicine. As part of a combined teaching, research, and community outreach effort, we provided genetic testing for a child with global developmental delay (GDD), dysmorphism, and intellectual disability (ID). Deletions involving the terminal region of 18q may lead to a variable phenotype including short stature, developmental delay (DD), hypotonia, hearing impairment and foot deformities (OMIM 601808). Duplications of the 5p terminal region cause DD, ID, and variable facial dysmorphism (OMIM 613174).

We present a 5-year-old Afro-Caribbean boy who has GDD, ID, hypotonia, feeding difficulties, facial dysmorphism and bilateral club feet. Speech development is limited to less than 5 words. He can walk for a short distance with an unstable gait. He has a high frontal hairline with a prominent widow's peak, frontal bossing, a flat and broad nasal bridge, up-slanting palpebral fissures, hypertelorism, and camptodactyly. His feet are internally rotated, and the right foot is worse than the left. Both of his parents are healthy. The family described a now deceased maternal aunt who they thought had Down syndrome, but to our knowledge, she was never medically evaluated. He has two maternal aunts who have primary infertility. Conventional karyotyping showed additional chromatin of unknown origin on the long arm of chromosome 18. Cytogenomic microarray analysis showed a 6.4 Mb deletion (chr18:71518518-77943115) on chromosome 18 and a 24 Mb duplication (chr5:8097721-32062984) of material from chromosome 5. We postulate that the combined clinical features of our patient are due to deletions of regions of chromosome 18 and the partial duplication of chromosome 5. We are now in the process of offering a karyotype to both parents. A definitive diagnosis is beneficial to these families and can assist in providing answers with reassurance, and allows for informed genetic counseling and recurrence risk assessment, especially if a balanced translocation is found in a parent as suspected.

## **PgmNr 2488: *FMR1* methylation and sizing PCR assays improve sensitivity and efficiency in clinical testing for fragile X syndrome, revealing some uncommon and unexpected results.**

### **Authors:**

J.A. Lee; R.C. Caylor; W. Hall; J.A.C. Coleman; R.J. Louie; M.J. Friez

[View Session](#)[Add to Schedule](#)

**Affiliation:** Molecular Diagnostic Laboratory, Greenwood Genetic Center, Greenwood, South Carolina.

---

Fragile X syndrome is the most common cause of autism worldwide, affecting 1 in 4000 males and 1 in 6000 females, and is caused by a CGG triplet repeat expansion in the *FMR1* gene. *FMR1*-related disorders also include fragile X-associated tremor ataxia syndrome and *FMR1*-related primary ovarian insufficiency. The disease mechanism involves aberrant methylation and expression of *FMR1*. A repeat size of 200 CGG repeats or greater is considered a full mutation, causing hyper-methylation of the *FMR1* promoter and transcriptional silencing. Premutations are between 55 and 200 CGG repeats and are generally not associated with abnormal methylation; they typically result in an unmethylated allele in males and partially methylated alleles in females due to X-inactivation, and are associated with increased expression. In our molecular diagnostic laboratory, we have implemented a more sensitive methylation PCR since July 2017, as well as a sizing PCR since November 2015, replacing the traditional Southern blotting and homebrew PCRs, respectively. By implementing these assays, we have been able to improve the efficiency of our workflow, as well as detect size and methylation mosaicism with a greater accuracy than was possible by traditional methods. We have also detected some uncommon and unexpected results, including males and females with size and methylation mosaicism as well as with fully-methylated premutations. Here we describe our experience to date, share some unusual findings, and reaffirm that methylation analysis be performed for any patient with a premutation-sized allele or greater.

## PgmNr 2489: CNV/SNV analysis of 844 ID genes vs. virtual panel testing based on HPO terms in singleton NGS referrals.

### Authors:

I. Simonic; H. Grayton; R. Newby

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medical Genetics laboratory, Cambridge University Hospital, Cambridge, United Kingdom, United Kingdom

---

Trio analysis of WES data represents the gold standard genetic test for complex referrals. It is costly and not always possible. Our clinical exome pipeline has been modified and validated for detection of SNV, CNV and AOH and utilised diagnostically for 844 gene panel testing of singleton referrals using NxClinical software. The average time for classification of roughly 50 SNV events per case and whole genome CNV analysis was 30 minutes per case. No variants of possible clinical relevance were detected in roughly 50% of cases. All CNVs previously detected by CMA were also detected using this pipeline, providing the CNV regions contained at least one sequenced gene. Furthermore, pathogenic SNVs and CNVs, which had not been identified from previous more targeted analysis, were detected in more than 10% of cases. An additional re-analysis of the clinical exome data was completed for all cases using the variant prioritisation based on HPO terms within NxClinical environment. All SNVs and CNVs reported by the diagnostic panel testing were also detected by HPO variant prioritisation. Several additional findings were also detected and classified as pathogenic and relevant to the reason for referral. Our approach clearly demonstrates that even in the instances where trio WES is not possible, our single assay approach can efficiently detect pathogenic and likely pathogenic variants, and copy number abnormalities involving known OMIM morbid genes with no significant additional analysis time.

## PgmNr 2490: Low-pass whole genome sequencing identified a single exon deletion in RPS6KA3 causing Coffin-Lowry syndrome.

### Authors:

W. He <sup>1</sup>; T. Chiang <sup>1</sup>; J. Shen <sup>1</sup>; X. Wang <sup>2</sup>; H. Wang <sup>2</sup>; Y. Yang <sup>1</sup>; X. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) AiLife Diagnostics, Pearland, TX.; 2) Sichuan Provincial Maternal and Child Health Hospital, Chengdu, China

---

Low-pass WGS has recently been reported to be able to detect CNVs at the resolution comparable with that of chromosomal microarray analysis (CMA). In the meantime, the coverage of CMA is variable across the genome depending on the probes included in the assay, low-pass WGS has much more uniform coverage and can potentially detect CNVs encompassing any genes. While low-pass WGS has been shown to effectively detect large CNVs, its utility in detecting smaller events is relatively unknown. Here we report the identification of a single exon deletion in RPS6KA3 gene by low-pass WGS in a male patient with global developmental delay, hypertelorism, hearing impairment, and hypotonia. At the mean coverage of 3X (paired end reads), WGS detected the hemizygous deletion encompassing exon 1 of RPS6KA3. The size of the deletion is about ~42 kb (chrX:20257168-20299113, hg19). RPS6KA3 is associated with X-linked Coffin-Lowry syndrome [MIM: 303600] or mental retardation X-linked 19 [MIM: 300844]. Coffin-Lowry syndrome typically presents with dysmorphic facial features including hypertelorism, sensorineural hearing loss, skeletal abnormalities, hypotonia, and neurologic problems, which are consistent with the clinical features of this patient. Interestingly whole exome sequencing (WES) was ordered concurrently for this patient. CNV analysis based on the WES depth of coverage revealed no reads aligned to exon 1 of RPS6KA3 in this patient, consistent with the WGS findings. Our results showed that low-pass WGS, which is usually used to detect gross deletions and duplications, can also detect smaller, intra-genic CNVs.



## PgmNr 2491: Whole exome sequencing identifies two novel de novo truncating mutations in UPF3B and NLGN3 in patients with intellectual disability.

### Authors:

A. Alkhateeb<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biotechnology and Genetics Department, Jordan University of Science and Technology, Irbid, Jordan; 2) Biomedical Sciences Department, Qatar University, Doha, Qatar

---

Three families with idiopathic intellectual disability has been recruited to identify their genetic causes. The families were non-consanguineous and had a single affected patient each. All had an uneventful pregnancy and delivery. Informed consent was obtained and all procedures followed were in accordance with the ethical standards of the Institutional Review Board. Whole exome sequencing was carried out for the patients and their parents. Whole exome capture and sequencing were performed at Centogene Laboratory. Approximately 60 MB of human exons were enriched from 1 µg of fragmented genomic DNA using the SureSelect Human All Exon V6 Kit followed by processing on NextSeq platform to obtain an average coverage of ~100X. Approximately 97% of targeted bases are covered > 10X. An in-house bioinformatics pipeline was applied including base-pair calling, alignment of reads to genome assembly GRCh37/hg19, filtering out low-quality reads, and variants annotation. All disease-causing variants in CentoMD, ClinVar, and HMGD were considered. Additionally, all variants in gnomAD database with a minor allele frequency of <1% were considered. Evaluation of identified variants was focused on coding exons and their flanking intronic bases with consideration of multiple inheritance patterns. Further, clinical data and family history were considered for evaluating identified variants. Sanger sequencing was performed to confirm identified variants. Only variations in genes potentially related to the patient's clinical phenotype were reported. Two of the families showed two novel de novo nonsense mutations, one in UPF3B gene (NM\_080632.2: c.724C>T, p.Arg242\*) in first family and in NLGN3 gene (NM\_181303.1:c.2525C>A, p.Ser842\*) in second family. Both of these mutations were not found in ExAC, Kaviar, 1000G, and Gnomad databases. The third family showed a known mutation in SYNGAP1 (NM\_006772.2: c.1735C>T, p.Arg579\*). All three genes are already implicated in the etiology of intellectual disability. In summary, we identified two novel de novo nonsense mutations in two genes and one known mutation in a third gene in three families with singleton patients with intellectual disability.

## **PgmNr 2492: A novel *de novo* RAC1 frameshift likely pathogenic variant in a patient with developmental delay and epilepsy suggests haplo-insufficiency as a new mechanism of disease in RAC1-related disorders.**

### **Authors:**

J. Neidich <sup>1</sup>; K. Belhassan <sup>1</sup>; J.L. Granadillo <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Pathology and Immunology, Washington University School of Medicine, St. Louis, MO; 2) Pediatrics, Washington University School of Medicine, St. Louis, MO

---

Rac family small GTPase 1(*RAC1*) is a ubiquitously expressed gene involved through its interaction with different genes in numerous cellular functions, including motility, invasion, migration, apoptosis, proliferation, growth, cell-cycle regulation, morphology, adhesion, neuronal polarization with axonal growth, and differentiation. Although *RAC1* is an extensively studied gene, only seven patients with distinct *de novo* missense pathogenic variants and developmental disorders with diverse phenotypes have been reported.[PMID: 28886345] Functional studies were compatible with either a dominant-negative or gain-of-function mechanism.

Here we report a 16 month-old female with global developmental delay and epilepsy, normal head circumference, hypotonia, Duane syndrome, bilateral fourth toe clinodactyly, difficulty feeding, horseshoe kidney, and recurrent ear infections. On exam she had dysmorphic facies with broad forehead, deep-set eyes, hypertelorism, down-slanting palpebral features, and mild micrognathia. Brain MRI identified a mildly decreased white matter volume with thinning of the corpus callosum. Chromosomal microarray was normal, and trio whole exome sequencing revealed a novel *de novo* *RAC1* likely pathogenic frameshift, denoted c.383\_384insGT; p.I129SfsX4 (NM\_018890.3). *RAC1* deletion/duplication analysis was negative.

*RAC1* is associated with autosomal dominant intellectual disability. The seven cases described in 2017 by Reijnders et al. presented with global developmental delay/intellectual disability, and either macrocephaly or microcephaly. However, other features were variably present, such as hypotonia, dysmorphic features, cardiac abnormalities, behavioral problems, stereotypic movements, epilepsy, cerebellar abnormalities and corpus callosum hypoplasia. Therefore, the disruption of the complex signaling pathway between *RAC1* and different proteins might result in complex and diverse clinical presentations. Contrary to previous reports, our patient was found to have a frameshift variant, which is expected to result in loss of function. This suggests that haplo-insufficiency is the most likely pathogenic mechanism in our patient. To our knowledge, this is the first report of a patient with a *RAC1*-related disorder due to a heterozygous loss-of-function variant.

## PgmNr 2493: Investigation of X-fragile syndrome in Brazilian patients.

### Authors:

G.F.S. Carvalho <sup>1</sup>; A.T. Dias <sup>1</sup>; S.N. Chehimi <sup>1</sup>; Y. Gasparini <sup>1</sup>; E.A. Zanardo <sup>1</sup>; M.V. Souza <sup>1</sup>; M.M. Montenegro <sup>1</sup>; A.M. Nascimento <sup>1</sup>; A.B. Freitas <sup>2</sup>; F.A.R. Madia <sup>1</sup>; G.M. Novo-Filho <sup>1</sup>; L.D. Kulikowski <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina FMUSP, Universidade de Sao Paulo, Sao Paulo, SP, BR; 2) Departamento de Obstetrícia e Ginecologia, Faculdade de Medicina FMUSP, Universidade de Sao Paulo, Sao Paulo, SP, BR

---

Background: Fragile X Syndrome (FXS) is the second leading cause of intellectual disability (ID), affecting 1: 5000-7000 men and 1: 4000-6000 women. An alteration of the *FMR1* gene at Xq27.3 cause the FXS. Most of patients present CGG expansion (> 200 triplets) in the 5'UTR of the gene, the other patients present *FMR1* mutations and duplication / deletion in the same genomic region. Diagnostic detection using commercial kits allows the unambiguous identification of patients and carriers, while prenatal and neonatal diagnoses are a challenge. Clinical features comprise global developmental delay, anxiety and hyperactive behavior. Characteristic facies include large ears, a long face, a prominent jaw and forehead and flat feet. Material and Methods: We evaluated 63 patients with intellectual disability associated with neuropsychomotor developmental delay and / or other (FXS) clinical features, both sexes and without age restriction, using FragilEase® kits (Perkin Elmer, Massachusetts, USA) or Amplidex® (Asuragen, Texas, USA). The results show four evils with full CGG expansion (> 200 repeats), two patients with intermediate alterations, ranging from 45 to 54 repeats and one patient with pre mutation (range from 55 to 200 repeats). Four females showed alleles in heterozygosis. Discussion: Our study revealed the prevalence of 6.3% of complete mutations, and for non-complete mutations, we identified a percentage of 11.1%. The obtained results corroborated data reported in the literature, but the overall detection rate was higher than the rates previously reported, highlighted the importance of the criteria used to select patients.

## PgmNr 2494: Detection of CGG repeats in Fragile X syndrome using CRISPR Cas 9 combined with single molecule sequencing.

### Authors:

P. Pugniere <sup>1</sup>; P. Lochu <sup>1</sup>; S. Kemeny <sup>1</sup>; J. Audoux <sup>2</sup>; S. Beaumeunier <sup>2</sup>; J.M. Holder <sup>2</sup>; N. Philippe <sup>1</sup>; G. Egea <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Medical Genetics, GenBio Laboratory, Clermont-Ferrand, Auvergne-Rhône-Alpes, France; 2) SeqOne, IRMB, CHRU de Montpellier - Hopital St Eloi, Montpellier, France

---

Fragile X Syndrome is caused by CGG expansion in the 5' untranslated region (5'UTR) of the fragile X mental retardation 1 (*FMR1*) gene. These alterations range from a permutation state (55-200 repeats) to a full mutation allele (>200 repeats). It is the most common cause of X-related inherited intellectual disability. Patients with fragile X syndrome present an hypermethylation of the CGG triplets leading to an epigenetic gene silencing of *FMR1* that induces the loss of the encoded fragile X mental retardation 1 protein, FMRP.

Current diagnostic techniques are primarily based on costly and time-consuming PCR reactions that are specific to either repeat detection or promoter methylation. Using third-generation high throughput sequencing, one can fully sequence single molecules without fluorescent labelling and PCR amplification, thus significantly reducing turnaround time and reagent costs. Single molecule sequencing enables real-time sequencing of DNA which facilitates the detection of structural variants (copy number variants, gene duplications, deletions, insertions, inversions, and translocations) or base modifications (methylation). Combined with the emergence of CRISPR-Cas 9 genome editing, selection and subsequent sequencing of these genomic regions of interest is now possible.

We present a new methodology that uses targeted read, single molecule sequencing to simultaneously identify and count CGG repeats, detect interrupting AGG sequences and provide insight into the methylation status of the *FMR1* gene with a single analysis. Moreover, the technique demonstrates high sensitivity to detecting mosaic alleles. The method was validated based on DNA samples with genotyped CGG repeats and specific cuts using CRISPR Cas 9. We further demonstrate that the identification of CGG repetitions using MinION (Oxford Nanopore Technologies) is as precise as conventional techniques based on PCR amplification. We have also been able to observe DNA methylation patterns found on mutated alleles with more than 200 CGG repeats.

The validation and implementation of this new methodology for Fragile X Syndrome diagnostic will be detailed. The application of the approach in diseases involving repeat-expansions, such as ataxic syndromes, muscle atrophy, Steinert's myotonic dystrophy or Huntington's chorea will also be discussed.

## PgmNr 2495: *ZNF407* variants: A new cause of autosomal recessive syndromic intellectual disability?

### Authors:

J. Sobreira <sup>1</sup>; M. Galvão de Oliveira <sup>1</sup>; S. Bragagnolo <sup>1</sup>; N. Sobreira <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centro de Genética Médica, Universidade Federal de São Paulo, São Paulo, São Paulo, Brasil; 2) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, EUA

---

Kambouris et al. (2014) described a consanguineous Arab family with a homozygous damaging *ZNF407*-p.S1685W variant and affected by an apparently novel autosomal recessive disorder characterized by cognitive impairment, failure-to-thrive, hypotonia and dysmorphic features including bilateral ptosis and epicanthic folds, synophrys, midface hypoplasia, downturned mouth corners, thin upper vermillion border and prominent ears, bilateral 5th finger camptodactyly, bilateral short 4th metatarsal bones, and limited knee mobility bilaterally. Here, we describe a 7-year-old Brazilian female, born to a consanguineous couple. She presented with cognitive impairment, microcephaly, strabismus, thick lips, abnormal implantation of her fingers, radio-ulnar synostosis and poor weight gain. As part of the Baylor-Hopkins Center for Mendelian Genomics we performed whole exome sequencing on the proband, her mother and father. Using the PhenoDB Variant Filtering tool we identified a rare, missense, homozygous variant on exon 1 of the *ZNF407* gene (c.G2405T; p.Gly802Val) in the proband that was heterozygous in her mother and father. In gnomAD, there are 20 alleles described with this variant but no description of a homozygous sample. This variant was not described in the Brazilian database, ABraOM (Online Archive of Brazilian Mutations), that contains genomic variants of 609 elderly Brazilian individuals from a census-based sample from the city of São Paulo. This variant's CADD score is 14.5, it is predicted to be damaging by DANN (0.9935), and SIFT (0.005, 0.022), and, its GERP score is 5.84. The *ZNF407* PLi score is 1 and its missense Z-score is 0,86. Based on these findings, we suggest that the homozygous *ZNF407*-p.Gly802Val variant identified in our patient is responsible for her phenotype and that together with the family described by Kambouris et al. (2014), our patient have a novel autosomal recessive syndromic intellectual disability disorder caused by homozygous or compound heterozygous variants in *ZNF407*.

## PgmNr 2496: Diagnostic yield of array CGH and focused clinical exome sequencing in patients with developmental delay/intellectual disability disorders.

### Authors:

G. Méjico <sup>1</sup>; I. Canonero <sup>1,2</sup>; M.F. Gosso <sup>1</sup>; B. Brun <sup>1</sup>; C. Rohr <sup>1</sup>; M.F. Madeira <sup>1</sup>; A.L. Ruggieri <sup>1</sup>; P. Cecatto <sup>1</sup>; I. Sfaello <sup>2</sup>; Z. Sfaello <sup>2</sup>; E. Cuestas <sup>2</sup>; A. Martínez <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Rosario, Héritas, Rosario, Santa Fe, Argentina; 2) Córdoba, Cetes, Córdoba, Córdoba, Argentina

---

**Objective:** To determine the detection rate and diagnostic/clinical advantages of implementing Array (aCGH) followed by Focused Clinical Exome Sequencing (CES) in individuals with unexplained developmental delay/intellectual disability (DD/ID).

**Methods:** Retrospective case series of individuals evaluated from 2016 until May 2019 by the Genetics and Neurology areas of our laboratory, in which diagnostic yield was performed after aCGH testing and followed by Focused CES in those who were negative. All of them have been proven negative by conventional cytogenetical analyses. Inclusion criteria: Studies that used aCGH and Focused CES to identify genetic abnormalities or pathogenic variants in patients with DD/ID and dysmorphia/congenital anomalies, in whom conventional cytogenetic analysis proved negative. Exclusion criteria: patients in whom genetic diagnostic had been carried out by other molecular studies.

**Results:** One hundred and ninety-five patients with DD/ID and dysmorphia were included. All of them were evaluated by karyotyping, aCGH and Focused CES on genes associated to DD/ID. Conventional cytogenetic analysis detected seven structural chromosomal rearrangements, of which only one was a balanced translocation. In forty-seven cases (24,10%) a pathogenic copy number variant was detected by aCGH and only eleven (5,60%) patients resulted in a variant of uncertain significance (VOUS). Otherwise, Focused CES on DD/ID genes yielded forty-eight cases (24,61%) with a pathogenic variant and eleven cases (5,60%) with a VOUS variant.

**Conclusion:** The combined diagnostic yield of causal genetic anomalies detected by aCGH or Focused CES on DD/ID genes in our case series was 60%. In order to improve this result, we are considering switching to clinical Whole Genome Sequencing with which, in addition to detecting CNVs and variants in coding sequences, could also identify intronic genetic variants and/or regulatory elements that we are not currently targeting due to both the intrinsic limitations of the techniques used and the enormous complexity implied in genetic counseling for WGS.

## PgmNr 2497: Deletion breakpoint detection in clinical exome sequencing.

### Authors:

R.I. Torene<sup>1</sup>; K. Galens<sup>1</sup>; C. Borroto<sup>1</sup>; J. Scuffins<sup>1</sup>; G. Douglas<sup>1</sup>; J. Keller-Ramey<sup>1</sup>; K. Beattie<sup>1</sup>; J. Kelsay<sup>2</sup>; G.B. Schaefer<sup>2</sup>; J. Juusola<sup>1</sup>; K. Retterer<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Data Science, Gene Dx, Gaithersburg, MD; 2) University of Arkansas for Medical Sciences, Springdale, AR

---

Copy number variant discovery from clinical NGS data is an essential diagnostic tool. Traditional variant callers, such as GATK, are tuned to detect smaller indel events, but are limited by read length. Read-depth based callers perform better with larger events. Single-exon or sub-exonic deletions are of clinical importance, but can be missed by such variant callers. We, therefore, developed a custom structural variant discovery tool, SCRAMble (Soft Clipped Read Alignment Mapper, [github.com/genedx/scramble](https://github.com/genedx/scramble)) that detects breakpoints from clusters of clipped reads.

We applied breakpoint deletion detection prospectively to our clinical exome diagnostic pipeline for 6385 cases, of which 1301 (20.4%) had a positive molecular diagnosis. Among the positive cases were 60 (4.6%) causative deletions. From this cohort, we confirmed and reported 10 breakpoint deletions, 4 of which were considered causative, 3 were considered possibly causative, and 3 were found in candidate genes. Altogether, causative breakpoint deletions accounted for 4/1301 (0.3%) positive diagnoses and 4/60 (6.7%) positive diagnoses resulting from deletions.

SCRAMble can detect deletions too large to be called as sequence indels and too small to be called by read depth as copy number variants. In fact, we found that 5/10 of the reported breakpoint deletions detected by SCRAMble were not called by an orthogonal NGS method. Among the 5 breakpoint-only deletions were 3 partial-exon deletions, 1 deletion fully encompassing 2 small (<150 bp) exons, and a partial deletion of each of 2 adjacent exons. We note that clinically reportable breakpoint deletions have a median size of 1.6 kb, while clinically reportable read depth deletions have a median size of 185 kb (Wilcox  $p < 0.005$ ).

In one case, an adult female presented with developmental regression in early childhood. Rett syndrome was suspected, but prior targeted sequencing of the *MECP2* gene yielded negative results by both Sanger and NGS panel testing. However, SCRAMble analysis of exome capture data identified a pathogenic, partial exon 4 deletion in *MECP2*, which was confirmed by targeted Sanger sequencing across the breakpoints.

Clipped read calling of deletions is a useful addition to clinical diagnostic NGS capture pipelines and can detect causative variants missed by typical sequence and copy number variant callers. SCRAMble increases the yield of causative deletions by 7.1% (4/56), thus leading to diagnoses, which were not previously evident.



## PgmNr 2498: Angelman like spectrum: Molecular diagnostic performance when the phenotype is evocative of an Angelman syndrome, but methylation is negative.

### Authors:

I. Canonero; G. Méjico; M.F. Gosso; M.F. Madeira; C. Rohr; B. Brun; A. Ruggieri; F. Fay; M. Vázquez

[View Session](#) [Add to Schedule](#)

**Affiliation:** HERITAS, Ocampo Bis 210, Rosario, Argentina

**Introduction:** Angelman syndrome(AS) is characterized by severe developmental delay/intellectual disability(DD/ID), speech impairment, gait ataxia, inappropriate happy demeanor that includes laughing, smiling, excitability, postnatal microcephaly and seizures. Diagnosis is established in patients with clinical diagnostic and molecular genetic testing with deficient expression of the maternally inherited *UBE3A* allele by the methylation study in the 15q11.2-q13 region (80%). *UBE3A* sequencing detects mutations in an additional of 11%. Therefore, molecular genetic testing identifies alterations in approximately 90% of cases. The remaining 10% with AS phenotype are associated with microdeletions /microduplications (CNVs) and point mutations constituting of spectrum Angelman Like Syndrome (ALS).

**Objective:** to determine the diagnostic performance of array Comparative Genomic Hybridization (aCGH) and Focused CES in patients with AS phenotype and negative methylation, and to identify the different entities that make up the spectrum(ALS).

**Methods:** Observational, retrospective, descriptive study of the implementation of molecular techniques and assessment of their diagnostic capacity, which included patients with AS phenotype, evaluated in Medical Genetics, and molecularly studied in the Héritas laboratory during 2016-2019. Molecular studies included methylation study, aCGH and Focused CES. We excluded patients with DD/ID without AS phenotype, and patients with clinical criteria but lacking molecular studies.

**Results:** 46 patients with AS phenotype were included. 21 cases (45,6%) were positive for methylation test (typical AS). By NGS ALS was identified in 14 cases (30.4%): 5 cases of typical Rett syndrome(*MECP2* gene), 2 FOXP1 related-syndromes, 2 GRIN2B related-syndromes, 2 Mowat Wilson syndromes(*ZEB2* gene), 1 Pitt Hopkins syndrome (*CNTNAP2* gene), 1 Phelan McDermid syndrome(PHMDs) and 1 case of epileptic encephalopathy (*STXBP1* gene). aCGH identified 4 pathogenic deletions (8.6%): 1 Kleeftstra syndrome and 3 PHMDs. Only in 7 cases (15%) the molecular diagnostic not be identified.

**Conclusion:** In patients with AS phenotype but with negative methylation test we have been able identify CNVs or pathogenic point mutations in 72% of cases, with a wide diagnostic spectrum of rare syndromes that we can currently diagnose, advise, and perform genetic counseling on thanks to diagnosis techniques based on aCGH testing followed by Focused CES.



## PgmNr 2499: Partial 5p duplication and 15q deletion in a family with a t(5;15)(p13.3;q26.3) translocation.

### Authors:

F.T. Bellucco; B. Favilla; E. Perrone; M.I. Melaragno

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Morphology and Genetics, UNIFESP - Universidade Federal de São Paulo, São Paulo, Brazil

---

Several patients with 5p duplication and 15q deletion have been reported in the literature, involving different chromosome regions and great variations in their clinical features. To the best of our knowledge, partial trisomy 5p associated with monosomy 15q was never reported. Here, we describe a family, characterized by G-banding karyotype and array technique, in which we identified a 30 Mb 5p15.33p13.3 duplication and a 2.5 Mb 15q26.3 deletion in three individuals (two siblings and their paternal uncle), due to a balanced familial translocation between 5p and 15q (found in the father and in the paternal grandmother). The siblings present neurodevelopment delay, intellectual disability, short stature, facial dysmorphisms (such as downslanting palpebral fissures, retrognathia and downturned corners of mouth), short neck, and brachydactyly with 5th finger clinodactyly. Although the boy presents obesity and hyperphagia, like his uncle, the girl does not present these phenotypes. Instead, she presents convergent strabismus and ocular hypertelorism, which were not identified in her brother. Duplications at 5p13.3-5p15.33 region result in mild and relatively indistinct phenotypes, in comparison with duplication at 5p10-5p13.1 segment including the *NIPBL* gene, which results in a more severe phenotype. Chromosome rearrangements involving 15q26, on the other hand, are often related to multiple congenital anomalies, such as growth delay, learning disabilities, unusual facial features and anomalies of the hands and feet. One of the most important genes that seems to be responsible for 15q26 deletion phenotype is the *IGF1R* gene, which has already been associated with growth retardation, developmental delay, clinodactyly, and brachydactyly. Despite presenting such phenotypes, our patients' deletion does not include *IGF1R* gene. They show a combination of phenotypic findings of both 5p duplication and 15q deletions described in the literature thus far. However, the genotype-phenotype correlation is still challenging, since there is no phenotype concordance among the patients, even between both siblings, who have the exact same rearrangement. These data emphasize the importance of detailed cytogenomic and clinical analyses for an accurate diagnosis, prognosis, and genetic counseling and provide an opportunity to improve genotype-phenotype correlations of partial 5p duplication and 15q deletion syndrome patients. Financial Support: FAPESP, Brazil.

## PgmNr 2500: Missense variants in the histone acetyltransferase complex component gene *TRRAP* cause autism and syndromic intellectual disability.

### Authors:

B. Cogné <sup>1,2</sup>; S. Ehresmann <sup>3</sup>; E. Beauregard-Lacroix <sup>3</sup>; J. Rousseau <sup>3</sup>; R. Redon <sup>2</sup>; S. Bézieau <sup>1,2</sup>; S. Kury <sup>1,2</sup>; P.M. Campeau <sup>3</sup>; TRRAP consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical genetics, CHU NANTES, NANTES, France; 2) l'institut du thorax, INSERM, CNRS, UNIV Nantes, Nantes, France; 3) Centre de Recherche du CHU Sainte-Justine et Université de Montréal, Montréal, QC, Canada

---

Acetylation of the lysine residues in histones and other DNA-binding proteins plays a major role in regulation of eukaryotic gene expression. This process is controlled by histone acetyltransferases (HATs) found in multiprotein complexes that are recruited to chromatin by the scaffolding subunit transformation/transcription domain-associated protein (TRRAP). *TRRAP* is evolutionarily conserved and is among the top five genes intolerant to missense variation. Through an international collaboration, 17 distinct de novo or apparently de novo variants were identified in *TRRAP* in 24 individuals.

A strong genotype-phenotype correlation was observed with two distinct clinical spectra. The first is a complex, multi-systemic syndrome associated with various malformations of the brain, heart, kidneys, and genitourinary system and characterized by a wide range of intellectual functioning; a number of affected individuals have intellectual disability (ID) and markedly impaired basic life functions. Thirteen individuals with this phenotype had missense variants between amino acids 1031 and 1159, clustering around the c.3127G>A p.(Ala1043Thr) variant identified in five unrelated individuals. The second spectrum manifested with autism spectrum disorder (ASD) and/or ID and epilepsy and was particularly associated for 7 individuals with missense variants clustering between amino acids 1859 and 1932. Facial dysmorphism was seen in both groups and included upslanted palpebral fissures, epicanthus, telecanthus, a wide nasal bridge and ridge, a broad and smooth philtrum, and a thin upper lip.

RNA sequencing analysis of skin fibroblasts derived from affected individuals skin fibroblasts showed significant changes in the expression of several genes implicated in neuronal function and ion transport. Thus, we describe here the clinical spectrum associated with *TRRAP* pathogenic missense variants, and we suggest a genotype-phenotype correlation useful for clinical evaluation of the pathogenicity of the variants. We will discuss new variants identified in a follow-up study and pathomechanistic perspectives.

## PgmNr 2501: Phasing heterozygous *NPC1* mutations by long range amplicon sequencing.

### Authors:

Q. Meng<sup>1</sup>; C. Brownstein<sup>2,3</sup>; F. Sedlazeck<sup>1</sup>; M. Mahmoud<sup>1</sup>; Y. Han<sup>1</sup>; H. Doddapaneni<sup>1</sup>; S. Jhangiani<sup>1</sup>; J. Gonzalez-Heydrich<sup>3</sup>; M. Hojlo<sup>2,3</sup>; E. Garvey<sup>2,3</sup>; C. Genetti<sup>2</sup>; D. Muzny<sup>1</sup>; R. Gibbs<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center (HGSC), Baylor College of Medicine, Houston, TX; 2) Manton Center for Orphan Disease Research, Division of Genetics and Genomics, Boston Children's Hospital, Boston MA; 3) Tommy Fuss Center for Neuropsychiatric Disease Research, Department of Psychiatry, Boston Children's Hospital, Boston, MA

---

The Niemann-Pick type C1 (NPC1) protein is an intracellular cholesterol transporter which plays an essential role in the maintenance of lipid homeostasis. *NPC1* genetic defects cause NPC1 disease, a fatal autosomal-recessive lipid-storage disorder characterized by progressive neurodegeneration. However, the highly variable disease phenotypes, including the wide range in age of onset, the challenges of annotation of new genetic variants, and the complexity of biochemical and cytologic lab tests to detect the compromised lipid trafficking have all contributed to the current inefficient ascertainment of genotype and phenotype association and the NP-C diagnosis. We report a child with two rare variants in *NPC1* (parents unavailable for testing). Marked behavioral and emotional dysregulation were noted at 18 months. Abilities in using verbal information and spatial and logic-based reasoning have decreased significantly since her previous testing (2.5 years previously). Hallucinations and delusions are partially suppressed on clozapine. To confirm the diagnosis, we established methods to phase the two variants of 16,474 bp apart on chromosome 18, without relying on trio analysis or statistical inference. The first variant, c.547G>A (p.Ala183Thr), has a reported allele frequency <0.001 in public databases (gnomAD, TopMed and 1000 Genomes Project) and is cataloged as of 'uncertain significance' by ClinVar. The second variant, c.1937G>A (p.Arg646His), may be benign as it is a conservative substitution and represented in ExAc 13 times. Despite the lack of functional evidence to support the pathogenicity of both variants, it is still of profound interest to determine if both affected *NPC1* alleles of the patient are in cis position so that the existence of one functional copy of *NPC1* gene can be confirmed. We selected PrimeSTAR GXL Polymerase to amplify the 17.8 kb amplicon from the patient sample. Purified PCR product was sequenced on the Pacbio Sequel with a yield of 19.2 Gb and a mean read length of 23.7 kb, providing a coverage of 28,938x for this genomic region. Using the Whatsap program we showed the two SNVs were on the same haplotype and the patient was not a compound heterozygote for this disorder. The treatment mode has been adjusted appropriately and the study highlights the importance of phase information in clinical diagnostic genetics. Long range amplification and long read sequencing can be used as an efficient and valuable option for this phasing.

## PgmNr 2502: Variability of average relative telomere length measurement using multiplex Luminex assay in DNA from different cell types and blood collection tube.

### Authors:

M.G. Kibriya; F. Jasmine; A. Shaikh; H. Ahsan; B.L. Pierce

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute for Population and Precision Health, Public Health Sciences, Univ Chicago, Chicago, Illinois.

**Background:** Relative Telomere Length (RTL) is a potential biomarker of aging and chronic disease. Previously we developed a non-PCR, probe-based high throughput RTL assay on Luminex platform suitable for large-scale studies. This multiplexed assay (measuring the telomere and the reference gene from the same DNA sample in a single well) requires ~50ng of DNA. We use branched DNA technology of QuantiGene chemistry for signal amplification. The “telomere length” is measured against a standard or reference DNA sample (hence the term “relative”). In this study we examined if different source of DNA from same individual’s blood sample could affect the RTL measurement in this Luminex based assay.

**Material Methods:** Blood from 11 individuals were collected in different tubes - (a) K2-EDTA tube, (b) Trace element tube with EDTA, (c) Heparinized tube, and (d) serum separator SST tube with gel and clot activator. The blood in K2-EDTA tube was used to separate (1) Peripheral Blood Mononuclear Cell (PBMC) and (2) granulocyte cell population using standard Ficoll gradient for DNA extraction. DNA from Whole blood (WB) was derived from (i) trace element EDTA tube, (ii) heparinized tube and (iii) the serum separating tube (SST) with clot activator. DNA was extracted using Flexigene DNA kit (5 samples/person). Quantification was done by Nanodrop spectrophotometry. We tested a total of 55 different DNA samples from the 11 individuals. RTL assay was performed using Luminex method - each was assayed in quadruplicate in four different plates simultaneously in a single batch by same individual.

**Result:** The RTL measurement in DNA from PBMC or Granulocyte was not different (1.039 SD 0.169 vs 1.047 SD 0.167, ANOVA  $p=0.911$ ). RTL measurements of Whole Blood DNA from EDTA, Heparin or SST tubes did not show any difference (0.976 SD 0.125, 0.999 SD 0.188 and 1.011 SD 0.136;  $p=0.857$ ). Regression using General Linear Model (GLM) showed that the variation in RTL measures was mainly attributed to “person-to-person” variation (69%) and the “tube-to-tube” variation contributed only 3.3% of the total variation. The precision of the assay was good to excellent with the Intraclass correlation coefficient (ICC) of 0.896 (95%CI 0.842 - 0.935) among the replicates.

**Conclusion:** The Luminex based multiplex assay for RTL yields similar RTL measures from an individual whole blood sample irrespective of blood collection tube (EDTA, Heparin or SST) or white blood cell type (PBMC or Granulocyte).

## PgmNr 2503: Biallelic variants in *COX4I1* associated with a novel phenotype of developmental regression, intellectual disability and seizures.

### Authors:

N. Pillai <sup>1,2</sup>; N. AlDhaherri <sup>1,2,3</sup>; R. Gosh <sup>1</sup>; J. Lim <sup>2,4</sup>; H. Streff <sup>1,2</sup>; A. Nayak <sup>2,4</sup>; B. Graham <sup>5</sup>; N. Hanchard <sup>1,2</sup>; S. Elsea <sup>1,2</sup>; F. Scaglia <sup>1,2,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Tx.; 2) Texas Children's Hospital, Houston, Tx; 3) Department of Pediatrics, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, UAE; 4) Department of Neurology, Baylor College of Medicine, Houston Tx, USA; 5) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA; 6) Joint BCM-CUHK Center of Medical Genetics, Prince of Wales Hospital, ShaTin, Hong Kong SAR

### Background:

Biallelic variants in *COX4I1* (OMIM: 123864) have been previously described in a single patient who presented with short stature, poor weight gain, dysmorphic features and features of Fanconi anemia. *COX4I1*, located at 16q24.1, encodes the subunit IV isoform 1, COX-IV subunit of cytochrome c oxidase (COX/Complex IV) in humans and other vertebrates. COX plays an important role in oxidative phosphorylation by transferring electrons from cytochrome c to molecular oxygen and contributes to a proton electrochemical gradient and formation of ATP. Here, we describe a novel *COX4I1* variant in two siblings who presented with developmental regression, seizures and pathognomonic changes in brain imaging resembling Leigh syndrome phenotype.

### Case reports:

First sibling is a 3 year old Iraqi male, born at 37 weeks of gestation to consanguineous parents. After initial normal growth and development, his motor skills regressed at 8 months of age. Brain MRI findings were suggestive of Leigh syndrome. He developed seizures at 2 years of age. A trio WES done as a part of his diagnostic evaluation showed a homozygous novel variant in *COX4I1*, c.454C>A (p.P152T). Untargeted metabolomics profile was done on plasma and CSF, which showed elevated lactate along with fumarate suggesting mitochondrial dysfunction.

His older brother, 11 years of age had the same phenotype with regression of skills at 11 months of age followed by seizures at one year of life. He had known familial mutation testing for *COX4I1*, which identified the same variant. A mitochondrial respiratory chain enzyme analysis done on the muscle biopsy specimen showed reduction in complex IV activity to 16% compared to normal controls, meeting modified Walker criteria further confirming the diagnosis of mitochondrial disease. Chromosome breakage studies done on both siblings were normal.

**Conclusions:** Here, we describe siblings with a novel homozygous variant on *COX4I1* presenting with encephalopathy, developmental regression, hypotonia, pathognomonic brain imaging findings resembling Leigh-like syndrome, expanding the known clinical phenotype associated with pathogenic variants *COX4I1*. We also emphasize the importance of using untargeted metabolomics analysis that

could provide functional evidence for pathogenicity of variants in mitochondrial syndromes.

## PgmNr 2504: Unexpected levels of benign heteroplasmic variation: A case of possible biparental inheritance of mtDNA?

### Authors:

L. Schultz-Rogers <sup>1</sup>; E. Perales Clemente <sup>2</sup>; J. Evans <sup>1</sup>; M. Schultz <sup>2</sup>; R. Dhamija <sup>3</sup>; E.W. Klee <sup>1,4</sup>; L. Hasadsri <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN.; 2) Division of Laboratory Genetics and Genomics, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.; 3) Department of Medical Genetics, Mayo Clinic, Phoenix, AZ.; 4) Department of Clinical Genomics, Mayo Clinic, Rochester, MN.

---

Strict maternal inheritance of mitochondrial DNA and subsequent homoplasmy is a well-established tenant of human genetics. On average 24% of individuals contain at least one heteroplasmic position (MAF $\geq$  10%), with generally no more than two heteroplasmic positions observed in a single individual. Multiple heteroplasmic variants are usually explained by sample degradation, contamination, or recent blood transfusion. We report on a 19 year old female with a complex phenotype raising suspicion for a mitochondrial disorder. From 2014 to 2019 the patient had mitochondrial sequencing performed 4 times by 3 different institutions. All 4 tests reported the presence of multiple heteroplasmic positions. The most recent sequencing done at our institution revealed 6 benign homoplasmic and 21 benign heteroplasmic variants observed at levels ranging from 10-20%. While the patient's primary haplogroup is X2, 8 of the heteroplasmic variants are associated with the haplogroup R or with H, J, or U which are all branches off of R. Another 10 variants aren't part of a specific haplogroup but are reported in individuals who are primarily Eurasian and in some cases African. This evidence led us to hypothesize that the heteroplasmic variants are likely derived from the same Middle Eastern haplogroup. As the patient is adopted we don't have data on the parents, however we do know that her mother is Syrian and her father is Iraqi which appears to fit with the two observed haplogroups. We believe these results are unlikely due to contamination as the results were confirmed in a separate urine specimen, neither specimen was run with other samples with the same haplotype, and testing of an independently collected specimen at another institution reported the same results. Neither of these specimens were collected within 4 months of a blood transfusion. Finally, two custom long-range PCR reactions were performed prior to sequencing to exclude interference from any nuclear genes or pseudogenes. Therefore, we propose that one possible explanation for this patient's unusual sequencing results is biparental inheritance of mtDNA. Recent work by Luo et al. provided convincing evidence of paternally inherited mitochondrial DNA in several unrelated families. Whether the unexplained excessive heteroplasmic positions in our patient are caused by biparental inheritance, a technical issue, or some other anomaly, it is a phenomenon worth investigating for its implications in clinical testing.

# PgmNr 2505: A novel liquid chromatography-tandem mass spectrometric method reveals decreased levels of guanidino compounds in the plasma of patients with arginase 1 deficiency treated with pegzilarginase.

## Authors:

G.A. Diaz <sup>1</sup>; A. Schulze <sup>2</sup>; M.C. McNutt <sup>3</sup>; E.L. Teles <sup>4</sup>; J.L. Merritt II <sup>5</sup>; G.M. Enns <sup>6</sup>; S.P. Batzios <sup>7</sup>; R.L. Conway <sup>8</sup>; S. Ferrati <sup>9</sup>; S. Rowlinson <sup>9</sup>; S. Alters <sup>9</sup>; J.E. Wooldridge <sup>9</sup>; R.T. Zori <sup>10</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) University of Toronto and The Hospital for Sick Children, Toronto, ON, Canada; 3) UT Southwestern Medical Center, Dallas, TX; 4) Centro Hospitalar de São João, Porto, Portugal; 5) University of Washington, Seattle, WA; 6) Stanford University, Stanford, CA; 7) Great Ormond Street Hospital NHS Trust, London, UK; 8) Children's Hospital of Michigan, Detroit, MI; 9) Aeglea BioTherapeutics, Austin, TX; 10) University of Florida, Gainesville, FL

## Background

Arginine and guanidino compounds (GCs) accumulate in the plasma of patients with arginase 1 deficiency (ARG1-D) and may contribute to the neurologic phenotype of the disease. Pegzilarginase (PZA) treatment has been shown to sustainably reduce arginine levels in a phase 1/2 study of patients with ARG1-D. To date, determination of GC levels has been challenging due to inadequately sensitive methodologies. We assessed plasma GC levels in patients with ARG1-D treated with PZA using a novel liquid chromatography-tandem mass spectrometric (LC-MS/MS) method.

## Methods

Blood samples for GC ( $\alpha$ -keto- $\delta$ -guanidinovaleric acid [GVA]; N $\alpha$ -acetyl-L-arginine [NAAArg]; R,S-argininic acid [ArgA]; guanidinoacetic acid [GAA]) analyses were collected from patients with ARG1-D receiving up to 7 single ascending doses of PZA every 2 weeks followed by a weekly repeat dose for up to 8 weeks before entering into a long-term extension study. Samples were collected in a tube containing dipotassium ethylenediaminetetraacetic acid (K<sub>2</sub>EDTA), N<sup>w</sup>-hydroxy-nor-arginine (nor-NOHA) and mannitol, and processed to plasma. Various concentrations of GCs were added to calibration standards prepared in a surrogate matrix, and quality control samples were prepared in pooled K<sub>2</sub>EDTA human plasma containing nor-NOHA, PZA, and mannitol. Fifty microliters of plasma were treated with 100  $\mu$ L of 1N hydrochloric acid containing internal standards, followed by 400  $\mu$ L of ice-cold 10% trichloroacetic acid. Supernatant derived from vortex-mix and centrifugation was analyzed by LC-MS/MS using Shimadzu Nexera UHPLC-Applied Biosystems/MDS Sciex API 5500.

## Results

Sixteen patients were recruited for the study; 69.0% were females and median age (range) was 15 (5–31) years. Baseline mean GVA, NAAArg, ArgA, and GAA levels (standard deviation) were 5.1 (1.67), 1.1 (0.48), 2.8 (1.21), and 3.1 (1.11)  $\mu$ mol/L, respectively, which decreased to 2.7 (1.26), 0.7 (0.46), 1.8 (0.87), and 2.0 (1.21)  $\mu$ mol/L, respectively, 2 weeks after the last repeat dose. The concentration range for validation of this method was 0.03–10  $\mu$ M for GVA, NAAArg, and ArgA, and 0.300–100  $\mu$ M for



GAA.

### *Conclusions*

Treatment with PZA was associated with marked reduction in plasma GC levels in patients with ARG1-D. Further studies are needed to confirm the relationship between GC levels and patients' outcomes. This LC-MS/MS method was effective in measuring GCs in this patient group and could represent a useful tool in conditions with altered GC levels.

## PgmNr 2506: Mitochondrial DNA mutations in Leigh syndrome: Our experience.

### Authors:

P.G.A. Dawod <sup>1,2</sup>; B. Rovcanin <sup>2</sup>; M. Brankovic <sup>2,3</sup>; A. Marjanovic <sup>2,3</sup>; M. Jankovic <sup>2,3</sup>; I. Novakovic <sup>2,3</sup>; F.I. Abdel Motaleb <sup>1</sup>; J. Jancic <sup>2,4</sup>; V. Kostic <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Biochemistry and Molecular Biology, Faculty of Medicine, Ain Shams University, Cairo, Egypt.; 2) Faculty of Medicine, University of Belgrade, Serbia.; 3) Neurology Clinic, Clinical Centre of Serbia.; 4) Child and Adolescent Neurology and Psychiatry Clinic, Belgrade, Serbia.

---

**Background:** Leigh syndrome (LS) is severe neurological disorder characterised by progressive loss of mental and movement abilities i.e. psycho-motor regression. First symptoms of LS become apparent in the early childhood with lethal outcome within two to three years. LS has heterogeneous genetic aetiology and could be associated with mutations in more than 75 genes. About 20% of LS cases are caused by mitochondrial DNA (mtDNA) mutations, in which the mutated or impaired mitochondria show defective mitochondrial protein synthesis with sequentially reduced activity of respiratory chain complex I.

**Objective:** The aim of this study was to perform mutational screening of entire mtDNA in Serbian patients clinically diagnosed with Leigh syndrome.

**Material and Methods:** Five patients included in this study were recruited from the child neurology hospitals in Belgrade, Serbia. All examined individuals had characteristic clinical presentation suggesting the presence of LS. Molecular genetic analysis of entire mtDNA was performed by Sanger sequencing.

**Results:** We have detected mtDNA mutation in one out of five LS patients. In that case sequence analysis showed homoplasmic m.8993 T>G substitution in *ATP6* gene. The T>G transition of the mitochondrial nucleotide pair 8993 causing change amino acid Leucine to Arginine is one of the most frequent mtDNA mutations in LS. Numbers of other variants were detected in various fragments of mitochondrial genome.

**Conclusion:** This is the first case of mtDNA mutation in LS in Serbia. Understanding of the genetic background of mitochondriopathies may further facilitate the diagnostic approach and open perspectives to future, possibly therapies.

## PgmNr 2507: Identification of an insertional duplication of the *AMT* gene in a patient with glycine encephalopathy.

### Authors:

L. Wang<sup>1</sup>; H. Dai<sup>1,2</sup>; L. Fleming<sup>3</sup>; W. Zhu<sup>1</sup>; S. Peacock<sup>1</sup>; K. Curry<sup>3</sup>; Y. Cao<sup>2</sup>; J. Condie<sup>3</sup>; G. Wang<sup>1</sup>; V. Patel<sup>1</sup>; L. Meng<sup>1,2</sup>; Y. Wang<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Baylor Genetics, Houston, TX.; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) St Luke's Children's Specialty Center, Boise, ID

---

Glycine encephalopathy, an inborn error of glycine metabolism, is defined by the deficiency in the activity of the mitochondrial glycine cleavage enzyme system and the aberrant accumulation of the undegraded glycine. It can be caused by the mutations in the genes encoding the protein components of the glycine cleavage system with an autosomal recessive inheritance pattern. Most patients have elevated glycine in plasma and CSF, and present with phenotypes such as lethargy, seizures, hypotonia, abnormal EEG and mental retardation. Death in infancy can be common for patients with the classic neonatal form of the disorder. Mutations in the *AMT* (aminomethyltransferase) gene, which encodes the T protein of the glycine cleavage system, account for approximately 20% of the affected patients. Here we report one newborn female with symptoms reminiscent of glycine encephalopathy. Whole exome sequencing only detected one novel missense variant c.665G>T (p.R222L) in the *AMT* gene which was inherited from the patient's mother. Follow-up array CGH using the MitoMet v3.3 oligonucleotide microarray identified a 0.7 Kb duplication including exon 5 and partial exon 6 of the *AMT* gene. Further gap-PCR and Sanger sequencing analysis at junction breakpoints confirmed that this duplication was inserted into intron 3 of the same gene. This intragenic duplication in the *AMT* gene reshuffles the exons and probably results in an absent or truncated protein. Based on the whole exome data, the duplication is in trans configuration with the missense variant. Copy number changes involving the *AMT* gene are very rare. To our knowledge, this is the first report of duplication of this gene ever identified in patients with glycine encephalopathy.

## PgmNr 2508: The correlation study between phenotypes and genotypes of 174 cases of mitochondrial diseases caused by mitochondrial DNA variations in children.

### Authors:

Y. Shi; F. Fang; Z. Liu; D. Shen; L. Dai; W. Zhang; J. Li; X. Ren; T. Han; C. Ding

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Neurology, Beijing Children's hospital, Capital Medical University, National Center for Children's Health, China, Beijing, China

**Objective:** To delineate the phenotypes and genotypes of children with mitochondrial DNA (mtDNA) variations, and analyze the correlation between them.

**Method:** The clinical and genetic data of patients with mitochondrial diseases caused by mtDNA variations in the department of Neurology, Beijing Children's Hospital from Jan 2001 to Feb 2019 was collected retrospectively. Pathogenicity analysis was performed in the three cases with variants of uncertain significance.

**Results:** There were total 174 diagnosed cases with mtDNA variations collected in our study. Male: Female=1.3:1. The average age of presentation was  $(5.67 \pm 4.07)$  years old, ranging from the neonatal to 15.1 years old. Twelve phenotypes were discovered in our study, including MELAS (n=82), MILS (n=64), MM (n=6), MERRF (n=5), LHON (n=3), KSS (n=5), MELAS and LS overlap syndrome (n=3), MERRF and LS overlap syndrome (n=2), RIRCD, DEAF, DMDF and Mitochondrial encephalomyopathy with deafness, epilepsy and developmental delay in 1 case respectively.

In our study, there were two kinds of mtDNA variations, including large-scale deletions and point mutations. The five deletions were all KSS. Among the point mutations, there were 13 genotypes and 29 allele changes. tRNA was the most common. The mutation ratios of Co<sub>2</sub> were higher than Co<sub>1</sub> and tRNA.

In MELAS and MELAS/LS, m.3243A>G was the "hot point", accounting for 92.7%. In MILS and LS/MERRF, *MT-ATP6* and m.9176T>C was the most common. MM were all caused by m.3243A>G; MERRF were all caused by m.8344A>G. The most common phenotype of tRNA variations was MELAS. The most common phenotype of Co<sub>2</sub> variations was MILS. And the only phenotype of Co<sub>1</sub> variations was MILS. The three cases with variants uncertain significance in our study were respectively m.9396G>A (MELAS), m.3955G>A (MILS), m.10407A>G (Mitochondrial encephalomyopathy with deafness, epilepsy and developmental delay).

### Conclusions:

MELAS and MILS were the most common phenotypes, accounting for 83%. tRNA was the most common in which the most common phenotype was MELAS. The most common phenotype of Co<sub>2</sub> variations was MILS. Moreover, the only phenotype of Co<sub>1</sub> variations was MILS. In MELAS and MELAS/LS, m.3243A>G was the "hot point". The mutation threshold was calculated into 35.63%. The most common pathogenic variations of MILS were Co<sub>2</sub> variations. The most common phenotype and allele change was *MT-ATP6* and m.9176T>C, only accounting for 14.1%.

## PgmNr 2509: Detection of succinic semialdehyde dehydrogenase deficiency in newborn dried blood spots.

### Authors:

M. Brown <sup>1</sup>; C. Turgeon <sup>2</sup>; P. Rinaldo <sup>2</sup>; P. Ashcraft <sup>3</sup>; E. Arning <sup>3</sup>; T. Bottiglieri <sup>3</sup>; A. Pop <sup>4</sup>; G.S. Salomons <sup>4</sup>; J.-B. Roullet <sup>1</sup>; K.M. Gibson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pharmacotherapy, College of Pharmacy and Pharmaceutical Sciences, Washington State University, Spokane, WA; 2) Biochemical Genetics Laboratory, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Baylor Scott and White Research Institute, Institute of Metabolic Disease, Dallas, TX; 4) Department of Clinical Chemistry, Metabolic Unit, VU University Medical Center & Amsterdam Neuroscience, Amsterdam, the Netherlands

**BACKGROUND & RATIONALE** – Succinic semialdehyde dehydrogenase deficiency (SSADHD) is a rare inborn error of  $\gamma$ -aminobutyric acid (GABA) metabolism, originating from inherited mutations in *ALDH5A1*, which manifests global developmental delays, motor dysfunction, intellectual disability, seizures and other behavioral abnormalities (Attri, 2017). Biomarkers for the disease include elevated GABA and its derivative  $\gamma$ -hydroxybutyric acid (GHB) in biological fluids. With the GHB-lowering antiepileptic vigabatrin being the only targeted therapeutic available, treatment is primarily symptomatic. Although there is no defined life expectancy of patients with SSADHD, there has been an increasing number of reports of sudden unexpected death in epilepsy (Pearl, 2011; Horino, 2016).

To date, SSADHD is not included in newborn screening (NBS) panels, as GABA and GHB are not routinely measured in newborn dried bloodspots (DBSs). As a result, families experience an approximate 2-year diagnostic delay resulting in profound stress, anxiety, and delayed treatment. **SPECIFIC AIMS** – We sought to determine whether SSADHD can be detected in the newborn period with novel metabolic analyses and biomarkers. Our objectives were two-fold: 1) to determine if SSADHD is associated with a unique metabolic profile in the newborn and post-newborn periods, and 2) to determine if GHB can reliably be detected in SSADHD patient DBSs.

**METHODS/DESIGN** – We obtained historical newborn DBSs from 7 SSADHD patients including 1<sup>st</sup> and 2<sup>nd</sup> screens from 3 of the 7 patients (age at collection between 48 and 340 hrs), and collected 17 post-newborn DBSs (0.8-38 years of age; median, 8.2 years) from patients with confirmed diagnosis of SSADHD. DBS amino acids, acylcarnitines and creatine were determined using currently available NBS methodology and compared to archival age-matched control data points (CLIR, Mayo Clinic). DBS GHB was determined using UPLC-MS/MS (Forni, 2013).

**RESULTS** – In newborn DBSs, GHB was 111-767 mM, exceeding the 99.99%-tile (78 mM). In contrast, C2-, C3-, C4- and C4-OH carnitines, creatine and ornithine concentrations were below normal. Low levels of C2- and C4-OH carnitines, ornithine, histidine and creatine were also observed in post-newborn DBSs.

**CONCLUSION** – Early detection of SSADHD in newborn DBS now appears plausible with quantitation of short-chain acylcarnitines, creatine and ornithine used as potential 1<sup>st</sup> tier screening tool and GHB as a 2<sup>nd</sup> tier method.

## PgmNr 2510: Deletion 21q22.3 with loss of 119 genes in a Colombian boy: A case report.

### Authors:

L.J. Moreno Giraldo <sup>1,2,3,4</sup>; D. Arturo-Terranova <sup>1,4</sup>; J.M. Satizabal Soto <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) postgraduate in biomedical sciences - Medical Genetics, Universidad del Valle, Cali, Colombia; 2) Universidad Santiago De Cali; 3) Universidad Libre; 4) Grupo de investigación Enfermedades Congénitas del Metabolismo

---

Chromosomal constitutional imbalances are frequently associated with learning disabilities, deterioration, congenital anomalies and deterioration of growth. Chromosomal aneuploidies, particularly those that involve chromosome 21, are one of the alterations most prevalent human chromosomes, and occur in approximately 1 of every 700 births. We present the case of a two-year-old boy with facial dysmorphism, deterioration of growth and developmental disorder. A karyotype in G banding and chromosomal analysis was performed using a comparative genomic hybridization matrix to make its correct diagnosis. The karyotype was performed in 15 metaphases, which reported a masculine 46, XY, r (21); Ring chromosome 21, which leads to the loss of genetic material. CGH evidenced the deletion of a region of approximately 4.5Mb on chromosome 21q22.3 with pathogenic classification; this alteration affects approximately 119 genes. This is a rare deletion in the long arm of chromosome 21. These types of deletions have been reported with clinical variability, but common features include craniofacial, skeletal malformations, cardiac and genitals with cognitive impairment. It is reported that under normal conditions there is a high interaction between genes product of the terminal deletion on chromosome 21q22.3. The condition in genes of the KRTAP family, characterized as a class of human genes that promote proliferation in epithelial cells in mammals, lead to significant damage to necessary biological functions such as keratinization, cell and epithelial cell differentiation, tissue and epithelial development, in the same way that the genes of the TFF family act. Also, the genes of collagen play important roles in maintaining the structure and function of the extracellular matrix. The members of the family of Collagen VI, such as COL6A1, COL6A2, form distinct networks of microfibrils in the connective tissue and are responsible for interacting with other components of the extracellular matrix; these last genes are closely related to POFUT2, ADARB1, SLC19A1 and PCBP3 in charge of processes mesenchymal and epithelial. The present analysis will help to determine a better phenotype-genotype correlation analysis in cases with partial monosomy 21. The future implementation of exome and RNA sequencing techniques, and analysis of their proteomic expression in a clinical context could lead to better analysis and interpretation of the genotype-phenotype correlation in cases similar to that described.

## PgmNr 2511: Phasing of genomic rearrangements reveal involvement of both homologous chromosomes in pre- and post-zygotic events.

### Authors:

C. Carvalho <sup>1</sup>; C. Beck <sup>1</sup>; Z. Akdemir <sup>1</sup>; F. Sedlazeck <sup>2</sup>; Q. Meng <sup>2</sup>; J. Hu <sup>2</sup>; H. Doddapaneni <sup>2</sup>; Z. Chong <sup>3</sup>; E. Chen <sup>1</sup>; M. Withers <sup>1</sup>; S. Jhangiani <sup>2</sup>; A. English <sup>2</sup>; Y. Han <sup>2</sup>; K. Chen <sup>4</sup>; D. Muzni <sup>2</sup>; R. Gibbs <sup>2</sup>; C. Shaw <sup>1</sup>; P.J. Hastings <sup>1</sup>; J.R. Lupski <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Molecular Human Genetics, Baylor Col Medicine, Houston, Texas.; 2) Human Genome Sequencing Center, BCM, Houston TX; 3) Department of Genetics and the Informatics Institute, the University of Alabama at Birmingham, Birmingham, AL; 4) The University of Texas MD Anderson Cancer Center, Houston, TX

---

Copy number variation (CNV), including duplications and triplications formed by microhomology-mediated break-induced replication (MMBIR) has been shown to be generated mainly by intrachromosomal events, i.e. only segments originating from the same homologue are part of the end product. The hypothesized reason is that template-switches (TS) within the same chromosome are thought to be facilitated by the physical proximity in the nucleus due to linkage as well as to *cis* intrachromosomal interactions. Autosomal triplications followed by runs of homozygosity (ROH) are clinically relevant exceptions: they are generated post-zygotically due to interhomolog TS which can lead to imprinting perturbations and associated diseases. To gain mechanistic insights into the formation of *de novo* autosomal copy number gains and complex rearrangements, we studied 11 trios (patient + parents) in which probands were referred with congenital malformations carrying rearrangements involving 17p11.2. Trios were studied by combining array comparative genomic hybridization (aCGH), long-and short-reads DNA sequencing technologies including a targeted 7Mb Illumina short read spanning the 17p interval and Pacific Biosciences SMRT sequencing approach. Phasing of the copy number gains were obtained by B-allele frequency analysis provided by trio joint-calling, whereas array data and PacBio were used to resolve CNV breakpoint junctions. Intrachromosomal rearrangements were observed in 6 out of 11 cases and interchromosomal events in 1 out of 11, confirming that a majority of events occurred within the same homolog. Surprisingly, though, in 4 out of 11 (36%) cases the CNVs constitute a mix of both intra- and inter-homologue amplified segments rather than originating from only one ancestral chromosome, all of them formed by pre-zygotic mutagenesis. Intriguingly, only a subgroup of the template-switching events led to CNVs, similar to what was previously observed for post-zygotic autosomal triplications. Our results indicate that high-resolution B-allele frequency analysis is a helpful tool to uncover CNV formation and reveal TS events that can result in increased rearrangement complexity. Importantly, DNA segments originating from two distinct parental chromosomes contributing to the formation of the same rearrangement strongly support a TS model of generating amplifications and have further implications for human diseases, including those resulting from perturbations of imprinting.

## **PgmNr 2512: Cousins with 4p- syndrome mosaicism and parents with balanced translocation: The importance of the family history in cases of structural chromosomal disorders.**

### **Authors:**

J.I.C.F. Neri; E.A. Gil

[View Session](#) [Add to Schedule](#)

**Affiliation:** Centro de Reabilitação Infantil, SESAP, Natal, Rio Grande do Norte, Brazil

---

4p- syndrome or Wolf-Hirschhorn syndrome (WHS) arises from a contiguous gene deletion of the distal short arm of chromosome 4 and is characterized by craniofacial features, growth impairment, intellectual disability, and seizures. Usually cases of chromosomal mosaicism occur due to post-zygotic error, for this reason a negligible risk of recurrence of that condition in the future offspring of the parents during their genetic counseling is reported. We present here the case of two cousins, a girl born in 2014 and a boy born in 2013, both presenting WHS in mosaic [46, XY, del (4) (p15.3) / 46, XY] with 20% of normal cells. They came for evaluation separately when they were about one year old and without identifying their kinship. The girl's father and the boy's mother are siblings and have balanced translocation. These cases illustrate the importance of the good anamnesis carried out with the deepening of the family history for similar cases as bases to a adequate genetic counseling.



## PgmNr 2513: Diagnostic approaches and yields following the detection of long continuous stretches of homozygosity (LCSH) in SNP microarrays.

### Authors:

Y. Fan; L. Wang; Y. Sun; W. Qiu; H. Zhang; B. Xiao; L. Han; R. Xu; F. Lin; X. Gu; Y. Yu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Xinhua Hospital affiliated to Shanghai Jiaotong University School of Medicine, shanghai, Shanghai, China

**Purpose:** Single-nucleotide polymorphism (SNP) microarrays can identify long continuous stretches of homozygosity (LCSH). LCSH is associated with risk of recessive disorders and uniparental disomy (UPD). However, the diagnostic utility of LCSH warrants further clinical investigation. The goal of this study is to evaluate the diagnostic yields of different approaches following LCSH detection, and recommend optimal reporting thresholds.

**Methods:** A cohort of 2226 index cases (mostly pediatric patients) with SNP microarray results were recruited. LCSH was defined as uninterrupted copy-number neutral homozygosity above 3Mb. A total of 1944 cases harboring at least one LCSH were further divided to three groups: 1) "Suspected consanguinity" group (total LCSH covered over 1.5% of autosomal region, n=18), followed by whole-exome sequencing (WES); 2) "Imprinting related" group (cases with LCSH on chr 6,7,11,14,15 and 20, n=26), followed by either methylation-specific MLPA or STR linkage analysis for UPD testing; 3) "Not-imprinting related" group (cases with LCSH on single chromosome not involving imprinting disorders, n=17), followed by WES. The diagnostic yields in these three groups and the LCSH pattern in the cases with confirmed diagnosis were analyzed. Potential reporting thresholds and corresponding missed diagnosis and false negative rates were presented.

**Results:** In suspected consanguinity group, WES revealed diagnostic variants on LCSH in 11.1% cases, and another 22.2% cases were diagnosed with LCSH-irrelevant variants, including de novo and autosomal dominantly transmitted variants. For cases harboring LCSH on single chromosome, similar yields after WES were found - 11.8% diagnosed with variants on prior detected LCSH, 17.6% diagnosed with variants irrelevant to LCSH. For cases with LCSH on imprinting related chromosomes, subsequent analysis confirmed UPD diagnosis in 48% cases, and the most prevalent diagnosis is maternal UPD of chr15 associated with Prader-Willi syndrome. Pattern analysis suggested a threshold of ~5Mb for terminal LCSH and ~15Mb for interstitial LCSH was associated with maximal UPD yield and minimal false positives.

**Conclusion:** WES following the detection of LCSH may identify diagnostic variants, but the yield is low and possibly not related to the original LCSH finding. UPD analysis following LCSH detection on imprinting-related chromosomes is highly recommended after setting optimal reporting thresholds.

## PgmNr 2514: Whole genome sequencing (WGS) as a first-line diagnostic test: Its success is in the details.

### Authors:

A. Kaplun; N. Neerman; C. Stanley; G. Faust

[View Session](#) [Add to Schedule](#)

**Affiliation:** Variantyx Inc., Framingham, Massachusetts.

---

### Background:

WGS is well positioned to become the clinical diagnostic standard for rare genetic disorders due to the benefits inherent in PCR-free, genome-wide sequencing combined with its continually decreasing cost. Initial published reports are supportive of this paradigm shift, yet all WGS assays are not created equal. We've undertaken the challenge to develop and validate a clinical WGS pipeline that supports simultaneous analysis and interpretation of multiple variant types, resulting in a clinically available alternative to sequential testing.

### Methods:

Clinical grade whole genome sequencing (WGS) is conducted on an Illumina platform using the Illumina TruSeq DNA PCR-Free Library Preparation Kit. With the PCR-free protocol, significantly less average coverage is required compared to panels and exomes. Specifically, we find that 30X mean mappable coverage provides a highly uniform sequencing depth with 97.3% of nucleotides covered at  $\geq 8X$  and 99.4% of HGMD and ClinVar annotated variants covered at  $\geq 8X$  - a depth that is sufficient for specific and sensitive variant calling. Following alignment of reads, a combination of open source and proprietary algorithms call small sequence change variants, structural variants, mitochondrial variants and tandem repeat expansions. Variants are subjected to annotation and filtering before being presented to our clinicians for interpretation within the context of the patient's clinical symptoms and medical history.

### Results:

In 2018, we obtained full CLIA accreditation and CAP certification for all four components of our WGS pipeline: small sequence changes, structural variants, mitochondrial variants and short tandem repeats. We will present a combination of secondary and tertiary validation results which demonstrate >99% sensitivity, specificity and PPV for single nucleotide variants; >95% sensitivity and specificity for indels up to 50 nucleotides; >96% clinical sensitivity for structural variants; the ability to detect mitochondrial heteroplasmy levels down to 5%; and >99% sensitivity for tandem repeat expansions spanning >20 known pathogenic loci. We will additionally present cases representative of those processed in the first 18 months that the full Genomic Unity™ test has been available for clinical use.

## PgmNr 2515: Genetic variants in a cohort of Colombian patients with VACTER/VACTERL association.

### Authors:

O.M. Moreno Niño <sup>1</sup>; A.I. Sánchez <sup>1</sup>; A. Rojas <sup>1</sup>; G. Giraldo <sup>1</sup>; A. Herreño <sup>1</sup>; A. Paredes <sup>1</sup>; M. Olaya <sup>1,2</sup>; Y. Vargas <sup>2</sup>; J.C. Prieto <sup>1</sup>; F. Suarez <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Medicine, Instituto de Genética Humana, Pontificia Universidad Javeriana Bogotá, Carrera 7, 40-62, Bogotá, Colombia; 2) Hospital Universitario San Ignacio, Carrera 7, 40-62, Bogotá, Colombia,

---

**Introduction:** VACTER-L association comprises different multisystem congenital malformations including vertebral anomalies, anal atresia, cardiac defects, trachea-oesophageal fistula with or without oesophageal atresia, renal malformations and limb defects. Prevalence has been estimated in 1:7000 to 1:40000 individuals. This heterogeneous condition has many overlapping defects with other malformation syndromes and it is usually considered as a diagnosis of exclusion. Fanconi anemia (FA) patients often exhibit birth defects suggesting VACTERL association. Because of this, FA must be considered among one of the first differential diagnosis. **Objective:** The aim of this study was to determine genetic abnormalities through banding-cytogenetics analysis, molecular cytogenetics (aCGH), and chromosomal breakage test for FA in patients with diagnostic criteria of VACTER/VACTER-L association and establish the relationship with the phenotype. **Methods:** This was a prospective study. Data regarding clinical features was obtained by physical examination or from clinical reports, and then it was analyzed to describe all clinical aspects. Complete cytogenetic investigation based on chromosomes with G-band staining was performed in all patients, as well as chromosome breakage studies for FA using standard procedures. **Results:** We evaluated 18 patients who met criteria for VACTER/VACTER-L association, 11 were female and 7 were male. Age range between 1 month and 16 years. 13(72.2%) patients presented with vertebral anomalies, 8(44.4%) anal atresia, 14(77.8%) congenital heart defects, 8(44.4%) oesophageal atresia, six of them having TE fistula, 10(55.6%) renal anomalies, and 10(55.6%) limb defects. Karyotypes were normal for all patients and chromosomal breakage tests were negative for FA. aCGH was performed in 15 patients, eight of them had alterations classified as Variants of Unknown Significance (VUS), and in one patient dup(17)(q12) classified as pathogenic and in other patient del(15)(q11.2) classified with variable effect.

**Conclusion:** All patients in our study exhibit a wide clinical spectrum. Chromosome aberrations and FA were ruled out in all of the patients. aCGH revealed microalterations that in two cases could explain the phenotype, while the VUS variants have not yet been described in patients VACTER/VACTER-L; these VUS affect some genes important in cell function and could be involved in embryonic development.

## PgmNr 2516: A quantitative, sensitive method for the confirmation of low-level chimeric RNAs identified by RNA-fusion detection in inherited conditions.

### Authors:

N. Vidal-Folch <sup>1</sup>; L. Schultz-Rogers <sup>3,4</sup>; G.R. Oliver <sup>3,4</sup>; E.W. Klee <sup>1,3,4,5</sup>; D. Oglesbee <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 2) Department of Medical Genetics, Mayo Clinic, Rochester, MN; 3) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 4) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 5) Department of Clinical Genomics, Mayo Clinic, Rochester, MN

---

**Introduction:** Chimeric RNAs (fusion transcripts) are frequently found in cancer tissues where their oncogenic, diagnostic and therapeutic relevance is widely recognized. Recently, fusion transcripts have also been reported in sporadic cases of developmental and neurological phenotypes including brain malformation, intellectual disability, schizophrenia, spastic paraplegia and autism spectrum disorder, as well as in normal physiology. When using fusion-detection methodology, confirmation of candidate hits with an orthogonal method is critical to elucidate true transcripts from false positive events. Traditional techniques, such as targeted PCR and Sanger sequencing, are often used for hit validation, but these platforms lack sufficient sensitivity and specificity for low-level chimerism. Here, we describe our droplet digital PCR (ddPCR) approach to RNA fusions that provides unsurpassed analytical sensitivity down to 4 supporting reads per 100 million total reads.

**Methods:** Chimeric RNAs were identified by a modified fusion-transcript detection algorithm coupled with a categorization workflow. A total of 11 candidate fusion transcripts with strong phenotypic relevance were discovered from 9 undiagnosed patients suspected of inherited disease. For fusion confirmation, G-block constructs for each candidate were synthesized. Each fusion transcript was amplified from patient cDNA. DdPCR was performed using EvaGreen Supermix, QX100 droplet generator and QX200 droplet reader (Bio-Rad).

**Results:** G-blocks showed positive amplification of all candidate RNA-fusion constructs. Five fusion transcripts (with 10, 11, 15, 31 and 32 supporting reads) were confirmed by both traditional PCR and ddPCR, providing the clinical diagnosis for two previously unsolved cases. Three fusion transcripts (with 4, 7 and 13 supporting reads) were confirmed only by ddPCR. Three fusion transcripts (with 5, 12 and 15 supporting reads) were not confirmed by either methodology. While the latter species appear to be artefactual hits, all three candidates contained reads across exon-exon boundaries consistent with RNA splicing products.

**Conclusion:** This ddPCR approach is a highly sensitive and specific method for confirming chimeric RNAs discovered in rare inherited disorders.

## PgmNr 2517: The VetSeq Study: Follow-up results from a feasibility study of expanded exome sequencing in the Veterans Health Administration.

### Authors:

L.E. Hull <sup>1</sup>; S. Advani <sup>1</sup>; M.H. Hawley <sup>2</sup>; A.M. Oza <sup>2</sup>; S. Amr <sup>2</sup>; P. Pomianowski <sup>1</sup>; J.L. Vassy <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) VA Boston Healthcare System, Boston, Massachusetts.; 2) Laboratory for Molecular Medicine, Partners HealthCare; 3) Department of Medicine, Harvard Medical School; 4) Division of General Internal Medicine and Primary Care, Brigham and Women's Hospital

---

### Background

Although exome sequencing (ES) has demonstrated utility in many diagnostic contexts, the range of secondary results that should be returned from ES is an open question. We are conducting a feasibility study of ES in the VA Boston Healthcare System (VABHS), returning indication-based and secondary results.

### Methods

In the VetSeq Study, providers refer patients for indication-based ES from a CLIA-certified lab through a research protocol. The report includes indication-related results, secondary pathogenic or likely pathogenic variants in >4600 genes associated with monogenic disease risk, carrier status variants, and pharmacogenomic results. Patients complete surveys and interviews before ES and 3 months after receiving results from their referring providers.

### Results

Two patients have completed the study. P1 is a 72-year-old man with decades of sensory neuropathy and myalgias. ES identified a pathogenic p.Arg298Cys variant in *LMNA* associated with Charcot-Marie-Tooth hereditary neuropathy type 2. Though some *LMNA* variants are associated with dominantly inherited neuromuscular disease, P1's variant is typically associated with autosomal recessive inheritance and was thus inconclusive for his presentation. Family testing determined the variant was absent in a daughter and grandson with similar symptoms. In his baseline interview, P1 had hoped ES would identify a diagnosis and possible treatment not for himself but possibly for his family or for other patients. In his follow-up interview, he was uncertain ES met these expectations, given the apparent lack of definitive familial benefits or scientific advancement. P2 is a 33-year-old woman with a history of severe thrombocytopenia in her and 1 of 2 sons. In her baseline interview, she hoped that ES would suggest a treatment to stabilize their platelet counts. ES identified a p.Tyr346Cys variant in *ETV6* that was present in her affected son but not her unaffected son. *ETV6* has been associated with autosomal dominant thrombocytopenia, but this variant was classified as uncertain significance due to insufficient evidence of pathogenicity. In her follow-up interview, P2 said that although ES did not conclusively identify a cause for the family's thrombocytopenia, learning she carries a pathogenic *HFE* variant associated with hereditary hemochromatosis was useful.

### Conclusions

It is feasible to introduce ES to VABHS clinical care. Patients might find value in outcomes unrelated to their own diagnosis and treatment.

## PgmNr 2518: Hepatic hemangioma in a patient with *de novo* mutation in NF1 gen.

### Authors:

M. Rodriguez Castro <sup>1</sup>; L.J. Moreno Giraldo <sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Universidad Libre, cali, Colombia; 2) Universidad del Valle, Medical Genetics; 3) Universidad Santiago de Cali; 4) Grupo de investigación Enfermedades Congénitas del metabolismo

---

Neurofibromatosis type 1 is a neurocutaneous disorder caused by mutation in the suppressor gene of neurofibromin 1 - NF1 located in the chromosome 17q11.2. This disease is of autosomal dominant inheritance with very variable expressivity and with progressive multisystemic commitment. Exist de novo mutations, these are generally related to alleles inherited from parents of very advanced age. More than 1289 mutations have been found in this gene but few have been related to the disease phenotype. The diagnosis is characterized by presence of more than 6 café-au-lait macules greater than 5 mm in pre-patients pubescent and 15 mm in diameter in post pubescent patients, axillary freckles or inguinal, 2 or more neurofibromas of any type or 1 or more neurofibromas plexiform, optic gliomas, 2 or more Lish nodules, bony dysplasias and presence of tumors malignancies of the peripheral nerve sheath. The most frequent neoplasms in the pediatric age are intracranial gliomas (20%), and other high and low grade gliomas, followed by malignant tumors of the peripheral nervous sheath, and rhabdomyosarcomas. With the objective of to determine the molecular etiology of a patient with hepatic hemangioma and neurofibromatosis through exomic studies. We studied the clinical case of a 3-year-old male patient with multiple café-au-lait spots throughout the body, of different sizes, freckles in the armpits and inguinal area, broad forehead, eyebrows arched, palpebral fissures down, hypertelorism and anteverted ears. With history of respiratory distress at birth, dermatitis, rhinitis, asthma, alteration in psychomotor development, with no family history of the disease and finding of Hepatic hemangioma in abdominal magnetic resonance. Clinical exome was requested where a de novo mutation is detected, with variant pathogenic in the NF-1 gene: c1381C>T; p.Arg461Ter in heterozygous, autosomal state dominant. The use of imaging techniques is essential to assess the

extension and progression of subcutaneous and plexiform neurofibromas, but not used routinely and there is no consensus on which patients should be monitored. the early presentation of hepatoblastoma without neurofibromas or hemangiomas cutaneous lesions have not been previously described; only one case has been described in the literature of cutaneous hemangioma associated with hepatic hemangioma in a patient with neurofibromatosis type 1 with this same mutation.

## PgmNr 2519: De novo KAT6B mutation identified with whole-exome sequencing in a boy with Say-Barber-Biesecker-Young-Simpson syndrome.

### Authors:

J.M. Satizabal Soto <sup>1,2,4</sup>; D. Arturo-Terranova <sup>1,4</sup>; L.J. Moreno Giraldo <sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) postgraduate in biomedical sciences - Medical Genetics, Universidad del Valle, Cali, Colombia; 2) Universidad Santiago de Cali; 3) Universidad Libre; 4) Grupo de investigación Enfermedades Congénitas del Metabolismo

---

Say-Barber / Biesecker / Young-Simpson syndrome (SBBYSS; OMIM 603736), variant of Ohdo syndrome is a rare syndrome with multiple congenital anomalies. The clinical diagnosis is usually based on a phenotype with a mask-like face severe blepharophimosis a bulbous nasal tip, feeding problems, long thumbs and big toes, and dislocated or hypoplastic patellae. . this disease is also associated with delayed development and intellectual disability, which are often severe. Many affected infants have weak muscle tone (hypotonia) that leads to breathing and feeding difficulties. We present the clinical case of a Colombian boy of 4 years with dysmorphic features, short stature, suction and swallowing disorder, bilateral hearing loss, bilateral cryptorchidism, and developmental delay. The boy has normal echocardiogram and echocardiogram results. Previous genetic testing (array-CGH, PTPN11 FOXL2 mutation analysis) gave normal results. We performed whole-exome sequencing and identified on chromosome 10 a heterozygous nonsense mutation in the KAT6B gene, NM\_012330.3 : c.3349 C>T (p.Gln1117\*). The mutation led to a premature stop codon. The analysis carried out in the parents confirms that the variant is *de novo* in the patient. To our knowledge, this mutation has not been reported before. Being a stop codon, this mutation is classified as probably pathogenic. several aspects found support for the causal relationship between the mutation in KAT6B and the correct diagnosis of our patient. KAT6B mutations have been described in other patients with SBBYSS. This disease is not found in the updated list of orphan diseases in Colombia. In the literature it is reported that typically the syndrome is caused by *de novo* pathogenic variants. The disease manifests in childhood with severe hypotonia and feeding difficulties. Skeletal problems such as hypoplasia are associated and in 50% of patients structural heart defects and dental anomalies are reported. In many cases it is also associated with intellectual disability and severe developmental delay. If a patient is suspected of SBBYSS, its recommend sequencing the KAT6B gene to find the correct diagnosis. In these cases where the CGH does not provide enough information, the complete sequencing of exome allows to provide a diagnosis correct, genetic counseling and personalized therapy.



## PgmNr 2520: Evaluation of clinically relevant CNVs from WES data in the cytogenomic routine.

### Authors:

E.A. Zanardo <sup>1</sup>; F.P. Monteiro <sup>2</sup>; A.T. Dias <sup>1</sup>; S.N. Chehimi <sup>1</sup>; Y.G. Oliveira <sup>1</sup>; L.A. Costa <sup>2</sup>; L.L. Ramos <sup>2</sup>; G.M. Novo-Filho <sup>1</sup>; G.F.S. Carvalho <sup>1</sup>; M.M. Montenegro <sup>1</sup>; M.V. Souza <sup>1</sup>; A.B. Freitas <sup>3</sup>; A.M. Nascimento <sup>1</sup>; J.P. Kitajima <sup>2</sup>; F. Kok <sup>2</sup>; L.D. Kulikowski <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina FMUSP, Universidade de São Paulo; 2) Mendelics Análise Genômica; 3) Departamento de Obstetrícia e Ginecologia, Faculdade de Medicina FMUSP, Universidade de São Paulo

---

**Introduction:** Detection of single nucleotide variants (SNVs) and copy number variations (CNVs) is essential for patient genotyping in cytogenomic diagnostic. Recently the screening of CNVs from whole exome sequencing (WES) data has become a common practice for routine diagnostic and allowed improving the accuracy of CNVs detection associated with phenotypes. **Methods:** In this study, we evaluated 38 patients with developmental delay and/or multiple congenital malformations and negative WES (Nextera Rapid Capture Exomes - Illumina) results for pathogenic SNVs. The WES data were submitted to ExomeDepth software in order to identify potential relevant CNVs. We performed genomic array using Infinium CytoSNP850K BeadChip (Illumina) and BlueFuse Multi v4.3 (BlueGnome - Illumina) software in all samples for comparison and confirmation of the CNVs. Those were classified as benign, VUS or pathogenic, and only the pathogenic CNVs were evaluated. **Results:** We identified 44.7% of the patients with pathogenic CNVs detected by both techniques including deletions and duplications in different chromosomal regions. Only one patient showed pathogenic CNV detected by array and missed by WES. Also, only the genomic array technique revealed cases with regions of homozygosity (8%) that led to correct diagnostic conclusion. Additionally, we performed a reevaluation and reanalysis of WES for normal results in attempt to reclassify variants originally classified as VUS based in relevant literature, updated available mutation/variant databases, and the latest research. The reanalysis improved the detection and cytogenomic diagnostic and showed 29.4% of the patients with SNV in a specific gene/region related to clinical phenotype. The extraction of CNVs information from WES data is an advantageous approach since it can improve the cost-effectiveness and reduces the number of genomic tests required in routine diagnostic. **Grants:** FINEP-CT INFRA0160/12-SP8; CAPES-Finance Code 001.

## PgmNr 2521: A comprehensive genome analysis approach using whole exome sequencing and microarrays reveals novel diagnoses in autopsy negative patients.

### Authors:

C. Chung <sup>1</sup>; A. Nesbitt <sup>1</sup>; D. Watson <sup>1</sup>; K. Stumm <sup>1</sup>; M. Harr <sup>1</sup>; X. Wang <sup>1</sup>; M. Khan <sup>1</sup>; J. Glessner <sup>1</sup>; H. Hakonarson <sup>1,2,3</sup>; A. Santani <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA.; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.; 4) Department of Pathology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

---

**Background** Thousands of deceased patients do not receive genetic testing and the underlying molecular etiology, despite a comprehensive autopsy, remains unknown. In the pediatric setting, a post-mortem molecular diagnosis can help families determine recurrence risk and support the development of management strategies for similarly affected relatives. This study aims to leverage a comprehensive genomics strategy including chromosomal microarray (CMA) and/or whole exome sequencing (WES) to determine the utility of this strategy in identifying pathogenic variants in a deceased pediatric cohort. **Methods** Subjects were enrolled in the research biorepository at the Center for Applied Genomics at Children's Hospital of Philadelphia. Eighty-one patients with unexplained death were selected based on pre-determined criteria related to phenotype and family history (55.6% male, 44.4% female). We performed CMA and WES analysis using a series of custom bioinformatics pipelines and algorithms. **Results** The cohort's phenotypic spectrum included multiple congenital anomalies (85.2%), neurodevelopmental disorders (45.7%) and craniofacial disorders (7.4%); 38/81 (46.9%) presented with overlapping features. CMA was performed on 81 patients and a pathogenic/likely pathogenic copy number change was found in 16% (13/81). A molecular diagnosis was identified using WES in 2/11 subjects initially sequenced. Results on WES analysis for the remaining 57 undiagnosed individuals will be presented. One patient, with prenatal findings of polycystic kidneys, oligohydramnios, pulmonary hypoplasia, mild increased NT and ventriculomegaly, had findings in 2 potentially relevant genes: a homozygous variant in *GLYCTK* (D-glyceric aciduria, AR) and a homozygous variant in *CEP290* (Meckel syndrome/Bardet-Biedl syndrome/Joubert syndrome, AR). The other patient, with renal dysplasia, gastroesophageal reflux, severe intellectual disability, focal epilepsy, microcephaly, small optic nerves and facial dysmorphism, carried a pathogenic variant in *DYRK1A* (Mental retardation 7, AD). **Conclusion** A comprehensive post-mortem genomics approach can help establish a molecular diagnosis for patients with uninformative autopsy results. We will provide the molecular diagnostic yield in this cohort and describe the unique challenges associated with performing accurate interpretation on a deceased patient cohort.

## PgmNr 2522: The spectrum of pathogenic variants and clinical features of individuals undergoing *IRF6* gene testing.

### Authors:

H. Cox; J. Dong

[View Session](#) [Add to Schedule](#)

**Affiliation:** Geneticists, PreventionGenetics, Marshfield, WI

---

**Background:** Pathogenic variants in the *IRF6* gene are associated with autosomal dominant isolated orofacial clefting, Van der Woude Syndrome (VWS) and Popliteal Pterygium Syndrome (PPS). Features of VWS may include lower lip pits, cleft lip, cleft palate, cleft uvula, submucous cleft palate and hypodontia. The PPS phenotype is more severe and may include orofacial clefting, abnormal external genitalia, syndactyly, ankyloblepharon, syngnathia, and pyramidal skin on the hallux. To date, over 300 pathogenic variants have been reported in the *IRF6* gene. This study reports the clinical features and spectrum of pathogenic variants in individuals referred for clinical *IRF6* gene testing. **Methods:** This study included 126 individuals who underwent *IRF6*-gene testing at a single clinical laboratory. Clinical information was obtained from provider-completed test request forms. Of the tested individuals, 54% (68/126) were male and 46% (58/126) were female with a median age at testing of 2.7 years of age (range 0.01-51 years). The majority of individuals (64%, 81/126) were ascertained based on a personal history of orofacial clefting. Clinical suspicion of VWS or PPS was reported for 13% (16/126) and 4% (5/126) of individuals, respectively. A positive family history of orofacial clefting was reported in 33% (42/126) of individuals. The *IRF6* gene was analyzed using Sanger sequencing and Next Generation Sequencing methods. Copy Number Variation (CNV) analysis was performed for 29% (37/126) of tested individuals. **Results:** Overall, 40 unique *IRF6* pathogenic variants were identified in 51 individuals (41%, 51/126) during the course of clinical testing. Pathogenic variants were detected in 37% (38/94) of individuals with a personal history of orofacial clefting, VWS or PPS. Of the pathogenic variants detected, 35% (14/40) localized to the winged-helix DNA-binding domain and 22.50% (9/40) localized to the SMIR/IAD protein-binding domain. Pathogenic missense (13/40), nonsense (12/40), frameshift (11/40) and splicing (4/40) were observed. No pathogenic CNV events were detected. **Conclusion:** The phenotype of individuals with an *IRF6* pathogenic variant was consistent with previous reports, with the majority of individuals presenting with a personal history of orofacial clefting. The overall positive rate was 41% with a clinical sensitivity of 37% for individuals with a personal history of orofacial clefting, VWS or PPS.

## PgmNr 2523: Heterozygous double gross deletions in *JAG1* in a patient with Alagille syndrome.

### Authors:

T. Togawa <sup>1</sup>; S. Ito <sup>1</sup>; K. Ohashi <sup>2</sup>; K. Ito <sup>1</sup>; T. Endo <sup>3</sup>; T. Sugiura <sup>4</sup>; S. Saitoh <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Pediatrics, Kainan hospital, Yatomi, Japan; 3) Department of Pediatrics, Komaki prefectural hospital, Komaki, Japan; 4) Department of Pediatrics, Sugiura Kodomo Clinic, Hekinan, Japan

---

Alagille syndrome (ALGS; OMIM 118450) is an autosomal-dominant multisystem disorder affecting the liver, heart, face, eyes, skeleton, and other organs. The prevalence is estimated to be approximately 1 in 30,000 live birth. ALGS is caused by a mutation in one of two genes in the Notch signaling pathway, *JAG1* or, rarely, *NOTCH2*. *JAG1* is located in 20p12 and consists of 26 exons. A mutation in *JAG1* can be identified in approximately 90% of clinically diagnosed patients with ALGS. To date, over 500 *JAG1* mutations have been reported, and approximately 10% of the mutations were gross deletions. Here, we present a patient with clinically diagnosed ALGS and with novel heterozygous double gross deletions in *JAG1*. The patient was a Japanese male from nonconsanguineous parents. He represented intrahepatic cholestasis and was introduced to our institution at 3 years in order to test molecular genetic analyses. His clinical features showed as follows; intrahepatic cholestasis with elevated serum liver enzyme, characteristic facial features including a high broad forehead and a triangular face, and dysplastic kidney. These features fulfilled clinical diagnostic criteria of ALGS. We carried out our targeted next-generation sequencing using the AmpliSeq and Ion Torrent Personal Genome Machine (ThermoFisher Scientific) system as we previously established. Sequenced reads were analyzed through Ion Reporter software (IR, ThermoFisher Scientific). Consequently, the IR called 2 heterozygous gross deletions in *JAG1*; 1616bp deletion in exon10 and 659bp deletion in exon 25 encompassing to exon 26. To eliminate false-positive mutations, we ascertained those by multiplex ligation-dependent probe amplification (MLPA) analysis using the SALSA MLPA probemix P184-C2 *JAG1* (MRC-Holland), and we successfully validated those two regions. Next, we validated the origin of the two deletions with the same method using gDNA from his parents. The examination revealed that his mother carried the both deletions. Thus, we confirmed that the double gross deletions were on the same allele which was transmitted from his mother. After the molecular testing, we carefully checked his mother, and she showed slightly elevated liver enzyme, agenesis of one kidney, and hypertension. These were compatible with clinical aspects of ALGS. In conclusion, this is the first report that a patient with clinically diagnosed ALGS had doubly mutated allele, including 2 gross deletions, in *JAG1*.

## PgmNr 2524: When a SV anomaly can hide another.

### Authors:

B. Chalot <sup>1,2,3</sup>; A.-L. Mosca-Boidron <sup>1,2</sup>; A. Vitobello <sup>1</sup>; D. Callegarin <sup>1,2</sup>; S. Verdez <sup>1,2</sup>; N. Marle <sup>1,2</sup>; E. Tisserant <sup>1</sup>; A. Plagos <sup>1</sup>; A.-M. Guerrot <sup>4</sup>; J. Puechberty <sup>5</sup>; P. Chambon <sup>6</sup>; G. Joly-Hélas <sup>6</sup>; J.-F. Deleuze <sup>7</sup>; R. Olaso <sup>7</sup>; A. Boland <sup>7</sup>; M. Poulleau <sup>2</sup>; T. Jouan <sup>1</sup>; C. Poe <sup>1</sup>; C. Thauvin-Robinet <sup>1,3</sup>; L. Faivre <sup>1,3</sup>; Y. Duffourd <sup>1</sup>; P. Callier <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Inserm UMR 1231 GAD, Université de Bourgogne Franche-Comté, Dijon, France; 2) Laboratoire de Génétique Chromosomique et Moléculaire, Plateau technique de biologie, Centre Hospitalier Universitaire Dijon Bourgogne, 21079 Dijon, France; 3) Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'Interrégion Est, FHU TRANSLAD, Centre Hospitalier Universitaire Dijon Bourgogne, 21079 Dijon, France; 4) Service de Génétique, CHU Rouen; 5) Service de Génétique Médicale, CHU Montpellier, Montpellier, France; 6) Laboratoire de Cytogénétique, CHU Rouen; 7) Centre National de Recherche en Génomique Humaine (CNRGH), Evry, France.

---

**Introduction:** Structural variations (SVs) such as apparently balanced translocations, insertions and inversions are evidenced by karyotype. Some publications showed the potential role of whole genome sequencing (WGS) to characterize SVs (Redin et al. 2017, Schluth-Bolard et al. 2018).

**Methods :** The goal is to use WGS to characterize SVs in 10 patients with SVs and an abnormal phenotype, but also with the potential to identify other hits. Whole genome analyses were performed using short reads with 40X depth. Bioinformatic analysis of the raw data was performed with Lumpy algorithms for translocations and inversions and ControlFreeC for CNVs.

**Results:** The WGS short-read identified the 10/10 SVs and allowed to propose a diagnosis for two cases. In one case, the translocation breakpoint interrupted the *GRIN2B* gene. In the second case, the *NBEA* gene was interrupted by the same mechanism. The disruption of these two genes was compatible with the phenotype of these two patients. Five cases had translocations that disrupted unknown genes and one without gene. For two cases, WGS analysis found no gene disruption or effect position, but found rare variants of *FBN1* and *CD151* genes respectively responsible for the phenotypic features. The first one suffered of marfanoid syndrome with cardiac complications associated with *de novo* t(1;5)(q25;q11). The second one presented with epidermolysis bullosa associated with an maternal inv(3)(p14.3p24.2).

**Conclusion:** The use of the WGS is therefore particularly effective in characterizing the breakpoints of the SVs. We report two rare cases of SVs associated with pathogenic SNV which explain the clinical presentation. These SNVs hidden by SVs will be more frequently discovered with the increased use of WGS in France thanks to the generalization of pangenomic NGS.

## PgmNr 2525: Emerging value of rapid whole genome sequencing in complex congenital heart disease with multisystem involvement.

### Authors:

S. Girgis; J. Sasaki; B. Totapally; M. Diaz-Barbosa; K. Schain; J. Runyon; S. Lala Kana; P. Jayakar

[View Session](#) [Add to Schedule](#)

**Affiliation:** Nicklaus Children's Hospital, Miami, Florida

---

**Background:** The utilization of rapid whole genome sequencing (rWGS) can provide a diagnosis in the setting of multisystem involvement of unclear etiology. However, at this time, the cost-benefit ratio of this test must be weighed in the diagnostic decision-making process. We present a patient with complex congenital heart disease, pulmonary hypoplasia and diaphragm abnormality, lacking a unifying diagnosis. Whole genome sequencing revealed de novo likely pathogenic c.2042C>A variant in *MYRF* gene, consistent with the patient's presentation of cardiac-urogenital syndrome.

**Case Report:** The patient presented is a female born at 38 weeks, to a primigravid mother, with hypoplastic left heart syndrome (HLHS) diagnosed prenatally at 20 weeks gestation. Postnatally, she had findings consistent with hypoplastic right lung secondary to congenital diaphragmatic hernia (CDH). She was initially placed on hospice care and discharged home at 4 days of age, then transferred to our facility at 10 days of age. Physical exam was unremarkable. Brain MRI, newborn screen, karyotype and microarray were normal. Given multisystem involvement, rWGS was ordered, which was significant for a de novo, likely pathogenic, heterozygous c.2042C>A (p.Ala681Asp) variant in the myelin regulatory factor (*MYRF*) gene.

**Discussion:** The *MYRF* gene encodes a precursor of the transcription factor which mediates autocatalytic cleavage of myelin regulatory factor, leading to activation of the transcription of central nervous system myelin genes. Variants in the *MYRF* gene lead to multisystem involvement including the cardiac, pulmonary and urogenital systems, namely cardiac-urogenital syndrome. With congenital heart disease being linked to many syndromes as in this case, to suspect an underlying genetic defect in the setting of negative microarray and karyotype is not unfounded. Further pursuit with rWGS may be of benefit when a unifying diagnosis is suspected. The utilization of rWGS earlier in the diagnosis process of a multisystem disease, as opposed to after non-diagnostic subsequent serial testing and panels, may provide a potential cost and time-saving alternative for early treatment modalities including appropriate genetic counseling and management of care.

**Conclusion:** The utilization of rWGS can be of benefit when presentation does not readily conform to a specific syndrome or disease but is suggestive of an underlying genetic etiology.

## PgmNr 2526: Pathogenic variants in 77 individuals with Coffin-Siris syndrome.

### Authors:

F. Sekiguchi <sup>1</sup>; N. Miyake <sup>1</sup>; Y. Tsurusaki <sup>1,2</sup>; N. Matsumoto <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Graduate school of medicine, Yokohama City University, Yokohama, Kanagawa, Japan; 2) Faculty of Nutritional Science, Sagami Women's University, Sagamihara, Kanagawa, Japan

---

Coffin-Siris syndrome (CSS, MIM#135900) is a congenital disorder characterized by coarse facial features, intellectual disability, and hypoplasia of fifth fingers and nails. Pathogenic variants were found in genes encoding BRG1- or HBRM-associated factors (BAF) complex. To date, more than 120 patients with pathogenic variants in 9 BAF-related genes have been reported. We previously reported 71 patients of whom 39 had pathogenic variants. We newly recruited 182 CSS-suspected patients. In 182 patients together with previously unresolved 32 patients, we found 77 pathogenic variants in 77 patients. Pathogenic variants in *ARID1B*, *SMARCB1*, *SMARCA4*, *ARID1A*, *SOX11*, *SMARCE1* and *PHF6* were found in 47, 8, 7, 6, 4, 1 and 1 patients, respectively. In addition, pathogenic CNVs including duplication of *SMARCA2* were found in 3 patients. No novel genes with pathogenic variants have been identified. Of note, we first found a partial deletion of *SMARCB1* in a CSS patient.



## PgmNr 2527: SouthSeq: Sequencing NICU newborns in the south.

### Authors:

M. Thompson <sup>1</sup>; K. Bowling <sup>1</sup>; C. Finnila <sup>1</sup>; S. Hiatt <sup>1</sup>; M. Amaral <sup>1</sup>; K. East <sup>1</sup>; M. Cochran <sup>1</sup>; V. Greve <sup>1</sup>; W. Kelley <sup>1</sup>; D. Gray <sup>1</sup>; J. Lawlor <sup>1</sup>; A. Cannon <sup>2</sup>; L. Hendon <sup>3</sup>; H. Janani <sup>4</sup>; L. Merin <sup>2</sup>; C. Tuura <sup>3</sup>; K. Laborde <sup>4</sup>; A. Hurst <sup>2</sup>; R. Savich <sup>3</sup>; S. Spedale <sup>4</sup>; G. Barsh <sup>1</sup>; B. Korf <sup>2</sup>; G. Cooper <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama; 2) University of Alabama at Birmingham, Birmingham, AL; 3) University of Mississippi Medical Center, Jackson, MS; 4) Woman's Hospital, Baton Rouge, LA

---

SouthSeq is a Clinical Sequencing Evidence-Generating Research (CSER2) effort to perform genome sequencing (GS) for infants with congenital anomalies or other signs of a genetic disorder. We aim to provide genomic testing across rural/minority populations that are historically under-represented in clinical and genomic research. Recruitment sites are located at hospitals in Louisiana, Mississippi and Alabama. The primary goals are to provide early diagnoses, demonstrate the utility of GS as a frontline test with potential to improve outcomes, and equip non-genetics providers to return GS results.

Sequencing and analysis of the infant's DNA is conducted at HudsonAlpha. When genetic variation associated with the child's symptoms is identified, clinical Sanger confirmation of the variant in the child and available parental DNA is conducted to determine inheritance. Also, families have the option to receive pathogenic/likely pathogenic (P/LP) secondary findings, which focus on medically actionable genes.

To date, 106 newborns have received sequencing and analysis. The average age of enrolled newborns is 39 days and 57% are male. 50% of enrollees are from minority populations. 60% of newborns were enrolled along with both biological parents. Genetic variation associated or potentially linked to the congenital phenotype has been observed in 49 newborns (46%); 35 newborns have received a result that is likely diagnostic (33%), 26 of which harbor variation in an autosomal dominant disease. Notably, four babies harbored variation in *CHD7* associated with CHARGE syndrome. The remaining nine likely diagnostic cases have recessive disease. Fourteen infants harbored variants leading to uncertain diagnosis (13%). Interestingly, we detected heterozygous genetic variation in recessive disease genes (*PALB2*, *TMCO1*, *TAPT1*, *MUT*, *ATPAF2*, and *PMM2*) in five newborns in which their phenotypes match well with those previously associated with the gene; however, no variation in the other allele could be identified. Eighty-eight percent of families elected to receive secondary findings. Three families (1.9%) received P/LP variation across four genes, including *APOB*, *MSH6*, *ACTC1*, and *LDLR*.

Preliminary results suggest that we have successfully deployed GS in NICU settings to obtain early genetic diagnosis for babies born with multiple congenital anomalies. We hope and expect that early genetic diagnosis will improve long-term outcomes for newborns with rare disease.



## PgmNr 2528: CNVs as genetic modifiers for congenital heart disease in a Brazilian 22q11.2 deletion syndrome cohort.

### Authors:

M. Zamariolli <sup>1</sup>; A.G. Dantas <sup>1</sup>; N. Nunes <sup>1</sup>; M. Moysés-Oliveira <sup>2</sup>; L. Spíndola <sup>1</sup>; V.L. Gil-Da- Silva-Lopes <sup>3</sup>; D.C.Q. Soares <sup>4</sup>; C.A. Kim <sup>4</sup>; M.I. Melaragno <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Morfologia e Genética, UNIFESP, São Paulo, Brazil; 2) Center for Genomic Medicine, Massachusetts General Hospital, Boston, USA; 3) Department of Medical Genetics, Faculdade de Ciências Médicas, Universidade Estadual de Campinas, Campinas, São Paulo, Brazil; 4) Genetics Unit, Instituto da Criança, Universidade de São Paulo, São Paulo, Brazil

---

The clinical heterogeneity found in patients with the 22q11.2 deletion syndrome is not directly correlated to the different classes of deletion sizes related to this locus (3 Mb, 1.5 Mb and atypical deletions). Congenital heart disease represents one of the most clinically relevant phenotypes and the etiology of its variability among patients is not fully understood. Although the dose-sensitive genes mapped in the deleted region are the most important candidates for the phenotype, it has been suggested that more complex molecular mechanisms are involved in the phenotypic variability of the syndrome, including variants in the 22q11.2 intact region and genetic modifiers in other regions of the genome, i.e. structural and single nucleotide variants. Previous work in the literature has shown that both rare and common copy number variants (CNVs), outside of the 22q11.2 deleted region, could play a role in this variability. Even though some candidate genes and regions have been described as modifiers for the cardiac phenotype, they account for only a small proportion of this incomplete penetrance. This suggests that existence of other modifiers for this condition and, therefore, studies in new cohorts are important to identify them. To identify new CNVs modifiers of the cardiac phenotype, in this study, we generated SNP-array data from 215 individuals, being 99 22q11.2 deletion patients (65% with congenital heart disease and 35% with normal cardiac phenotype) and 116 controls using the CytoScan HD platform (Affymetrix). Modifying CNVs, outside of the 22q11.2 region, were detected using two different tools (PennCNV and CNV Workshop). Only the CNVs detected by both algorithms were considered for subsequent analysis. The genetic content of the CNVs revealed candidate genes that could be associated with the risk or protection for cardiac malformations in the patients, potentially acting as genetic modifiers for congenital heart disease in the 22q11.2 deletion. To our knowledge, this is the first comprehensive study in a Brazilian 22q11.2 deletion syndrome cohort. This study will leverage valuable knowledge for the better understanding of the role of genetic modifiers involved in the expressivity of cardiac alterations in the 22q11.2 deletion syndrome.

## PgmNr 2529: *De novo* genome assembly and phasing for undiagnosed conditions.

### Authors:

J.T. Shieh <sup>1,2</sup>; M. Penon-Portmann <sup>2</sup>; M. Levy-Sakin <sup>3</sup>; A. Slavotinek <sup>1,2</sup>; H. Perry <sup>6</sup>; D. Belefond <sup>2</sup>; A. Sharo <sup>4</sup>; R. Gallagher <sup>1,2</sup>; S.E. Brenner <sup>4</sup>; Z. Qi <sup>5</sup>; J. Yu <sup>5</sup>; J. Tenney <sup>2</sup>; O. Klein <sup>1,2,6</sup>; D. Martin <sup>8</sup>; P. Kwok <sup>1,3,7</sup>; D. Boffelli <sup>8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 2) Division of Medical Genetics, Department of Pediatrics, Benioff Children's Hospital, University of California San Francisco, San Francisco, CA; 3) Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA; 4) Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA; 5) Department of Laboratory Medicine, University of California San Francisco, San Francisco, CA; 6) Craniofacial Center, Craniofacial Biology and Department of Orofacial Sciences, University of California San Francisco, San Francisco, CA; 7) Department of Dermatology, University of California San Francisco, San Francisco, CA; 8) Children's Hospital Oakland Research Institute, Benioff Children's Hospital Oakland, Oakland, CA

---

Genome analyses that generate very long haplotypes may facilitate diagnostic testing, particularly through the detection of structural variants. In this study, we probed the ability of *de novo* genome assemblies to improve identification of genetic lesions responsible for undiagnosed conditions. We hypothesized that a single genetic test could detect all types of genetic variants, including coding, noncoding, and structural variants. We used a hybrid linked read/optical mapping approach on undiagnosed cases, using a trio design. Potential diagnostic candidate variants were verified using secondary testing and reviewed by a multidisciplinary team. Clinically significant structural rearrangement and copy number variants were detected, and we were able to determine novel phasing data that discerns allele-specific mutations. We conclude that *de novo* genome assemblies can provide detailed diagnostic information beyond that provided by typical clinical sequencing technologies.

## PgmNr 2530: Diagnostic yield of array-CGH in Colombian patients: Beyond the American College of Medical Genetics recommendations.

### Authors:

A. Ortiz <sup>1,2</sup>; I. Riveros-Barrera <sup>1</sup>; B.V. Arevalo-Jaimes <sup>1</sup>; D. Valero-Rubio <sup>1</sup>; J.M. Galvez <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Gencell Pharma, Bogotá. Colombia; 2) Clínica de Marly, Bogotá. Colombia; 3) Fundación CardiolInfantil-Instituto de Cardiología, Bogotá. Colombia

---

**Introduction:** Array-based comparative genomic hybridization (aCGH) is a method for high-resolution evaluation of chromosomal copy number variants (CNVs). In the clinical setting, the American College of Medical Genetics (ACMG) recommends aCGH as a first-tier test in patients with global development delay (DD), intellectual disability (ID), autism spectrum disorders (ASD), and congenital malformations (CM). This study evaluated the diagnostic yield of aCGH in a group of Colombian patients in relation to the ACMG indications. **Methods:** A retrospective study of postnatal aCGH cases evaluated in a Colombian genetic laboratory from April 2017 to March 2019 was performed. Clinical significance of CNVs was determined using open access databases (DGV, Decipher, SFARI, ClinGen). A medical record review was conducted to collect clinical patient's data. Frequency and clinical correlation analysis were carried out. **Results:** Of the 606 included patients (61,6% male, 38,4% female) 93,8% had one or more of the ACMG recommendations, being CM associated to DD/ID (41,2%) and CM only (23,5%), the most common. In the studied population, 124 pathogenic variants were identified in 106 patients (17,5%): 18,1% pathogenic variants among patients with one or more ACMG criteria (n=569) vs 8,1% within patients without ACMG indications (n=37). Interestingly, in this last group, 3,4% had short stature and/or epilepsy as a medical reason for aCGH, which showed a diagnostic yield of 17% and 24%, respectively, when associated to an ACMG recommendation. In patients with ASD, we found a 6% diagnostic performance, similarly to the 7-9% previously described. Among pathogenic variants, 69,4% were deletions and 30,6% duplications, contrary to VUS where 34,1% were deletions and 65,9% duplications, suggesting that deletions have a stronger and better studied deleterious effect than duplications. Most frequently affected chromosomes were 1, 7, 17, 22 and X, corresponding to the main known chromosomal syndromes (e.g. del-22q11). **Conclusions:** The aCGH diagnostic yield was 17,5% in this study, similar to the 18% previously reported. Patients following ACMG recommendations have a higher probability to have pathogenic findings than patients without any of those indications. Based on the high yield identified in epilepsy and short stature, we recommend that aCGH should be also considered as a first-tier test for patients with these clinical findings.

## PgmNr 2531: Importance of cytogenomic investigation of pigmentary mosaicism associated a global developmental delay in patients with blood karyotype normal result.

### Authors:

Y. Gasparini <sup>1,2</sup>; M.M. Montenegro <sup>1,2</sup>; G.M. Novo-Filho <sup>1</sup>; J.R.M. Ceroni <sup>2</sup>; A.M. Nascimento <sup>1,2</sup>; E.A. Zanardo <sup>1</sup>; S.N. Chehimi <sup>1,2</sup>; A.T. Dias <sup>1</sup>; G.F.S. Carvalho <sup>1</sup>; M.V. Souza <sup>1</sup>; A.B. Freitas <sup>3</sup>; F.A.R. Madia <sup>1</sup>; R. Honjo <sup>2</sup>; C.A. Kim <sup>2</sup>; L.D. Kulikowski <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratorio de Citogenomica, Departamento de Patologia, Faculdade de Medicina FMUSP, Universidade de Sao Paulo, Sao Paulo, SP, BR; 2) Unidade de Genetica, Departamento de Pediatria, Instituto da Crianca, Hospital das Clinicas HCFMUSP, Faculdade de Medicina, Universidade de Sao Paulo, Sao Paulo, SP, BR; 3) Departamento de Obstetrícia e Ginecologia, Faculdade de Medicina FMUSP, Universidade de Sao Paulo, Sao Paulo, SP, BR

---

Pigmentary mosaicism is associated with different cutaneous patterns of pigmentation in skin cells. Most of the cases present additional extracutaneous abnormalities including global developmental delay usually affecting the central nervous system and/or the musculoskeletal system. Usually, patients with those clinical manifestations begin the genetic investigation using peripheral blood karyotype. But recent studies showed that despite a karyotype from peripheral blood without alterations, genetic investigation in other tissues, as fibroblasts and saliva, may demonstrate a pigmentary mosaicism that explains the phenotype. We report four patients with global developmental delay and hyperchromic and/or hypochromic cutaneous spots. Cytogenetic analysis was performed in fibroblasts obtained from spots and peripheral blood. All patients analyzed showed normal karyotyping results from peripheral blood, but presented mosaicism in skin fibroblasts, either in hyperchromic and/or hypochromic cutaneous spots. G-banding karyotype analysis revealed one patient with 12 trisomy mosaicism in hypochromic spots (46,XX[36]/47,XX,+12[14]) and in hyperchromic spots (46,XX[34]/47,XX,+12[16]); one with 8 trisomy (47,XY,+8[25]) in all spots; one with triploidy (69,XXX[20]) in all spots; one with Mosaic Variegated Aneuploidy Syndrome (46,XY[33]/random losses[9]/hyperdiploidies[3]/premature separation of chromatids[4]/47,XY,+7[1]); and one case of hermaphroditism in a male patient with (46,XX[15]) in all spots. Array in the DNA extracted from skin fibroblasts, peripheral blood and saliva may reveal if those patients present a genomic mosaicism besides the chromosomal mosaicism present in the cutaneous spots. Our results highlight that although mosaicism is a relatively frequent phenomenon in the population it is not usually detected in peripheral blood but in other tissues. The cytogenomic investigation in different tissues for patients with global developmental delay and cutaneous spots is essential since it has allowed conclusive diagnosis as well as genetic counseling for the families.

## PgmNr 2532: High usability of NGS diagnostic in Slovenian children with neurodevelopmental disorders.

### Authors:

D. Krgovic <sup>1,2</sup>; S. Stangler Herodez <sup>1,2</sup>; A. Golub <sup>1</sup>; N. Kokalj Vokac <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Medical Genetics, University Medical Centre Maribor, Maribor, Slovenia; 2) University of Maribor, Faculty of Medicine, Department for Molecular Biology, Maribor, Slovenia

---

### Introduction:

Since the implementation of molecular karyotyping in diagnostics of neurodevelopmental disorders (NDDs), the understanding of the genetic causes of these diseases has significantly improved. However, with the introduction of the Next Generation Sequencing (NGS), the setting of the genotype-phenotype correlations has become even more straightforward. By determining causative mutations in individual genes, it enabled us to understand the role of individual genes in the aetiology of the disease, and significantly increased the diagnostic yield for genetic diagnostic of NDDs in Slovenian children.

### Methods:

In this study, we present a small series of paediatric patients with various NDDs where molecular karyotyping as the first-tier diagnostic test in our laboratory did not show the presence of microdeletion or microduplication, which could explain the patients' clinical picture. By using the panel sequencing of 4813 genes (TruSight One Illumina Kit) and focusing on phenotype driven analysis, we were able to identify the genetic origin of the disorder (pathogenic Single Nucleotide Variant – pSNV) in 29% (N 13/45) of screened patients. In 24% (N 11/45) of patients, a SNV classified as variant of unknown significance (VOUS) has been determined within the disease-associated genes.

### Discussion:

Although a small series of Slovenian paediatric patients were analysed using target NGS sequencing, our study illustrates the high usability of this method in diagnostic of NDDs. With phenotype-genotype driven analysis, we were able to identify rare and clinically poorly defined syndromes in previously undiagnosed patients. Among them, disorders such as Kabuki and Noonan syndrome, Menkes disease, mutations in *IQSEC2* gene causing the Rett-like phenotype etc. were identified. A high percent of pSNV are also partly due to the selection of suitable patients for NGS testing based on their good clinical characterization. Our study confirmed a high efficiency of NGS in diagnostic of NDDs and enabled the diagnosis of rarely described cases which will be presented.

## PgmNr 2533: Genetic/epigenetic diagnosis of facioscapulohumeral muscular dystrophy (FSHD) via optical mapping.

### Authors:

Y.-W. Chen <sup>1,2</sup>; S. Bhattacharya <sup>1</sup>; M. Almalvez <sup>1</sup>; S. Sen Chandra <sup>1</sup>; A. Liu <sup>1</sup>; B. Dufault <sup>1</sup>; H. Sharim <sup>3</sup>; Y. Ebenstein <sup>3</sup>; E. Vilain <sup>1,2</sup>; H. Barseghyan <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Medicine Research, Children's Research Institute, Children's National Medical Center, Washington, DC 20010, USA; 2) Department of Genomics and Precision Medicine, School of Medicine and Health Science, George Washington University, Washington, DC 20052, USA; 3) School of Chemistry, Center for Nanoscience and Nanotechnology, Center for Light-Matter Interaction, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat Aviv 6997801, Israel

---

FSHD has a prevalence of 1:8,000-20,000 in the general population and is recognized as one of the most common forms of neuromuscular disorders. FSHD is primarily (95% of cases) caused by a deletion of the D4Z4 macrosatellite repeats (each unit 3.3 kb long) on chromosome 4p35, resulting in 1-10 D4Z4 repeats (healthy range 11-100). The D4Z4 region in healthy individuals has been shown to be transcriptionally repressed. Studies have shown the region becomes transcriptionally de-repressed and DNA demethylated when the number of D4Z4 repeats is shortened to 10 units or less. Current standard genetic testing for FSHD is to detect contracted D4Z4 array using Pulsed-Field Gel Electrophoresis (PFGE) in combination with Southern Blotting, which is time consuming, labor intensive and in some cases not precise enough for FSHD diagnosis. To address these diagnostic limitations, we used a novel optical genome mapping approach that combines both genetic and epigenetic variations in the telomeric region of chromosome 4q35 for FSHD diagnosis. Optical mapping captures a pattern of fluorescent labels in long DNA molecules (>150kb), in nanochannel arrays for *de novo* genome assembly and structural variant (SV) calling. This technology offers substantial advantages over the current clinical diagnostic practice as it has been successfully used for identification of SVs in Duchenne muscular dystrophy and more recently in FSHD cases. However, these tests were focused only on genotypic diagnosis of FSHD based on D4Z4 repeat array. Here, we incorporated dual high molecular weight DNA labeling for both assessment of D4Z4 repeat array as well as its methylome based epigenetic profiling. In the protocol, long DNA molecules are first nicked/labeled with BspQI/BssSI endonucleases and labeled with red fluorescent nucleotides. Second, the same DNA molecules undergo treatment with M.TaqI methyltransferase that attaches green fluorescent cofactor onto non-methylated CpGs in ATCG sequences throughout the genome. Imaging and analysis of the sample library revealed the methylation pattern as well as the number of D4Z4 repeat units. We successfully identified the molecular diagnosis in all tested cases (i.e. constriction of D4Z4 array) and show that dually labeled optical mapping for genetic/epigenetic markers in FSHD provides the underlying disease-causing mechanistic information that can be used to identify disease onset, severity and prognosis.

## PgmNr 2534: Toward comprehensive interpretation of *RYR1* variants associated with malignant hyperthermia susceptibility.

### Authors:

J.J. Johnston<sup>1</sup>; D. Ng<sup>1</sup>; S.G. Gonsalves<sup>1</sup>; B. Brandom<sup>2,3</sup>; R.T. Dirksen<sup>4</sup>; S. Riazi<sup>5</sup>; L.A. Saddic III<sup>6</sup>; N. Sambuughin<sup>7</sup>; R. Saxena<sup>8</sup>; J. Weber<sup>9</sup>; H. Rosenberg<sup>10</sup>; L.G. Biesecker<sup>1,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Departments of Anesthesiology and Nurse Anesthesia, University of Pittsburgh, Pittsburgh, PA; 3) North American MH Registry of MHAUS, Gainesville, FL; 4) Department of Pharmacology and Physiology, University of Rochester Medical School, Rochester, NY; 5) Department of Anesthesiology, University of Toronto, Toronto, Canada; 6) Department of Anesthesiology, University of California Los Angeles, Los Angeles, CA; 7) Consortium for Health and Military Performance, Uniformed Services University Health Science, Bethesda, MD; 8) Partners Healthcare, Boston, MA; 9) Prevention Genetics, Marshfield, WI; 10) MH Association of the United States, Sherburne, NY; 11) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

---

*RYR1* is one of the 59 genes recommended by the American College of Medical Genetics and Genomics (ACMG) for opportunistic screening, it is recommended that only known pathogenic (KP) variants be returned. Nearly all malignant hyperthermia (MH) associated variants are rare amino acid substitutions making interpretation challenging. To address this challenge, we set out to work toward a complete interpretation of known *RYR1* variants, which could be made available to the genetics community. We set out to do this through review by a ClinGen expert panel including medical geneticists, clinical molecular geneticists, physiologists, and anesthesiologists. Variants associated with malignant hyperthermia susceptibility (MHS), and other *RYR1*-associated dominant phenotypes were identified from the European Malignant Hyperthermia Group's (EMHG) list of "diagnostic mutations", the Human Gene Mutation Database (DM), GeneReviews, and ClinVar (likely pathogenic/pathogenic), (n=302). The ACMG-AMP criteria for variant interpretation were modified based on the biology and genetics of *RYR1* and MHS. Proposed *RYR1* specific rules were applied to 48 variants from the EMHG list of "diagnostic mutations". Resulting classifications were discussed by the *RYR1*-MHS expert panel to reach a consensus classification and refine the rules. Further revision of *RYR1*-specific rules included reducing PM2 to supporting and re-evaluating PS4, BA1, and BS1. The prevalence of MHS has been estimated to be 1 in 2-3,000 based on genetic studies. Using allele counts in the non-Finnish European (NFE) population in gnomAD for EMHG variants assessed as pathogenic (30 variants, total of 54 alleles) or likely pathogenic (15 variants, total of 22 alleles) a prevalence of MHS due to variants in *RYR1* in the NFE population was predicted to be 1 in 849 (maf=0.0007, 95% CI). As only EMHG variants were included, we considered this a minimum prevalence, we therefore set BA1 to be twice this value at 0.0014. The maximum reported allele contribution for any single allele is 21%, BS1 was set to 0.0003. Finalized assessments will be made available to the genetics community through ClinVar. Access to variants curated by this process will allow more efficient review of secondary variants and support improved interpretation of *RYR1*/MHS

genomic testing results and improve the ability to implement opportunistic screening for *RYR1*.



## **PgmNr 2535: High throughput analysis of tandem repeat contraction associated with facioscapulohumeral muscular dystrophy (FSHD) by optical mapping.**

### **Authors:**

J. Wang; E.T. Lam; A.W.C. Pang; T. Wang; H.B. Sadowski; A.R. Hastie; M. Borodkin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bionano Genomics, San Diego, California.

---

Facioscapulohumeral Muscular Dystrophy (FSHD) is one of the most common forms of muscular dystrophy. FSHD symptoms include progressive muscular degeneration, weakness, and atrophy. A genetic disease, FSHD could be inherited and could impact multiple members of a family. Non-inherited FSHD also occurs. Genetic testing is the most reliable way to confirm a diagnosis. Contraction of D4Z4, a tandem repeat region on chromosome 4q35, is diagnostic, when present on a permissive haplotype, for FSHD and can be assayed for by Southern blot.

We have developed a new workflow based on optical mapping on the Bionano Genomics Saphyr platform, which offers several advantages. Based on specific labeling and mapping of ultra-high molecular weight DNA in nanochannel arrays, optical mapping provides a higher resolution determination of D4Z4 and other tandem repeat expansions and contractions, even those which span more than 100 kilobase pairs. Analyses showed that molecules could span and be uniquely mapped to this region, and the repeat region could be accurately sized while also defining permissive and non-permissive alleles (4qA and 4qB). Because of the single-molecule nature of the platform, it is even possible to detect and quantify mosaicism in the repeat length. Also, the data could help resolve borderline or indeterminate Southern blot results. To better define outliers, we have developed a database of typical D4Z4 lengths based on control samples. The data can be imported into a graphical user interface tool for visualization and curation.

Bionano offers sample preparation, DNA imaging and genomic data analysis technologies combined into one streamlined workflow that enables high-throughput analysis of tandem repeat regions of interest. Together, these components allows for efficient analysis of diseases associated with repeat expansion and contraction.

## PgmNr 2537: Whole exome study of early onset Alzheimer's disease.

**Authors:**

J.H. Kim

[View Session](#) [Add to Schedule](#)**Affiliation:** Ilsan hospital, Koyang, KyungKi, Select a Country

---

**Background :** Alzheimer disease (AD) is the most common cause of dementia and heritability is about 70%. The development of genetic research techniques has revealed a number of related genes, but it still cannot explain all of the heritability of AD. Early onset AD (EOAD) accounts for about 5% of all AD patients. The heritability of EOAD is as much as 100%. Thus, genetic studies for EOAD can reveal important genetic factors for AD.

**Methods :** DNA was extracted from the blood of 17 EOAD or mild cognitive impairment with age < 65. Exome sequencing was performed in TheragenEtex Bio Institute. The data was analyzed by GATK version 3.8. We also analyzed copy number variations (CNV) using xhmm. The results were filtered out according to mutation frequencies reported in gnomAD. Gene functions were predicted by SNPEFF and ANNOVAR.

**Results :** Among genes related with AD, two nonsynonymous mutations in *ABCA7* (Glu316Lys and Pro1396Leu), one in *SORL1* (Ile2033Leu) and one in *TREM2* (Ser183Cys) were found. As far as genes related with frontotemporal lobar degeneration, mutations in *TBK1* (Arg724Cys) and *GRN* (Asn537Lys) were found. We checked CNV of genes related with cognitive impairments. Duplications or deletion of *ABCD1* gene causing adrenoleukodystrophy were found. Duplication was found in the *MECP2* gene causing angelman syndrome.

**Conclusions :** There were no *APP*, *PSEN1* or *PSEN2* mutations in WES in 17 EOAD patients. For *ABCA7*, *SORL1*, *TREM2*, *TBK1*, and *GRN* mutations found in patients, further functional studies are needed to reveal their association with disease. In addition, large-scale studies are required for the CNVs of the *ABCD1* and *MECP2* genes.

## PgmNr 2538: Revealing repeats: Toward a PCR-based test for autosomal dominant spinocerebellar ataxias 1, 2, 3, 6, & 7.

### Authors:

M. Deshotel <sup>1</sup>; M. Jama <sup>1</sup>; G. Pont-Kingdon <sup>1</sup>; P. Bayrak-Toydemir <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Genetics, ARUP Laboratories, Salt Lake City, UT; 2) Department of Pathology, University of Utah, Salt Lake City, UT

---

Spinocerebellar ataxias (SCAs) belong to a family of neurological diseases that lead to progressive cerebellar atrophy and eventual loss of peripheral motor functions. Of the roughly 40 hereditary SCAs described to date, the subtypes SCA1 [MIM 164400], SCA2 [MIM 183090], Machado-Joseph disease/SCA3 [MIM 109150], SCA6 [MIM 183086], and SCA7 [MIM 164500] are the most common autosomal dominant ataxias worldwide and are caused by pathogenic expansion of exonic CAG repeats in SCA-associated genes. Symptoms of these five subtypes are clinically heterogeneous, and de novo mutations are known to arise, so molecular genetic testing is required for differentiation. Here, we describe the development of a test for the quantification of CAG repeats in the SCA-associated genes *ATAXIN1* (SCA1), *ATAXIN2* (SCA2), *ATAXIN3* (SCA3), *CACNA1A* (SCA6), and *ATAXIN7* (SCA7). Quantification of CAG repeats was performed using a chimeric triplet repeat-primed PCR method, which was originally developed at ARUP Laboratories for detection of repeat expansions in Fragile X, Huntington disease, and myotonic dystrophy-1. Forward primers were designed with HEX, NED, and FAM fluorescent dyes to facilitate multiplexing, and reverse primers were designed with a CAG<sub>n</sub> repeat region and flanking region. The amplicons were separated by capillary electrophoresis, and the characteristic ladder of peaks counted to determine the number of CAG repeats in each allele. Repeats in all five SCA-associated genes can be quantified in just two multiplex PCR reactions, an improvement over previously published methods. Our method also allows detection of CAT interruptions in *ATAXIN1* and CAA interruptions in *ATAXIN2*, both of which have clinical utility for patients. We have detected CAG expansions with as many as 60 (SCA1), 44 (SCA2), 72 (SCA3), 22 (SCA6), and 62 repeats (SCA7) in patient samples, but our method is expected to reveal even larger pathogenic expansions. To verify correct repeat allele sizing, deidentified normal patient samples with homozygous alleles were confirmed by Sanger sequencing. Nine samples from the Coriell Institute and three CAP proficiency testing samples were also tested to verify correct repeat allele sizing. These results indicate that the triplet repeat-primed PCR technique is a robust and flexible method that allows for accurate quantification of CAG repeats and CAA/CAT interruptions in SCA-associated genes.

## **PgmNr 2539: Pulse-field capillary electrophoresis of repeat-primed PCR amplicons for analysis of large repeat in spinocerebellar ataxia type 10.**

### **Authors:**

V. Hashem; A. Tiwari; B. Bewick; T. Ashizawa

[View Session](#) [Add to Schedule](#)

**Affiliation:** Neurology, Houston methodist research institute, Houston, Texas.

---

Large expansions of microsatellite DNA cause several neurological diseases. In Spinocerebellar ataxia type 10 (SCA10), the repeat interruptions change disease phenotype; an (ATTCC)<sub>n</sub> or a (ATCCT)<sub>n</sub>/(ATCCC)<sub>n</sub> interruption within the (ATTCT)<sub>n</sub> repeat is associated with the robust phenotype of ataxia and epilepsy while pure (ATTCT)<sub>n</sub> may have reduced penetrance. Large repeat expansions of SCA10, and many other microsatellite expansions, can exceed 10,000 base pairs in size. Conventional next generation sequencing (NGS) technologies are ineffective in determining internal sequence contents or size of these expanded repeats. Using repeat primed PCR (RP-PCR) in conjunction with a high-sensitivity pulsed-field capillary electrophoresis fragment analyzer (FEMTO-Pulse, Agilent, Santa Clara, CA), we successfully determined sequence content of large expansion repeats in genomic DNA of SCA10 patients and transformed yeast artificial chromosomes containing SCA10 repeats. This RP-FEMTO is a simple and economical methodology which could complement emerging NGS for very long sequence reads such as SMRT sequencing (PACBio) and Nanopore sequencing technologies.

## PgmNr 2540: Molecular testing of ataxia-related conditions: Experience from the University of Chicago Genetic Services Laboratories.

### Authors:

M. Sanyoura <sup>1</sup>; S. Andani <sup>1</sup>; H. Subramanian <sup>1</sup>; M. Helgeson <sup>1</sup>; S. Tawde <sup>1</sup>; K. Arndt <sup>1</sup>; P. Sodadasi <sup>1</sup>; B. Deml <sup>1</sup>; D. del Gaudio <sup>1</sup>; D. J. Waggoner <sup>1</sup>; C. M. Gomez <sup>2</sup>; S. Das <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Univ Chicago, Chicago, Illinois.; 2) Department of Neurology, University of Chicago, Chicago, Illinois

Ataxias represent a clinically and genetically complex group of neurological disorders and may be hereditary or sporadic. The ataxia symptom can be isolated or part of a multisystemic neurological disorder, presenting at ages ranging from infancy to adulthood, and can be dominant, recessive, X-linked or mitochondrial. Genetic testing for hereditary ataxias remains challenging mostly due to the significant clinical overlap and genetic heterogeneity with more than 500 genes reported to be associated with ataxia or ataxia-like presentation. Targeted exome sequencing is regarded as a powerful diagnostic tool for heterogeneous neurological disorders such as ataxia, but limited in its ability to detect repeat expansions which are estimated to account for over 50% of dominant hereditary ataxias. In 2014, we developed a targeted exome panel consisting of a predefined and regularly updated set of genes known to be associated with ataxia and ataxia-related conditions. In 2017, we implemented repeat expansion testing to allow for more comprehensive testing and enhancement of our diagnostic yield. Here, we present the results of 411 patients with ataxia and ataxia-like conditions, referred for genetic testing between April 2014 and May 2019. Overall, 53% of patients were female, 47% were male, and patient age ranged from 1 to 88 years. Of the 411, 68% (279) were referred for exome sequencing, 25% (102) for repeat expansion testing, and 7% (30) for both. Clearly pathogenic and suspected diagnostic variants were identified in 167 patients (41%), an impressive yield given the heterogeneity of our cohort. A total of 80 different genes were implicated in our cohort. 14 of the 167 patients had pathogenic repeat expansions in either spinocerebellar ataxia genes or the *FXN* gene. The 7 most common genes with abnormalities detected were *CACNA1A* (14), *SPG7* (13), *SYNE1* (10), *ADCK3* (6), *ATXN2* (6), and *ATP1A3* (5). Together, these genes accounted for 38% of positive cases.

With new gene discoveries and consequent updates to our exome panel, re-analysis of negative cases can be performed and was instrumental in the identification of pathogenic variants in *EBF3* in three patients with early onset ataxia. The ongoing discovery of novel gene-disease associations are expected to increase the diagnostic yield of targeted clinical exomes. Our lab is in the process of updating our bioinformatic pipeline to implement an automated reanalysis process for negative ataxia exomes.

## **PgmNr 2541: Phenotype influences the likelihood of a diagnostic result in a sequencing test of 2,300+ genes associated with autism spectrum disorders and/or intellectual disability.**

### **Authors:**

T. Brandt; A. Shanmugham; J. Scuffins; J. Juusola; D. McKnight

[View Session](#) [Add to Schedule](#)

**Affiliation:** Gene Dx, Gaithersburg, Maryland.

---

Autism spectrum disorders (ASD) and intellectual disability (ID) are common and frequently co-morbid disorders with a high degree of genetic heterogeneity. Exome sequencing is reported to have a positive diagnostic result (PDR) of ~14-33% for ASD/ID cohorts. However, studies demonstrating the utility of genetic testing for subgroups of patients within this wide phenotypic spectrum are rare. The objective of this study was to determine the utility of genetic testing for patients with ASD and/or ID. This is a retrospective study of genetic testing results of 2,497 patients with ASD/ID/developmental delay referred to our clinical laboratory. For each of these patients, 2,300+ genes associated with ASD and/or ID were sequenced and analyzed. The PDR for all cases was ~16%; however, the PDR varied depending on the clinical information provided. Patients with ID and multiple additional malformations or other significant medical issues (e.g. dysmorphic features, seizures, hypotonia), but not autistic features, had the highest PDR (~36%, 46/128), whereas patients with autistic features and multiple additional medical issues in the absence of ID had a significantly lower PDR (~15%, 57/386,  $p < 0.0001$ ). Patients with isolated ID or ASD had a PDR of ~23% (10/44) and ~4% (19/492), respectively. The PDR for patients <2 years old with developmental delays was ~27% (27/99). High genetic heterogeneity was observed among positive cases; however, 22 different genes were responsible for >40% of positive cases. In summary, genetic testing for patients with ID and additional major medical issues revealed a genetic etiology of disease in over one-third of cases. A genetic etiology was less likely to be identified in patients with ASD. The presence of additional phenotypic findings did correlate with an increased PDR. Therefore, this study supports the hypothesis that monogenic etiologies are relatively uncommon findings in patients with isolated ASD. In addition, although there is vast genetic heterogeneity for ID and ASD, findings in just over 20 genes accounted for a significant portion of positive cases, suggesting a minimum number of genes that should be tested for these patients.

## PgmNr 2542: Ten years of experience in molecular genetic diagnostics of Rett syndrome in Bulgaria.

### Authors:

T. Todorov <sup>1</sup>; P. Dimova <sup>3</sup>; V. Bojinova <sup>4</sup>; R. Tincheva <sup>5</sup>; A. Todorova <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human genetics, GGenetic Medico-Diagnostic Laboratory Genica, Sofia, Bulgaria MDL Genica, Sofia, Bulgaria; 2) Department of Medical Chemistry and Biochemistry, Sofia Medical University, Sofia, Bulgaria; 3) Department of Neurosurgery, University Hospital "Saint Ivan Rilski", Sofia, Bulgaria; 4) Neurological Hospital "St. Naum", Sofia, Bulgaria; 5) Department of Clinical Genetics, University Pediatric Hospital Sofia, Sofia, Bulgaria

---

Rett syndrome (RTT) is caused by pathogenic variants in the methyl-CpG-binding protein 2 (*MECP2*) gene. The *MECP2* gene has some unique characteristics: 1) it is mainly affected by de novo pathogenic variants, due to recurrent independent mutational events in a defined "hot spot" regions or positions; 2) complex mutational events along a single allele are frequently found in this gene; 3) most pathogenic variants arise on paternal X chromosome. The recurrent point mutations involve mainly CpG dinucleotides, where C>T transitions are explained by methylation-mediated deamination. The complex mutational events might be explained by the genomic architecture of the region involving the *MECP2* gene. The finding that most spontaneous mutations arise on paternal X-chromosome supports the higher contribution of replication-mediated mechanism of mutagenesis.

Here we present our ten years of experience in the molecular genetic diagnostics of RTT. Altogether 131 unrelated families with clinical suspicion of RTT were screened for mutations in the *MECP2* gene by Sanger sequencing, MLPA or NGS. The diagnosis was confirmed in 55 families (42%). The detected pathogenic variants were distributed in the following categories: 46 point mutations, small deletions, small duplications; 1 indel and 3 large deletions within *MECP2* gene; 1 large duplication covering *MECP2* gene; 1 Xp deletion involving the genes *ARX* and *CDKL5*. From the *MECP2* negative cohort, 2 patients turned out to carry pathogenic variants in the *CDKL5* gene and 1 patient was positive for pathogenic variant in the *FOXG1* gene. A male patient with RTT-like features was screened by whole exome sequencing and pathogenic variant in the *MECP2* gene was found; inherited from asymptomatic mother. A large pedigree with more than 5 affected males (severe mentally retarded) turned out to be positive for large duplication, covering the genes *SLC6A8*, *L1CAM*, *MECP2*, *TKTL1*, *FLNA*, *GDI1*. Although the clinical diagnosis of RTT is rather uniform, the genetic verification of RTT and RTT-like cases (both female and male) requires different diagnostic approaches due to the large spectrum of pathogenic variants associated with the disease.

## PgmNr 2543: Design and validation of a single customized next-generation sequencing capture library for neurodevelopmental, cardiac, immune, and inflammatory bowel diseases.

### Authors:

Y. Bayram<sup>1,2</sup>; J. Barrows<sup>1,2</sup>; W. Khan<sup>1,2</sup>; R. Fox<sup>1</sup>; L. Shi<sup>1,2</sup>; G. Cai<sup>1,2</sup>; X. Cai<sup>1,2</sup>; K. Manheimer<sup>1</sup>; G. Mendiratta-Vij<sup>1,2</sup>; W. Jiang<sup>1,2</sup>; Y. Yang<sup>1,2</sup>; Y. Sha<sup>1,2</sup>; M. Arif<sup>1,2</sup>; H. Wan<sup>1,2</sup>; E. Spencer<sup>2</sup>; N. Cody<sup>1,2</sup>; X. Lu<sup>1,2</sup>; J. Kim<sup>1,2</sup>; D. Hibbs<sup>1,2</sup>; M. Croken<sup>1,2</sup>; J. Holt<sup>1</sup>; A. McLellan<sup>1,2</sup>; S. Van den Berg<sup>1,2</sup>; E. Schadt<sup>1,2</sup>; L. Edelman<sup>1,2</sup>; F. Suer<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sema4, a Mount Sinai venture, Stamford, CT, 06902; 2) Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, 10029

**Background:** Cost-effectiveness of next-generation sequencing (NGS) libraries has increased as sequencing costs continue to decline. Neurodevelopmental, cardiac and immunodeficiency diseases share overlapping genes and phenotypes that facilitate consolidation into a single sequencing library. Locus heterogeneity and phenotypic overlap make NGS ideal for diagnostic testing of such disorders. Targeted NGS captures relevant coding, intergenic, intronic and problematic regions to increase diagnostic yield. Genomic diagnosis enables personalized care.

**Methods/Results:** Primary literature, public database, expert opinion and commercially-available panel review provided potential genes. These genes and HGMD known/potential disease-causing mutations/phenotypes were split into subgroups based on gene/mutation/HPO relationships. Genes, source lists and mutations were then consolidated into preliminary lists with total and filtered P/LP variants, variant type and detectability by NGS, and CNV detection availability. Genes were computationally scored for actionability by tallying filtered variants and commercial availability. These data and published rules for clinical utility assessment guided laboratory directors and fellows in manually scoring the preliminary gene list; 885 had sufficient clinical actionability.

Agilent SureSelectQXT libraries were sequenced on Illumina NovaSeq 6000 instruments. Saliva, peripheral blood, prenatal, and Coriell cell line samples (576) were assessed for performance. Reproducibility, intra and inter-run genotyping concordance were >99%. All known P/LP small insertion/deletions and single nucleotide variants (SNVs) in 115 samples were successfully detected. The estimated analytical test sensitivity was 99% for SNVs and 95% for small insertions/deletions; the overall detection rate was >97%. CNV detection using an eXome Hidden Markov Model (XHMM) with a baseline of 346 negative samples was validated with 174 positive (variants ≥ two probes) samples. The sensitivity of XHMM was >94%, specificity was >99%, and accuracy was >99%.

**Conclusion:** Test orders change frequently during the diagnostic odyssey. A single panel library allows pivoting on changed orders. A larger library broadens the sensitivity, specificity, and accuracy. As the number of genes increases in large diagnostic panels, addressing variants of uncertain clinical significance remains an important area of development that will benefit from public efforts such as ClinVar/ClinGen.



## PgmNr 2544: When less is more: A novel whole gene deletion of *DHX30* confers an apparent smaller-effect relative to typical missense, which results in NEDMIAL-associated phenotypes.

### Authors:

J. Murry<sup>1</sup>; I. Mannucci<sup>2</sup>; H.-J. Kreienkamp<sup>2</sup>; D. Lessel<sup>2</sup>; P.B. Shieh<sup>3</sup>; T. Althoff<sup>4</sup>; J. Abramson<sup>4</sup>; A. Slavotinek<sup>5</sup>; J.A. Martinez-Agosto<sup>6,7</sup>; S.F. Nelson<sup>1,3,6,7</sup>; F. Quintero-Rivera<sup>1,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Pathology and Lab. Med., UCLA Center for Health Sciences, David Geffen School of Medicine, University of California Los Angeles; 2) Institute for Human Genetics. University Medical Center Hamburg-Eppendorf. Hamburg, Germany; 3) Dept. of Neurology, David Geffen School of Medicine, University of California Los Angeles; 4) Dept. of Physiology, David Geffen School of Medicine, University of California Los Angeles; 5) Dept. of Pediatrics, University of California San Francisco; 6) Dept. of Human Genetics, University of California Los Angeles; 7) UCLA Clinical Genomics Center, David Geffen School of Medicine, University of California Los Angeles

The RNA helicase *DHX30*, of the DExH-box protein family, is predicted to mediate ATP dependent unwinding of RNA secondary structure and is highly expressed during brain neurogenesis. Five unique *de novo* missense *DHX30* variants have been described in patients with Neurodevelopmental Disorder with Severe Motor Impairment and Absent Language (NEDMIAL syndrome, MIM: 617804), conferring impaired ATPase activity or RNA binding, implicating the helicase domain of its encoded protein, suggesting that accumulated mutant protein leads to a global decrease in protein synthesis (Lessel et al 2017). However, the consequences of *DHX30* deletion in humans are unknown. Chromosomal microarray analysis detected a *de novo* 1.1 Mb loss of 3p21.31 in a 3-year-old with hypotonia, developmental delay (DD), and speech delay with motor impairment, experiencing difficulty climbing stairs, and falling frequently. Brain MRI detected Chiari malformation and hypomyelination. Of the deleted genes, *DHX30* was associated with NEDMIAL, whose phenotype partially overlapped with our patient's, and was predicted to be dosage-sensitive. This is the first report of full gene deletion and a milder presentation compared to others with missense variants in *DHX30*.

In parallel, clinical exome sequencing found novel missense *DHX30* variants in two patients with intellectual disability, DD and NEDMIAL spectrum. Both, a *de novo* p.R782Q and p.T739A localize to the helicase domain. 3D-Protein modeling for both supports helicase localization, predicting ATPase deficiency. Functional analysis of the missense mutations will be presented. Helicase-inactivating missense variants may exert a spectrum of dominant negative phenotypes and could suggest that over-expression of the mutant helicase protein potentially interferes with normal RNA metabolism, and may be one way in which dominant-negative missense variants may lead to a greater defect compared to that of a heterozygous deletion allele with  $\leq 50\%$  protein activity. We report 1) the first complete null *DHX30* allele conferring a milder phenotype than the severity associated with missense SNVs in patients who are non-verbal have no independent walking, and present with delayed myelination with cerebellar or cerebral atrophy, and dilated ventricles; 2) Two novel missense

variants associated with NEDMIAL. All three patients add to the limited phenotypes of patients with NEDMIAL and spectrum, guiding clinical interpretation for these and other novel *DHX30* variants.

## PgmNr 2545: Age-related molecular diagnosis of 539 patients with hereditary spastic paraplegia using next generation sequencing gene panel.

### Authors:

H.Z. Elloumi; R. Brandon; T. Brandt

[View Session](#) [Add to Schedule](#)

**Affiliation:** GeneDx, Gaithersburg, MD

Hereditary spastic paraplegia (HSP) comprises a diverse group of neurodegenerative motor neuron disorders characterized by progressive lower limb spasticity and weakness. Age of onset can range from early childhood to late adulthood. The genetic basis of HSPs is heterogeneous.

We performed a retrospective study to evaluate the diagnostic yield of HSP genetic testing. Data represent an unselected cohort of 539 individuals sent to our clinical diagnostic laboratory for HSP genetic testing using a next-generation sequencing plus deletion/duplication analysis panel including up to 42 genes. The average age at testing was 39 y.o. (range 1-85).

We obtained a positive diagnostic rate (PDR) of ~24% (128/539). The PDR based on age at testing is as follows: 22% (38/172) for children (1-17 y.o.), 28% (40/142) for younger adults (18-50 y.o.), and 22% (50/225) for older adults (51-85 y.o.). Of positive findings, 73% were identified in genes associated with autosomal dominant disorders and 45% were found in the *SPAST* gene. Autosomal recessive (AR) disorders accounted for 24% of positive cases and X-linked disorders represented only 3%. It is interesting to note that *SPG7* mutations were the most frequent etiology of AR-HSP in our cohort (13/31), followed by *SPG11* (9/31). The most common molecular diagnoses in children were in the following genes (from most to least prevalent): *SPAST*, *ATL1*, and *KIF1A*. For younger adult patients, common causative genes were *SPAST*, *SPG11*, *SPG7*, and *KIF5A*. The most likely causative genes were *SPAST*, *SPG7*, *REEP1*, and *BSCL2* for adults age 51+. Other interesting age-related trends were observed. For instance, this panel only identified *ATL1*-related diagnoses in children (n=11), *SPG11* or *CYP7B1*-related diagnoses in younger adults (n=9 and 3, respectively), *BSCL2* positive cases in older adults (n=3), and *SPG7*-related diagnoses in adults (n=13). Diagnostic pathogenic/likely pathogenic variants in *SPAST*, *REEP1*, *KIF1A* and *KIF5A* were identified in all age groups.

Amongst diagnostic variants, missense variants were the most frequent (~43%), followed by Indels (~20%), copy number variations (CNVs) (~15%, with the vast majority involving *SPAST*), nonsense variants (~12%), and splice site variants (~9%).

In summary, this data showed the utility of genetic testing for HSP, a complex disorder with a challenging diagnosis, and how positive findings vary with patient age.

## PgmNr 2546: *SMN1* and *SMN2* copy number quantification using a novel digital PCR platform.

### Authors:

L. Jiang; M. Dueck; R. Lin; S. Gallagher; A. Zayac; P. Hung

[View Session](#) [Add to Schedule](#)

**Affiliation:** COMBiNATi, Burlingame, California.

---

Spinal muscular atrophy (SMA), a genetic neuromuscular disorder and the leading cause of infant mortality, is largely caused by the loss of the survival motor neuron gene 1 (*SMN1*). Survival motor neuron gene 2 (*SMN2*) has been found to produce partial function and can compensate for *SMN1* deletion. Accurate quantification of *SMN1* and *SMN2* copy numbers provides critical diagnostic and prognostic values for the disease. SMA has become an actionable disease with the advancement of treatments. Newborn screening of SMA is increasingly practiced in many states and other countries. Current screening methods are complex and time consuming. Here we present two duplex PCR assays capable of absolute quantification of *SMN1* and *SMN2* copy numbers on a novel digital PCR platform to meet the growing demand for SMA Newborn screening.

The duplex assays consist of *SMN1* or *SMN2* hydrolysis probe assay with a commercially available Rnase P TaqMan copy number reference assay. We simplified the assay workflow by eliminating the restriction digestion recommended for copy number assays. We also evaluated the assays using a DNA concentration (~2ng/ul) resembling newborn dried blood spots. Additionally, accuracy and repeatability studies were conducted on 10 human genomic DNA samples with characterized *SMN1/SMN2* copy numbers from the Coriell Institute Cell Repository. All the tests were conducted on a novel digital PCR platform with one-step walkaway workflow. Data were analyzed using platform-specific software that identifies positive dPCR partitions, applies Poisson Statistics for absolute quantification of copy numbers and can also display individual amplification curves to provide confidence in assay performance.

Both assays demonstrated accurate and consistent *SMN1/SMN2* quantification of all samples with a simple workflow and low DNA concentration. For *SMN1*, the assay correctly assessed the copy number of the 10 samples - zero (n=5), one (n=3), two copies (n=1) or three copies (n=1). For *SMN2*, the assay also exhibited robust performance for the 10 samples containing 2 zero copy, 1 one copy, 3 two copies, 3 three copies, and 1 five copies. *SMN1/SMN2* results were 100% consistent across 3 repeat experiments on the instrument. Additionally, the data were obtained in under 90 minutes.

We demonstrated two duplex SMA newborn screening assays that accurately and reliably quantify *SMN1* and *SMN2* copy numbers on a novel *digital PCR* within 90 minutes, *with a single-step and walkaway workflow*.

## PgmNr 2547: Full genome analysis for identification of single nucleotide and structural variants in genes that cause developmental delay.

### Authors:

H. Kao <sup>1</sup>; N. Lee <sup>2,3</sup>; P. Hwu <sup>2,3</sup>; P. Kwok <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) 1. Institute of Biomedical Sciences, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, No. 8, Chung-Shan S. Rd., Zhongzheng Dist., Taipei 10041, Taiwan; 3) Department of Pediatrics, National Taiwan University Hospital, No. 8, Chung-Shan S. Rd., Zhongzheng Dist., Taipei 10041, Taiwan; 4) Cardiovascular Research Institute, Institute for Human Genetics, and Department of Dermatology, University of California, San Francisco, San Francisco, 94143, CA, USA

---

Developmental delay caused by gene mutations are hard to identify due to phenotypic and genetic heterogeneity. While whole exome sequencing (WES) has been used to identify causal mutations, the majority of cases remain unsolved. In this study, we test the hypothesis that full genome analysis (phased whole genome sequencing with assembly based on linked-read sequencing and mapping technologies) of trios provides the best chance of identifying gene mutations, and especially structural variations, that cause developmental delay.

Eight trios with negative finding by WES were subjected to (1) whole genome sequenced (WGS) to 60X coverage on the 10x Genomics linked-read platform, and (2) mapped to 90X coverage on the Bionano Genomics optical mapping platform. The de novo, phased genome assemblies of the patient and his/her two biological parents were compared and the single nucleotide variants (SNVs) and structural variants (SVs), including de novo variants, were identified. After filtering out synonymous SNVs, and common SNVs/SVs against public databases, the remaining variants were further evaluated according to the three possible genetic models: autosomal recessive, autosomal dominant and X-linked recessive. In the autosomal dominant and X-linked recessive models, de novo variants were selected for careful analysis. In each instance, we examined genes affected by SNVs, SVs, or a combination of SNVs and SVs.

Candidate variants (both SNVs and SVs), predicted as deleterious by multiple tools in the web-based WANNONAR suite, were selected for genotype-phenotype correlation in consultation with the clinicians and a literature search of the candidate genes were performed. Three WES-negative cases that were successfully diagnosed by this full genome analysis approach will be presented.

## PgmNr 2548: Copy number variations contribute to neurodevelopmental disorders.

### Authors:

A.R. Patel <sup>1,2</sup>; S.S. Chettiar <sup>2</sup>; D.D. Jhala <sup>3</sup>; R. Uppala <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetic Laboratory, Department of Anatomy, B.J. Medical College, Civil Hospital Campus, Ahmedabad, Gujarat, India; 2) Department of Genetics, GeneXplore Diagnostics and Research Center Pvt. Ltd., Ahmedabad, Gujarat, India.; 3) Human Genetic Centre, Department of Zoology, Gujarat University, Navarangpura, Ahmedabad, Gujarat, India.; 4) Department of Obstetrics and Gynecology, Oakland University- William Beaumont School of Medicine, Oak, MI, USA.

---

Copy Number Variants (CNVs) detection of by Chromosomal Microarray Analysis (CMA) testing is a powerful diagnostic tool is now well accepted as genetic diagnostic tests for various disorder including such as Developmental delay (DD), Intellectual disabilities(ID) and Autism Spectrum Disorders(ASD). CMA offers high resolution scan of whole genome which, significantly contribute to understand the etiology of these complex neurodevelopment impairments. Considering the detection efficiency of CMA and its impact on medical management, CMA is first-choice of test for in-depth analysis of children with these disorders.

In the present study, we have investigated 130 children (83 Males and 47 Females; Age ranges from: 2.0 yrs.-17yrs.) diagnosed with Developmental delay, Autism spectrum disorder or intellectual disabilities were recruited for the genetic analysis. The microarray platform utilized in this study was specifically designed to increase detection of CNVs in genomic regions of demonstrated relevance to DD/ID/ASD. Array CGH was performed on oligonucleotide microarray platform (G2565CA, Microarray Scanner System, Agilent (USA). Array data was analyzed using cytogenomics software (CytoGenomics v4.0.3, Agilent technologies).

CMA results revealed a total of 55 CNVs in 46 patients (35.38%). Among these, 32 (69.56%) signified as known diagnostic CNVs contributing to the clinical representation, 23 were non-diagnostic CNVs (variants of unknown significance (VOUS)). Fourteen patients (30%) with developmental delay showed 15 CNVs, 18 patients (40%) with intellectual disabilities showed 24 CNVs and 14 patients (30%) with Autism Spectrum Disorders (ASD) showed 16 CNVs. Imbalances of chromosome 2,15 and 22 were found to be most common. CNVs were detected more in males than females. The likelihood of an abnormal microarray results increased with the number of clinical abnormalities.

CMA has demonstrated great value in the clinical assessment of neurodevelopmental disorders. The optimized CMA platform can increase the diagnostic yield. Pathogenic findings give insights into the etiology of patients' neurodevelopmental conditions, and in many cases positively impact medical management decisions. The development of novel and accurate methods to interpret the potential pathogenicity of VOUS will further enable patients and their physicians to realize the maximum benefits of genetic testing for clinical care.

## PgmNr 2549: Utility of gene panel testing in children with seizure onset after 2 years of age: Results from a European and Middle Eastern epilepsy genetic testing program.

### Authors:

T.-P. Alastalo <sup>1</sup>; K. Gall <sup>1</sup>; E. Izzo <sup>2</sup>; K. Alakurtti <sup>1</sup>; E.H. Seppala <sup>1</sup>; L. Koskinen <sup>1</sup>; J. Koskenvuo <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Blueprint Genetics, San Francisco, CA; 2) BioMarin Pharmaceutical Inc., Novato, CA

---

**Background:** Epilepsy is one of the most common childhood-onset neurological conditions with a genetic basis. Genetic diagnosis provides potential for etiologically-based management and treatment. Existing research has focused on early-onset (<2 years) epilepsies while data regarding later-onset epilepsies is limited. Program goals: Determine, in a selected pediatric epilepsy cohort, the overall and actionable molecular diagnostic (MDx) yield and the CLN2 disease MDx yield. CLN2 is a severe, rapidly progressive neurodegenerative disease with onset of seizures at/after 2 years and average age-of-diagnosis of 5 years.

**Methods:** Blueprint Genetics' next-generation sequencing (NGS)-based 283-gene epilepsy panel was used. Copy number variant (CNV) detection from NGS data was included. Variant interpretation was performed according to ACMG guidelines. Program results (Oct/2017-Nov/2018) are reported from 210 patients (Europe, Middle East) with inclusion criteria: Age 24-60 months, first seizure at/after 24 months, and at least one additional finding. The program was sponsored by BioMarin Pharmaceutical Inc.

**Results:** Median age-at-testing: 42 months; median age-of-first-seizure-onset: 30 months; average delay from first seizure to comprehensive genetic testing: 10.3 months. Genetic diagnosis was established in 42 patients; 20.0% MDx yield. CNVs were reported in 26.2% of diagnosed patients; 27.3% of CNVs identified were intragenic. MDx included 5 CLN2 (*TPP1* gene) diagnoses, 4 MECP2, 3 SCN1A, 3 Angelman syndrome, 2 each of CHD2, KCNA2, MFSD8, SCN2A and STXBP1.

**Conclusion:** This program demonstrates the clinical utility of a comprehensive epilepsy gene panel for patients with first seizures at/after 2 years for MDx of pediatric epilepsy and CLN2 disease to guide management and treatment.



## PgmNr 2550: Exome sequencing in neurogenetics solves challenging cases.

### Authors:

Q. Thomas<sup>1,2,3</sup>; A. Vitobello<sup>1,4</sup>; F. Tran Mau-Them<sup>1,4</sup>; A. Fromont<sup>5</sup>; M. Giroud<sup>3</sup>; B. Daubail<sup>3</sup>; A. Jacquin-Piques<sup>3</sup>; M. Hervieu-Begue<sup>3</sup>; T. Moreau<sup>5</sup>; G.-V. Osseby<sup>3</sup>; S. Nambot<sup>1,2</sup>; A.-L. Bruel<sup>1,4</sup>; A. Sorlin<sup>1,2,4</sup>; P. Callier<sup>1,2,4</sup>; A.-S. Denomme-Pichon<sup>1,4</sup>; L. Faivre<sup>1,2</sup>; Y. Béjot<sup>3,6</sup>; C. Philippe<sup>1,4</sup>; C. Thauvin-Robinet<sup>1,2,4</sup>; S. Moutton<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Inserm UMR1231 team GAD, University of Burgundy and Franche-Comté, F-21000 Dijon, France; 2) Genetics Center, FHU-TRANSLAD, Dijon Bourgogne University Hospital, F-21000 Dijon, France; 3) Neurology, Dijon Burgundy University Hospital, Dijon, Burgundy, France; 4) Functional Unity of innovative diagnosis for rare disease, Dijon Bourgogne University Hospital, F-21000 Dijon, France; 5) Service de Pathologies Inflammatoires du système nerveux central, Neurologie Général, Dijon University Hospital; 6) Dijon Stroke Registry, EA7460, Pathophysiology and Epidemiology of Cerebro-Cardiovascular Diseases (PEC2), University Hospital of Dijon, University of Burgundy, Dijon, France

Most neurogenetic pathologies share two major characteristics: first, within a given framework their clinical presentations are highly uniform and non-specific e.g. gait disorders and limb incoordination for cerebellar ataxias. Second, their molecular bases are quite vast with large genetic and multiple modes of inheritance, making them difficult for physicians to choose between countless syndromes and genes.

After years of clinical classifications, the last decades have been marked by the clinical implementation of massive parallel sequencing technologies including exome sequencing (ES). Its implementation in neurogenetics led to the discovery of numerous new genes, phenotypes expansion and markedly increased diagnostic yield.

We performed ES as a first- or second-tier test in a cohort of 52 patients with neurodegenerative diseases: 13 cerebellar ataxia, 12 neuromuscular disorders, 10 spinocerebellar ataxia, 8 spastic paraplegia, 2 movement disorders and 7 labelled “others” in regard of complex clinical features. This study revealed a 42% overall diagnostic yield with 6 interesting atypical findings.

In a 13 year-old girl, with suspicion of mitochondrial cytopathy for years because of isolated cerebellar ataxia subsequently associated with psychomotor regression, cerebellar atrophy, hearing impairment, extensive leukoencephalopathy and gradient-echo sequences hyposignal of the basal ganglia, ES identified a homozygous *CRAT* missense variant inherited from related parents in favor of Neurodegeneration with Brain Iron Accumulation.

In an 11 year-old girl referred for the diagnosis of a cerebellar ataxia with basal ganglia and midbrain hypersignal, ES with a dedicated pipeline analysis allowed to identify a mitochondrial DNA missense variant in the *MT-ATP6* gene.

In a family referred for familial gait disorder (spastic paresis of lower limbs in a 46 year-old man and hypotonic cerebellar ataxia in his daughter), ES identified a causal *BSCL2* missense variant in the father while his daughter, who inherited this variant, also carried an additional missense variant in the *MT-ATP6* gene leading to a mitochondrial defect explaining her current symptoms.



ES is a powerful tool for diagnosing neurodegenerative diseases, allowing not to be restricted to a given panel and thus to identify molecular bases of very interesting findings such as differential diagnosis, expansion of clinical spectrum, multiple diagnosis within a given family and even a given patient.

## PgmNr 2551: Detection of spinal muscular atrophy genotypes in a highly multiplexed test.

### Authors:

L. Wu <sup>1</sup>; J. Winters <sup>2</sup>; L. Le <sup>1</sup>; L. Hasadri <sup>2</sup>; A. McDonald <sup>1</sup>; R. Varma <sup>1</sup>; A. Roter <sup>1</sup>; J. Schmidt <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Thermo Fisher Scientific, Santa Clara, California.; 2) 2Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905

---

Spinal muscular atrophy is a debilitating disease causing lack of spinal muscle control and infant mortality in the most extreme cases. Normal spinal nerve-muscle connections require the presence of protein produced by the SMN1 gene (Spinal Muscle Neuron 1 Gene). Symptoms can be reduced in severity by the presence of extra copies of the SMN2 gene, which produces a protein with reduced activity. The SMN1 and SMN2 genes are nearly identical in sequence (99.6% identity). Accurate detection of the number of SMN1 and SMN2 genes in a sample is critical to the correct characterization of the SMA genotype. Approximately one in 6,000 babies are born with the disease and one in 40 individuals are carriers of a defective SMA1 gene.

For analytical verification, 386 DNA samples were characterized using two methods: the Applied Biosystems CarrierScan® Assay and droplet digital PCR (ddPCR). In this data set, 7 samples were identified as having SMA carrier status. 19 samples were identified as having the SMN1 g.27134T>G mutation that is in linkage-disequilibrium with the silent-carrier chromosome, a chromosome containing 2 copies of the SMN1 gene in cis in some populations.

The CarrierScan and ddPCR data were 100% concordant for SMA carrier state and g.27134T>G genotype. Concordance for SMN1 copy number was 94% and concordance for SMN2 copy number was 82%.

The results show that the highly-multiplexed CarrierScan assay is very effective and accurate for detecting SMA carrier genotypes.

## PgmNr 2552: The *c.1353C>G* (p.D451E) variant of *SLC6A1* gene is associated with myoclonic-atonic epilepsy and absence seizures in a family with autosomal dominant mode of inheritance.

### Authors:

M. Hajianpour <sup>1</sup>; D. Johnson <sup>2</sup>; S. Yoon <sup>1</sup>; J. Culp <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics/ Medical Genetics, East Tennessee State University - Quillen College of Medicine, Johnson City, Tennessee.; 2) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) ETSU Family Physicians of Bristol, Bristol, TN

**Myoclonic-atonic epilepsy (MAE)** is an autosomal dominant disorder characterized by onset of absence and myoclonic seizures in early childhood. Patients have delayed development before the onset of seizures and show varying degrees of intellectual disability following seizure onset. MAE is caused by heterozygous mutation in the *SLC6A1* gene on chromosome 3p25. We present a family with multiple affected individuals presenting with variable expression of MAE and with an autosomal dominant mode of inheritance; all carrying a variant of uncertain significance (VUS) of *SLC6A1* gene.

**Proband** is an 11-year-old female with intractable seizures, developmental delay, intellectual disability (IQ 45), ADHD, disruptive mood dysregulation, bicuspid aortic valve, SVT, joint hypermobility, and dysmorphic features (dolichocephaly, low-set ears, high nasal bridge, high-arched palate, retrognathia, pes planus). She was born at 32 weeks of gestation, to a 19 yo, G3, P2 mother. Birth weight was 5 lbs.

### Pertinent Family History:

- **Full brother;** 12 yo with seizures and ID (IQ 40)
  - **Maternal half-brother;** 14 yo with “drop-type absence seizures” and ID (IQ 48), autism spectrum disorder (ASD), ADHD, and bipolar disorder
  - **Full sister;** 9 yo with joint hypermobility, no seizures, IQ 80 (unaffected)
  - **Father** has ID
  - **Mother** has joint hypermobility, mild ID, and possible seizures
- All children are adopted by a single family (Parents unavailable for testing).

Molecular studies for ID panel of 140 genes (and Whole Exome Sequencing) showed a variant of uncertain significance (**VUS**) of *SLC6A1* gene: *c.1353C>G* (p.D451E), which is present in **two affected full-siblings**, their **affected maternal half-brother** and their **mother** (obligate carrier-not tested), but not present in their **unaffected sister**.

The affected family members show similar but **variable** clinical features, and all carry the *SLC6A1* gene variant p.D451E, which is most likely the cause of the phenotype in this family with an autosomal dominant mode of inheritance.

*SLC6A1* encodes a GABA transporter that removes GABA from the synaptic cleft. Pathogenic mutations of *SLC6A1* gene present with variable expression of **myoclonic-atonic epilepsy, absence seizures, developmental delay, and intellectual disabilities**, as seen in affected members of this family. The phenotype is segregating with the p.D451E variant indicating it is a likely pathogenic variant (VLP). To our knowledge, this variant has not previously been characterized as pathogenic, or benign.

## PgmNr 2553: Phenotypic expansion of TAF1-related syndrome: *TAF1* variants and clinical phenotypes.

### Authors:

H. Cheng <sup>1</sup>; S. Capponi <sup>2</sup>; E. Wakeling <sup>3</sup>; E. Marchi <sup>4</sup>; Q. Li <sup>5</sup>; M. Zhao <sup>6</sup>; C. Weng <sup>7</sup>; S. Piatek <sup>8</sup>; H. Ahlfors <sup>8</sup>; R. Kleyner <sup>9</sup>; A. Rope <sup>10</sup>; A. Lumaka <sup>11</sup>; P. Lukusa <sup>11</sup>; K. Devriendt <sup>12</sup>; J. Vermeesch <sup>12</sup>; J. Posey <sup>1</sup>; E. Palmer <sup>13</sup>; S. Moutton <sup>14</sup>; P. Mark <sup>15</sup>; L. Adès <sup>16</sup>; S. Sandaradura <sup>16</sup>; J. Lupski <sup>1</sup>; T. Roscioli <sup>17</sup>; P. Agrawal <sup>18</sup>; A. Kline <sup>19</sup>; K. Wang <sup>6</sup>; H. Timmers <sup>2</sup>; G. Lyon <sup>4,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA; 2) Medical Faculty-University of Freiburg, Freiburg, Germany; 3) North West University Healthcare NHS Trust, Harrow, UK; 4) Institute for Basic Research in Developmental Disabilities (IBR), Staten Island, USA; 5) University of Toronto, Toronto, Canada; 6) Children's Hospital of Philadelphia, Philadelphia, USA; 7) Columbia University Medical Center, New York, USA; 8) Great Ormond Street Hospital, London, UK; 9) Stanley Institute for Cognitive Genomics, Cold Spring Harbor Laboratory, USA; 10) Genome Medical, South San Francisco, USA; 11) University of Kinshasa, DR Congo; 12) University of Leuven, Leuven, Belgium; 13) University of New South Wales, Randwick, Australia; 14) Bordeaux University Hospital, Bordeaux, Bordeaux, France; 15) Spectrum Health, , Grand Rapids, USA; 16) The Children's Hospital at Westmead, Sydney, Australia; 17) Sydney Children's Hospital, Randwick, Australia; 18) Boston Children's Hospital, Boston, USA; 19) Harvey Institute for Human Genetics at GBMC, Baltimore, USA.

---

**Purpose:** We recently described a new neurodevelopmental syndrome caused by mutations involving the X-linked gene, *TAF1*, in 11 families. These individuals with missense variants or a duplication involving *TAF1* - encoding a subunit of basal factor TFIID for RNA polymerase II transcription - were shown to present early in life with hypotonia, developmental delay, and characteristic facial features, followed by later diagnoses of intellectual disability (ID) and/or autism spectrum disorder (ASD). We investigated *TAF1* rare variant genotype/phenotype correlations, hypothesizing a broader clinical phenotypic spectrum for this disorder, by identifying additional families through a genotype-first approach.

**Methods:** Families were identified through an international collaboration and primarily clinically-based sequencing. Variants of interest were identified with a range of bioinformatic approaches, followed by molecular modelling for those variants falling within structurally characterized domains of *TAF1*. Phenotyping cluster analysis was also performed using Human Phenotype Ontology (HPO) terms.

**Results:** We studied a total of 27 families, each with a distinct variant in *TAF1*. We used familial segregation analysis, clinical phenotyping, bioinformatics, molecular modelling, and clustering approaches to determine whether these variants contribute to the clinical phenotypes. We determined that five of these variants are likely benign, ten of the *de novo* variants are likely pathogenic, and the remaining twelve maternally inherited variants remain of uncertain significance. Two of the *de novo* variants occurred in females with significant skewing of X-chromosome inactivation. Phenotypes associated with *TAF1* variants show considerable pleiotropy and clinical

variability in their presentation, including autism spectrum disorder, brain anomalies (e.g. ventriculomegaly and corpus callosum abnormalities), seizures, hearing loss, and cardiac malformations.

**Conclusion:** Our allelic series data reveal that likely damaging TAF1 variant alleles can result in a broad phenotypic spectrum for TAF1 syndrome, whilst also broadening the range of molecular defects in TAF1 observed in human neurodevelopmental disorders. It also illustrates the current difficulties in assigning pathogenicity to inherited missense variants.

## PgmNr 2554: Targeted resequencing identifies genes with recurrent variation in cerebral palsy.

### Authors:

C.L. van Eyk<sup>1,2</sup>; M.A. Corbett<sup>1,2</sup>; M.S.B Frank<sup>1,2</sup>; D.L. Webber<sup>1,2</sup>; M. Newman<sup>3</sup>; J.G. Berry<sup>1,2</sup>; K. Harper<sup>1,2</sup>; B.P. Haines<sup>1,2</sup>; G. McMichael<sup>1,2</sup>; J.A. Woenig<sup>1,2</sup>; A.H. MacLennan<sup>1,2</sup>; J. Gecz<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Adelaide Medical School, University of Adelaide, Adelaide, South Australia, Australia; 2) Robinson Research Institute, Faculty of Health and Medical Sciences, University of Adelaide, Adelaide, S.A., Australia; 3) Alzheimer's Disease Genetics Laboratory, Centre for Molecular Pathology, School of Biological Sciences, University of Adelaide, Adelaide, South Australia, Australia; 4) South Australian Health and Medical Research Institute, Adelaide, S.A., Australia

---

A growing body of evidence points to a considerable and heterogeneous genetic aetiology of cerebral palsy (CP). To identify recurrently variant CP genes we designed a custom gene panel of 112 candidate genes. We tested 366 clinically unselected singleton cases with CP, including 271 cases not previously examined using next generation sequencing technologies. Overall, 5.2% of the naïve cases (14/271) harboured a genetic variant of clinical significance in a known disease gene, with a further 4.8% of individuals (13/271) having a variant in a candidate gene classified as intolerant to variation. In the aggregate cohort of individuals from this study and our previous genomic investigations, six recurrently hit genes contributed at least 4% of disease burden to CP: *COL4A1*, *TUBA1A*, *AGAP1*, *L1CAM*, *MAOB* and *KIF1A*. Significance of Rare VARIants (SORVA) burden analysis identified four genes with a genome-wide significant burden of variants, *AGAP1*, *ERLIN1*, *ZDHHC9* and *PROC*, of which we functionally assessed *AGAP1* using a zebrafish model. Our investigations reinforce that CP is a heterogeneous neurodevelopmental disorder with known as well as novel genetic determinants.

## PgmNr 2555: Spinal muscular atrophy screening workflow for rapid whole genome sequencing.

### Authors:

S. Batalov; M. Tokita; S. Chowdhury; Y. Ding; L. Puckett; C. Konersman; J. Friedman; M. Gaughran; N. Veeraraghavan; S. Kingsmore

[View Session](#) [Add to Schedule](#)

**Affiliation:** Rady Children's Institute for Genomic Medicine (RCIGM), 7910 Frost St, San Diego, CA 92123

---

Spinal muscular atrophy (SMA) is a neuromuscular disorder that results in the loss of motor neurons and progressive muscle wasting. SMA is caused by a biallelic loss or mutations in *SMN1* gene, which encodes SMN, a protein necessary for survival of motor neurons. A gene paralog (*SMN2*) that differs in five non-protein-changing base pairs has mostly aberrant splicing and its copy number may dramatically modify disease severity. With the advent of new therapies for SMA, a rapid diagnosis of SMA is crucial to optimize patient outcomes.

Rapid Whole-Genome Sequencing (rWGS) technology has transformed medicine as a cost-effective approach to detecting pathogenic variants on a genomic scale in patients with genetic diseases. The ability to detect *SMN1* and *SMN2* copy numbers using rWGS in one test is highly desirable to increase the diagnostic yield, yet is typically hindered by inability to correctly map short reads by conventional alignment software due to high degree of sequence similarity. We present a computational workflow that robustly estimates *SMN1* and *SMN2* copy numbers from a standard Illumina short-read rWGS.

**Method:** Following the conventional short-read alignment, an estimate of the high-quality alignment read counts for two specific loci is performed. While the entire sequence of the two genes is essentially identical, the 270 base pair area near exon 7 with five conserved differences is useable for correct unambiguous assignment of the reads to the correct gene paralog (under the condition of now standard low sequencing error rate, <0.5%). The read count model has been trained on 900 normal and eight true positive samples and normalized to different read lengths (2x100, 2x150 bp). The read counts are stochastic by the nature of their generation by random shearing, but at the level of target average genomic coverage of  $\geq 40\times$  are statistically significantly robust to determine *SMN1* and *SMN2* ploidy levels of 0, 1, 2, and 3+.

**Results:** 31 clinically confirmed retrospective SMA samples with various ploidy (including 8 DNA samples purchased from Coriell) were used for validation. For clinical correlation, *SMN1* and *SMN2* copy numbers were analyzed using custom-designed MLPA reagents, according to the manufacturer's recommendations. The MLPA method was perfectly concordant with the computational estimates. The computational workflow robustly estimates *SMN1* and *SMN2* copy numbers and is validated and approved for all prospective clinical samples at RCIGM.



## PgmNr 2556: IFIH1 gain of function variants: Further delineation of multiple faces of type I Interferonopathy.

### Authors:

S. Shimada<sup>1</sup>; J. Stephen<sup>2</sup>; S. Nampoothiri<sup>3</sup>; L.A. Wolfe<sup>1,4</sup>; C. Toro<sup>1,4</sup>; C.J. Tiffit<sup>1,4</sup>; D.R. Adams<sup>1,4</sup>; W.A. Gahl<sup>1,2</sup>; MCV. Malicdan<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) 1.NIH Undiagnosed Disease Program, Common Fund, Office of the Director and the National Human Genome Research Institute, National Institutes of Health, Bethesda, USA; 2) 2.Medical Genetics Branch, national Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 3) 3.Department of Pediatric Genetics, Amrita Institute of Medical Sciences and Research Center, Kerala 682041, India; 4) 4.Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

*IFIH1* encodes the protein interferon-induced helicase C domain-containing protein 1, otherwise known as melanoma differentiation-associated protein 5 (MDA5), which is a cytosolic sensor of double-stranded(ds) RNA as innate immune receptor. Enhanced type I interferon signaling secondary to gain-of-function mutations in *IFIH1* results in a range of neuroinflammatory phenotypes including Aicardi-Goutieres syndrome (AGS) and multi-system phenotypes like Singleton-Merten syndrome (SMS). Here we describe the clinical spectrum of three patients from three independent families with gain of function *IFIH1* mutations identified by exome sequencing.

All three probands have a typical AGS phenotype with various neurological and systemic manifestations, and brain imaging consistent with intracranial calcification. Probands 1 and 2, harboring heterozygous *IFIH1* variants (c.1009A>G;p.Arg337Gly and c.2342G>A;p.Gly781Glu, respectively) presented with an AGS and SMS overlapping phenotype, manifesting with abnormal dental findings (periodontitis and moderate enamel hypoplasia) and cutaneous findings (psoriasis and chilblain lesions). Atypical presentations were also noted; proband 1 had recurrent non-diabetic pancreatitis while proband 2 had axonal sensory motor neuropathy. Proband 3, with a heterozygous c.2335C>T;p.Arg779Cys variant, showed early-onset epileptic encephalopathy (EOEE) with severe global developmental delay.

To determine the cellular consequences of these variants, we measured the interferon activity in mammalian cells overexpressing the different mutations as compared to full length, wild-type *IFIH1*. Based upon this luciferase-based reporter assay, all three variants had increased interferon activities, confirming a gain-of-function mechanism. We hypothesize that these variants will lead to the upregulation of interferon-stimulated gene transcription.

Our results further delineate the spectrum of *IFIH1*-associated type 1 interferonopathy, ranging from neuroinflammation in SMS to an autoimmune phenotype.

## PgmNr 2557: Identifications of novel mutations and copy number variation (CNV) in children affected with Tay-Sachs disease from India.

### Authors:

M. Mistri <sup>1</sup>; M. Kamate <sup>2</sup>; M. Kabra <sup>3</sup>; R. Puri <sup>4</sup>; G. Katta <sup>5</sup>; S. Hariharan <sup>6</sup>; S. Nampoothiri <sup>7</sup>; F. Sheth <sup>1</sup>; J. Sheth <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) FRIGE's Institute of Human Genetics, Ahmedabad, Gujarat, India; 2) KLES Prabhakar Kore Hospital, Belgaum, Karnataka, India; 3) All India Institute of Medical Sciences (AIIMS), New Delhi, India; 4) Sir Ganga Ram Hospital, New Delhi, India; 5) Kasturba Medical College, Manipal, Karnataka, India; 6) SAT Hospital, Thiruvananthapuram, Trivandrum, India; 7) Amrita Institute of Medical Science & Research Centre, Kerala, India

---

Tay-Sachs disease (TSD) is an autosomal recessive storage disorder caused by impaired activity of the lysosomal enzyme hexosaminidase-A (EC3.2.1.52) due to the mutation in HEXA gene. As per HGMD database, ~190 mutations have been reported from different ethnic groups, although very few disease-causing mutations are known in Indian TSD patients. Hence, the present study was undertaken to identify more novel mutations in HEXA gene in Indian TSD patients to evaluate genotype-phenotype correlation and to establish the common disease-causing mutations that can be utilized for carrier screening in the mass population. The overall study detected 26 mutations belonging to 35 affected patients with infantile-onset TSD, through bidirectional sequencing of HEXA gene and MLPA-based approach (MRC-Holland, P199-B) were also applied to investigate for the potential occurrence of large HEXA deletions/duplications. The most common mutations detected in six individual [18%] was c.1278insTATC (p.Y427IfsX5) while p.E462V mutation was found in 5 individuals [14%]. Overall, 17 patients harbors 15 novel variants, including 7 missense variants [p.V206L, p.Y213H, p.R252C, p.F257S, p.C328G, p.G454R, p.P475R], 4 nonsense variant [p.S9X, p.E91X, p.W420X, p.W482X], 2 splice site variants [c.347-1G>A, c.460-1G>A] and 2 small deletion [c.1349delC (p.A450VfsX3) and c.52delG (p.G18Dfs\*82)]. While remaining 18 patients harbors 11 previously reported variants that includes 6 missense variants [p.M1T, p.R170Q, p.D322Y, p.D322N, p.E462V, p.R499C], 1 nonsense variant [p.Q106X], 2 splice site variants [c.1073+1G>A, c.459+4A>G], one 4bp insertion [c.1278insTATC (p.Y427IfsX5)] and one large deletion of exon 1. In silico analyses of all novel mutations of HEXA gene were shown to be probably damaging with a deleterious effect on protein function. Protein homology modeling studies further established the effects of novel mutations occurred at highly evolutionarily conserved and functionally active domain residues in the protein leading to conformational changes in HEXA protein. In conclusion, Indian infantile TSD patients provide newer insight into the molecular heterogeneity of the TSD. Combining present study and our earlier studies, we have observed that 67% genotypes found in Indian TSD patients are novel which are associated with severe infantile phenotypes. Present study again reconfirms that, 2 variants E462V and c.1277\_1278insTATC are most common in nearly 32% (11/35) of Indian TSD patients.

## PgmNr 2558: Massively parallel sequencing: A robust approach to genotyping the myotonic dystrophy type 2 locus in the general population.

### Authors:

T. Oyeniyi <sup>1</sup>; R. Krahe <sup>2</sup>; D. Monckton <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Molecular, Cell and System Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom; 2) 2Division of Basic Science Research, University of Texas MD Anderson Cancer Center, Houston, TX

Myotonic dystrophy type 2 (DM2) is an autosomal neuromuscular disorder caused by a microsatellite expansion in intron 1 of the *CNBP* gene. The DM2 locus is a compound microsatellite having a complex array of repeats with three motifs arranged as (TG)<sub>n</sub>(TCTG)<sub>n</sub>(CCTG)<sub>n</sub>. The mutation is an expansion of the CCTG motif from the non-disease associated <CCTG<sub>75</sub> into the pathogenic range of an average of CCTG<sub>500</sub>. The conventional molecular method of diagnosis is fragment length analysis by PCR, gel electrophoresis and/or Southern hybridization. However, these methods do not account for the highly polymorphic TG and TCTG motifs of the repeat array. This makes it hard to distinguish between two non-disease associated alleles that are identical by state (length) but are different by structural (sequence) arrangement. With this limitation, unaffected heterozygous individuals may falsely appear as homozygous requiring further analysis to confirm their true genotype. This fast and accurate method for genotyping non-disease associated alleles at the DM2 locus. Primers were barcoded such that individual samples were identified for analysis in a pool of many, up to 384 individuals of the general Scottish population were screened at one sequencing run. To evaluate the sensitivity of the assay, 15 individuals from Texas who had been previously genotyped by fragment length analysis, were screened.

Sequencing revealed a typical sequence arrangement of (TG)<sub>n</sub>(TCTG)<sub>n</sub>(CCTG)<sub>n</sub> and an 8-bp flanking sequence TCTGTCTC downstream of the last CCTG repeat. The allele range at the three motifs was (TG)<sub>10</sub> to (TG)<sub>30</sub>, (TCTG)<sub>2</sub> to (TCTG)<sub>13</sub> and (CCTG)<sub>8</sub> to (CCTG)<sub>27</sub>, the modal alleles being (TG)<sub>21</sub>, (TCTG)<sub>9</sub> and (CCTG)<sub>15</sub>, and together, these three constituted the most frequent haplotype. A 12-bp sequence interruption (GCTG)<sub>1</sub>(CCTG)<sub>1</sub>(TCTG)<sub>1</sub> was observed within the CCTG motif in most alleles, while a few alleles had none or more insertion(s) of the interrupting sequence. In terms of total allele length, a unimodal distribution was observed amongst the cohorts, within a range of 106 bp to 186 bp with the mode at 138 bp. Of the total population, 5.8 % was homozygous by sequence as opposed to the 9 % homozygous by length.

NGS as an approach to genotyping the DM2 locus is robust and accurate, it offers a resolution to the base pair and remarkably high throughput that cannot be attained with the traditional fragment length analysis methods.

## PgmNr 2559: Analysis of neuromuscular problems using NGS panels in Ecuadorian population.

**Authors:**

V. Espin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genetics, Hospital Carlos Andrade Marin, Quito, Pichincha, Ecuador

---

The study and analysis of neuro-muscular problems is a very complex task. There are many difficulties to get a precise diagnosis because clinical presentations are quite similar among these diseases. In many cases this has capital importance in order to be able to establish the natural history of the disease, recurrence risk and sometimes to take choices of reproductive medicine.

Molecular technology has made the diagnostic approach easier.

We analyzed 27 patients who presented with neuromuscular symptoms at the Genetics Service of the Hospital Carlos Andrade Marín (Quito) between in 2017-2018.

An NGS panel of 123 genes was used, finding a conclusive diagnosis in 16 patients (Detection rate 59.25%). If the probable ones (Variable Unknown Significance VUS) are added, the diagnosis rate would reach 88.8%.

Ecuador is not a wealthy country. The option of applying this technology to patients with suspected hereditary neuromuscular disorder becomes a feasible and accessible option to reach diagnosis in a large number of these cases.

## PgmNr 2560: Dissecting the genetic basis of focal cortical dysplasia: A large cohort study.

### Authors:

S. Baldassari <sup>1</sup>; T. Ribierre <sup>1</sup>; E. Marsan <sup>1</sup>; H. Adle-Biassette <sup>2,3</sup>; S. Ferrand-Sorbets <sup>4</sup>; C. Bulteau <sup>4</sup>; N. Dorison <sup>4</sup>; M. Fohlen <sup>4</sup>; M. Polivka <sup>3</sup>; S. Weckhuysen <sup>1,5</sup>; G. Dorfmueller <sup>4</sup>; M. Chipaux <sup>4</sup>; S. Baulac <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Sorbonne Université, UPMC Univ Paris 06, UMR S 1127; INSERM, U1127; CNRS, UMR 7225; Institut du Cerveau et de la Moelle épinière (ICM), Hôpital Pitié-Salpêtrière, F-75013, Paris, France.; 2) INSERM UMR 1141, Hôpital Robert-Debré, F-75019, Paris, France; Faculté de Médecine Denis Diderot, Université Paris 7, Paris, France.; 3) Service d'Anatomie et de Cytologie Pathologiques, Hôpital Lariboisière, APHP, F-75010, Paris, France; 4) Department of Pediatric Neurosurgery, Rothschild Foundation Hospital, F-75019 Paris, France.; 5) Department of Neurology, University Hospital Antwerp, Belgium.

---

Genetic malformations of cortical brain development, such as focal cortical dysplasia (FCD) and hemimegalencephaly (HME), are major causes of severe pediatric refractory epilepsies subjected to neurosurgery. Neuropathological hallmarks of FCDs include enlarged dysmorphic neurons (DNs) and balloon cells (BCs). Here we aim to provide a comprehensive assessment of the contribution of germline and somatic variants in a large cohort of FCD/HME patients.

We enrolled in a monocentric study 80 children with drug-resistant epilepsy and a postsurgical neuropathological diagnosis of FCD1, FCD2 or HME. We performed targeted gene sequencing ( $\geq 2000\times$  read depth) on matched blood-brain samples to search for low-allele frequency variants in mTOR pathway and FCD genes.

We were able to elucidate 29% of FCD1 patients and 63% of FCD2/HME patients. FCD1 cases carried somatic loss-of-function variants in the N-glycosylation pathway-associated *SLC35A2* gene. FCD2/HME cases carried somatic gain-of-function variants in *MTOR* and its activators (*AKT3*, *PIK3CA*, *RHEB*), as well as germline, somatic and two-hit loss-of-function variants in its repressors (*DEPDC5*, *TSC1*, *TSC2*). We show that panel-negative FCD2 cases display strong pS6-immunostaining, stressing that all FCD2 belong to mTORopathies. Analysis of microdissected cells demonstrated that DN and BCs carry the pathogenic variants. We further observed a correlation between the density of pathological cells and the variant detection likelihood. Single cell microdissection followed by sequencing of enriched pools of DN unveiled a second-hit somatic loss-of-heterozygosity in a *DEPDC5* germline case.

This study indicates that FCD1 and FCD2/HME are two distinct genetic entities: a subset of FCD1 are related to glycosylation defects, while all FCD2/HME are somatic mTORopathies. We provide a framework for efficient genetic testing in FCD/HME, linking neuropathology to genetic findings and emphasizing the relevance of molecular evaluation in the pediatric epileptic neurosurgical population.

## PgmNr 2561: A novel candidate gene for hereditary spastic paraplegia in a consanguineous Malian family.

### Authors:

S. Diarra <sup>1,2</sup>; C. Grunseich <sup>1</sup>; C. Blackstone <sup>1</sup>; T. Coulibaly <sup>2</sup>; S.H. Diallo <sup>3</sup>; S. Diallo <sup>3</sup>; L. Cisse <sup>2</sup>; O. Coulibaly <sup>4</sup>; C.A.K. Cisse <sup>2</sup>; A. Yalcouye <sup>2</sup>; A.J.K. Ouabo <sup>2</sup>; C.O. Guinto <sup>2</sup>; G. Landoure <sup>1,2</sup>; K.H. Fischbeck <sup>1</sup>; The H3Africa Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurogenetics Branch, NIH, Bethesda, MD.; 2) Service de Neurologie, CHU du Point "G", Bamako, Mali; 3) Service de Neurologie, CHU de Gabriel Touré, Bamako, Mali; 4) Service de Chirurgie Pédiatrique, CHU de Gabriel Touré

**Introduction:** Hereditary spastic paraplegias (HSPs) are a heterogeneous group of neurodegenerative disorders characterized by lower extremity spasticity. At least 69 genes have been associated with the disease, but many other clinically characterized forms remain with no molecular diagnosis. In addition, very little data about HSPs exist in sub-Saharan Africa.

**Objective:** To characterize a Malian family with HSP and identify the underlying genetic defect.

**Methods:** After consent, patients went through a thorough neurological examination. Blood chemistries, brain and spinal imaging were done to exclude common causes. EEG was also performed to assess epileptiform activity. DNA was extracted for genetic analysis and fibroblasts were collected from patients and relatives for cell studies.

**Results:** We identified a consanguineous family in which three siblings presented with early onset HSP features. Symptoms started with toe walking followed by frequent falls, and worsened overtime. In addition, they presented generalized seizures. Neurological examination found brisk reflexes in four limbs with spastic gait, lower limb weakness, and equine feet. The oldest patient presented also problems with speech and behavioral disorder, and both had cognitive decline. Vitamin E levels and HTLV-1 serology, causing tropical paraplegia, were normal. Brain imaging showed thin corpus callosum with cortical and cerebellar atrophy, and EEG reported generalized spike-waves in the oldest patient, and slow background in the youngest. SPG gene panel testing (58 genes + *SPG4* del/dup + *mtDNA*) was negative. Exome sequencing in two affected individuals showed variants in two different genes. These variants segregated with the disease in this family. One of these variants, in the *AP2A2* gene, is not present in SNP databases including dbSNP, gnomAD, ExAc, and 1000Genomes. In addition, the gene has not been previously associated with any disease, but has a highest expression in spinal cord and fetal brain. Moreover, the protein product is a member of a complex known to be associated with other HSPs. The substituted amino acid lies in a critical part of a highly conserved domain and likely has a role in endocytosis.

**Conclusion:** Our findings suggests a novel candidate gene for HSP. Further studies are underway to evaluate this variant in the Malian population and its functional effects in cell culture.

## PgmNr 2562: Large pedigree subfamily's whole exome trio analyses from isolated South-Eastern Moravia (Czech Republic) population indicate heterogeneous genetic nature for Parkinsonism development.

### Authors:

R. Vodicka <sup>1</sup>; R. Vrtel <sup>1</sup>; K. Kolarikova <sup>1</sup>; K. Mensikova <sup>2</sup>; P. Kanovsky <sup>2</sup>; J. Stellmachova <sup>1</sup>; M. Prochazka <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Medical Genetics, University Hospital, Olomouc, Czech Republic; 2) Department of Neurology, Faculty of Medicine and Dentistry, Palacky University, University Hospital, Olomouc, Czech Republic

---

We described higher prevalence of parkinsonism in the small isolated southeastern Moravian region in the Czech Republic and in addition with well assembled eleven generation large pedigree.

The study aims to set proper and accessible subfamily trios from the pedigree to find shared genetic variants potentially associated with parkinsonism using whole exome analyses.

We used NGS Ion AmpliSeq Exome method (IonTorrent) for five subfamily trios. Each trio comprised of two affected and one healthy person. DNA exome libraries were sequenced on IonPI chips. Variants were predicted using pipeline Torrent Suite/Ingenuity Variant Analysis/Two case/control analysis. Final filtering was done with respect to population frequency (Global MAF?1%), variant effects and biological context (Parkinsonism responsible genes). Last filter was done with respect to the segregation of the disease within particular subfamily.

Almost whole exome was sequenced with coverage 1-20 and 90 % of exome was covered more than 20x in all samples. Together more than 70.000 variants with average base coverage depth 75 were analyzable in all trios before filtering.

There were found no one founder pathogenic variant in the subfamilies through the pedigree. Each subfamily trio shows different set of suspected variants. Trio A shares 2 variants with trio D (novel variant NM\_002386.3:c.322G>A;p.A108T in the gene *MC1R/RP11-566K11.2* and rare variant NM\_015210.3:c.1445C>T;p.A482V in the gene *MTCL1* ), trio B shares 1 rare variant with trio C (NM\_001256864.1:c.1817A>C;p.H606P in the gene *DNAJC6*). In addition, in trios C and E there were found two novel gene *CSMD1* variants NM\_033225.5: c.3335A>G,p.E1112G and c.4071C>G; p.I1357M respectively.

It can be concluded that genetic contribution to the disease in our large pedigree could be heterogeneous. Detailed whole exome analyses in genetically isolated parkinsonism patients could contribute to further understanding of the molecular-genetic mechanism and background of the disease. This study was supported by MH CZ – DRO (FNOL, 00098892).



## PgmNr 2563: SMA screening in 10439 newborns with a novel capillary electrophoresis based method.

### Authors:

H. Zhu; L. Wang; Y. Chen

[View Session](#) [Add to Schedule](#)

**Affiliation:** Prenatal Diagnosis Center, The Sixth Medical Center of PLA General Hospital, Beijing, Beijing, China

---

**Objectives:** Spinal Muscular Atrophy (SMA) is a relatively common autosomal recessive disorder. It has a carrier frequency of 1/40 to 1/60 and an incidence rate of about 4 to 10 in 100,000 in newborns, which are basically the same among various population. Currently, there is no effective SMA neonatal screening method. To detect SMA, MLPA and qPCR are commonly used in clinical practice. However, both of them have their limitations. In this study, we have developed a novel rapid, simple and accurate method based on CE platform. To the best of our knowledge, no similar method has been previously reported.

**Methods:** Previous methods quantitatively detect the copy number of SMN genes by comparing results of targets and a control gene. We took Survival of Motor Neuron 1 telomeric Pseudogene (SMNP), which was highly homologous to SMN1/2, as the control. Using the same primer set, we could amplify the target and control with equivalent primer binding affinity, which greatly improved the stability and quantitative performance of the detection system. This new method was validated using 313 samples, which were all tested by qPCR in parallel. All SMA-positive samples were then confirmed by MLPA. After validation, we tested 10493 samples and compared the results from new method and qPCR. All SMA-positive samples were confirmed by MLPA. Incidence rate and carrier frequency of Chinese Han population were determined based on these results.

**Results:** In the validation of 313 samples, 5 SMA carriers and 308 normal samples were detected by CE method, which was consistent with the result given by qPCR. The 5 SMA carriers were confirmed by further MLPA method. In the test of 10493 samples, CE method successfully detected 98.3% (10315/10493) samples, with one SMA patient and 179 carriers. These 180 samples were then verified by MLPA method and showed highly concordant results, which were better than those results from qPCR method. According to the above data, the frequency of SMA carriers in Chinese Han population was 1.73%.

**Conclusions:** For the first time, we have established a method for quantitative detection of copy number of *SMN* gene based on CE platform. The accuracy of this new method was demonstrated by a developmental validation and a large population test, indicating our novel method was accurate, stable and affordable for SMA screening.



## PgmNr 2564: Clinical validation and implementation of a highly efficient and sensitive dual molecular diagnostic assay for myotonic dystrophy type 1.

### Authors:

J. Alaimo<sup>1,2,3</sup>; I. Thiffault<sup>1,2,3</sup>; M. Farooqi<sup>1,2,3</sup>; C. Saunders<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pathology and Laboratory Medicine, Children's Mercy Hospital, Kansas City, MO.; 2) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO.; 3) University of Missouri-Kansas City School of Medicine, Kansas City, MO.

---

Myotonic dystrophy type 1 (DM1) is caused by a (CTG)<sub>n</sub> repeat expansion in the 3' UTR of *DMPK* and is an important and often overlooked consideration in the work up of a hypotonic infant. (CTG)<sub>n</sub> repeats are refractory to detection by short read sequencing and therefore, require other specialized methods for accurate quantification including PCR and Southern blot analysis. Current laboratory-developed PCR tests are limited to amplification of >100 repeats and require multiple reactions for either expansion detection or accurate sizing due to amplification dropouts. Furthermore, Southern blotting is labor-intensive, expensive, and has a significantly reduced turnaround time. Here, we describe the clinical validation and implementation of a new commercially available PCR assay that overcomes these significant hurdles. Triplet-primed PCR was performed on all samples and resolved using capillary electrophoresis. The assay generated numerical values for alleles up to and including 200 repeats and a categorical value for alleles >200 repeats to facilitate genotyping. Size estimation beyond 200 repeats were flagged by both an expanded stutter pattern and a corresponding pile-up peak. Larger expansions were resolved using an agarose gel electrophoresis method. A validation set of 23 samples was tested, including a subset obtained from previous CAP proficiency testing. Samples were selected to provide multiple representatives in each numerical category throughout the dynamic range which included the normal (5-34 repeats), premutation (35-49 repeats), and various disease ranges (>50 repeats). We observed 100% concordance between our results for this sample set and previously reported results with 100% sensitivity, specificity, accuracy, and precision. In addition, we were able to clearly resolve zygosity in all samples. Mosaicism of at least 10% was detectable. Another major advantage of this testing is a total hands-on time requirement of only 60 minutes, a significant improvement over dual PCR / Southern blot methods. Overall, this assay resulted in a faster, accurate, and cost-effective approach for reaching a DM1 molecular diagnosis and highlights the continuing advances in the molecular diagnostic testing space.

## PgmNr 2565: Beyond SNPs and INDELs: Identifying repeat expansions in an unresolved ataxia exome cohort.

### Authors:

H.P. Subramanian; S. Andani; D. del Gaudio; S. Das

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Human Genetics, The University of Chicago, Chicago, Illinois.

Ataxias are a group of neurological disorders with tremendous genetic heterogeneity. Around 50-60% of dominant hereditary ataxias are estimated to be caused by repeat expansions while the majority of other ataxias are due to mutations in one of several hundred genes. The diagnostic yield in ataxia patients through exome sequencing has been shown to be around 40-50%, leaving a large number of patients with no diagnosis. The possibility exists of patients with repeat expansions that have not been detected by exome sequencing. With the advent of new bioinformatic algorithms to identify repeat expansions from sequencing data, we sought to identify the presence of such expansions in exome negative ataxia cases.

We re-analyzed exome sequencing data of 183 ataxia patients with no clear pathogenic variants in a curated set of ataxia genes. We utilized the exSTRA R package (Tankard *et al.* 2018, The American Journal of Human Genetics, 103(6), 858-873) to identify repeat expansions in coding regions of certain spinocerebellar ataxia genes (*ATN1*, *ATXN1*, *ATXN2*, *ATXN3*, *ATXN7*, *CACNA1A*, *TBP*). Repeat primed PCR (RP-PCR) was performed to confirm and size the repeat expansions.

exSTRA works by identifying outliers in the cohort. 16 such outliers were initially identified and 3 out of 16 outliers were confirmed to be a true repeat expansion by RP-PCR. Two samples had a full expansion repeat in *CACNA1A* (11,22 repeats) and *ATXN7* (11,76 repeats) respectively and one sample had a reduced penetrance repeat in *ATXN2* (31,32 repeats). We are in the process of performing similar analysis in additional 33 ataxia patients negative by exome sequencing. In addition, we performed a preliminary analysis to screen for the presence of the recently described intronic AAGGG repeat expansion in the *RFC1* gene in patients with cerebellar ataxia, neuropathy, vestibular areflexia syndrome (CANVAS). Analysis of off-target reads in 27 samples indicates the presence of the AAGGG repeat in 3 samples that needs to be confirmed by RP-PCR. Our results on these ongoing studies will be presented.

Reanalysis of exome sequencing data appears to be a promising screening tool for the detection of repeat expansions resulting in an increase in diagnostic yield. With the emergence of better detection algorithms, the ability to use exome data for both mutation detection as well as repeat expansions will be useful for diagnostic testing purposes and allow for more comprehensive and cost-effective testing.

## PgmNr 2566: Comparison of Stereocilin gene copy number detection by multiplex ligation-dependent PCR amplification (MLPA) and multiplexed droplet digital PCR (ddPCR).

### Authors:

R. Margraf<sup>1</sup>; R. Mao<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah.; 2) Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah

---

Stereocilin gene (*STRC*) mutations are a common cause of moderate autosomal recessive hearing loss. 11% of all individuals with isolated non-syndromic hearing loss have a large *STRC* deletion, ranging from partial to full gene. The *STRC* gene is 98% homologous with its pseudogene (*pSTRC*), which includes a 100% homologous region from exons 1 to 15. *STRC* deletions cannot be accurately detected using NGS or microarray due to the *pSTRC* pseudogene homology. This report compares two assays that could accurately and specifically detect *STRC* deletions even with the *pSTRC* homology: Multiplex Ligation-Dependent PCR Amplification (MLPA) and copy number droplet digital PCR (ddPCR) assays. The *STRC* MLPA uses the SALSA MLPA probemix P461-A1 DIS from MRC- Holland. The MLPA had 5 probes located within or near *STRC* exons 19 (2 probes), 23, 24, and 25; and included probes in flanking genes, *CKMT1B* and *CATSPER2*. A multiplex *STRC* copy number ddPCR assay was developed on the Bio-Rad QX200™ Droplet Digital™ PCR system. The ddPCR assay had four targets for copy number: three within *STRC* (exon 19, near exon 23, and intron 25) and *CATSPER2* gene exon 7. Each of the *STRC* ddPCR assays were nearby an MLPA probe location.

Fifty-two de-identified samples were tested for copy number by *STRC* MLPA and ddPCR and results were compared for all samples. The *STRC* copy number overall and also the copy number per similar targeted exon locations were in agreement for 51 of 52 samples. Thirty-three samples were wild type, 10 were heterozygous *STRC* deletion, 4 had homozygous deletion, and 4 had a duplication. The one discrepant sample was wild type by ddPCR but was heterozygous deletion for one MLPA probe at exon 19, and had the same respective results upon repeats. This MLPA exon 19 location overlaps by 3 nt to the ddPCR assay, so the discrepant result may be a real deletion present only under the MLPA probe or a polymorphism affecting the MLPA result only. To test for assay specificity to *STRC* versus *pSTRC*, the four homozygous *STRC* gene deletion samples were run at 2x the recommended DNA input, and these samples remained at 0 copies at all *STRC* probe locations in both assays. In conclusion, 98% of the samples had concordance in *STRC* copy number between MLPA and ddPCR assays, and both assays showed specificity to the *STRC* gene versus *pSTRC*.

## PgmNr 2567: The more the merrier? Comparison of the diagnostic yield between focused vs comprehensive panels for inherited retinal disorders.

### Authors:

M. Pantrangi <sup>1</sup>; M. Reeves <sup>2</sup>; K. Goetz <sup>2</sup>; J. Schroeder <sup>1</sup>; J. Rath <sup>1</sup>; G. Londre <sup>1</sup>; J. Weber <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Prevention Genetics, LLC, Marshfield, Wisconsin.; 2) eyeGENE, National Institute of Health, National Eye Institute, Bethesda, MD

---

Inherited retinal disorders (IRDs), such as Retinitis pigmentosa, Stargardt disease, and Leber congenital amaurosis (LCA), are the leading cause of blindness in the Western world, affecting ~1 in 3,000 individuals. IRDs have significant phenotypic overlap and molecular testing can help confirm a clinical diagnosis or achieve an accurate molecular diagnosis. In addition, an accurate diagnosis may be required to help guide treatment (e.g., *RPE65* gene therapy Luxturna<sup>TM</sup>). Identifying the genetic cause for IRDs is challenging due to genetic heterogeneity. To date, ~300 loci have been mapped and over 250 genes associated with IRDs have been identified (<https://sph.uth.edu/retnet/>).

PreventionGenetics (PG) offers a comprehensive IRD panel, which includes analysis of the mutational hotspot ORF15 region in *RPGR* and sequencing of over 300 genes with copy number variant (CNV) detection. This panel is an upgrade from our previous tier system of testing which included 107 common IRD genes in Tier 1 and 173 less common IRD genes in Tier 2. Tier 2 genes were tested only when Tier 1 was negative or indeterminate. Test outcomes were considered indeterminate if a single pathogenic variant was found in an autosomal recessive gene along with a variant of uncertain significance (VUS) or if only VUS were detected.

To date, we have tested ~530 individuals with IRDs. The positive rate between the Tier 1 testing (76 individuals) and the 300+ gene panel testing (455 individuals) was compared. Interestingly, the positive rate for Tier 1 testing alone was 47% and the indeterminate rate was 51%, while the positive rate for the 300+ gene panel was 42% and its indeterminate rate was 58%. The genes with the highest positive rates included *USH2A* (20%); *ABCA4* (15%); *EYS* (~6%); *RPGR*, *CEP290*, *CRB1*, and *PRPF31* (each ~4%); *RPE65*, *CERKL*, and *CLN3* (each ~3%); *BBS1* (2.5%); *CYP4V2*, *DHDDS*, *FAM161A*, *PCARE*, *PDE6B*, *RDH12*, and *RHO* (each ~2%); *CHM* and *NR2E3* (both ~1.5%); and *PROM1* and *TOPORS* (both ~1%). All of these genes had previously been included in the Tier 1 test, with the exception of *BBS1* and *CYP4V2*.

The current study suggests that including a larger number of genes increases the VUS burden. However, the positive rate is comparable between the focused panel with the most common IRD genes and a larger comprehensive panel. In summary, we present data to demonstrate the clinical validity of focused multigene panels for the diagnosis of clinically and genetically heterogeneous IRDs.

## PgmNr 2568: Large structural variants account for a substantial fraction of unsolved cases of inherited retinal disease.

### Authors:

S.P. Daiger <sup>1,2</sup>; L.S. Sullivan <sup>1</sup>; S.J. Bowne <sup>1</sup>; E.L. Cadena <sup>1</sup>; D. Koboldt <sup>3</sup>; E.A. Pierce <sup>4</sup>; D.G. Birch <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics Center, School of Public Health, The Univ. of Texas Health Science Center, Houston, TX; 2) Ruiz Dept. of Ophthalmology and Visual Science, The Univ. of Texas Health Science Center, Houston, TX; 3) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH; 4) Ocular Genomics Institute, Massachusetts Eye and Ear Infirmary, Boston, MA; 5) Retina Foundation of the Southwest, Dallas, TX

---

**Background.** Inherited, degenerative retinal diseases (IRD's) affect more than 1.5 million people worldwide and are a major cause of legal blindness and loss of vision in young adults. Currently, there are a number of registered clinical trials and approved drugs for treating IRD patients with known disease-causing genes and mutations. Thus identifying the underlying cause of disease is not only of benefit to patients and families, but is a prerequisite for treatment. Panel testing of known IRD genes using next-generation sequencing (NGS) detects the cause in from 60% to 80% of cases, depending on the population tested. Among the remaining cases we have identified several large, complex, disease-causing structural variants not readily detected by conventional NGS.

**Results.** Our research focuses on families with autosomal dominant retinitis pigmentosa (adRP) and related dominant retinopathies. To date we have identified the cause in approximately 80% of cases. Our goal is to find the causes in the remaining 20%. Methods include linkage mapping, retinal targeted-capture NGS, whole-exome and whole-genome NGS, quantitative-PCR and quantitative-NGS, and Linked-Read 10X Genomics Chromium™ sequencing. Among the unsolved adRP cases one family was found to have a balanced translocation of the q terminus arms of chromosomes 2 and 4 involving 35 Mb and 73 Mb respectively. The likely cause of disease in this family is dysregulation of the *LRAT* gene on chromosome 4. The family also reports an increased incidence of miscarriages. In a second family, with dominant macular degeneration, we found a 60 kb tandem duplication of the *PRDM13* gene on chromosome 6q. The tandem copies of the *PRDM13* gene appear to be intact with unaltered flanking regulatory sequences. The likely cause of disease in the family is overexpression of the *PRDM13* gene. The family has clinical features of North Carolina Macular Dystrophy, with appearance of unique macular pits at birth, suggesting an *in utero* effect. Finally, we and other investigators have reported megabase deletions of the *PRPF31* gene on chromosome 19q as another cause of adRP.

**Conclusions.** These findings suggest that a substantial fraction of the unsolved IRD cases have structural mutations, in known disease-causing genes, not easily detected by conventional NGS. Further, large structural variants may have additional consequences such as increased miscarriages.

## PgmNr 2569: Detection, characterization, and breakpoint refinement of balanced rearrangements by optical mapping in clinical cases.

### Authors:

A. Hastie <sup>1</sup>; S. Schwartz <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bionano Genomics, San Diego, California.; 2) Laboratory Corporation of America, Research Triangle Park, North Carolina

---

The current standard of care for molecular diagnosis of many constitutional syndromic disorders and reproductive disorders is often chromosomal karyotype analysis, fluorescence in situ hybridization (FISH), and chromosomal microarray (CMA). Each of these methodologies has shortcomings that whole genome optical mapping (Bionano Genomics) is able to alleviate. Studies have clearly shown that approximately 6% of fetuses with *de novo* balanced rearrangements have serious congenital anomalies. With the advent of CMA an associated deletion was identified in some patients, but in most cases no underlying cause for the anomalies was found. In this study we demonstrate the ability of Bionano optical mapping to detect balanced translocations in five samples, all of them corresponding to translocations found by karyotyping. However, because Bionano Optical mapping has much higher resolution compared to karyotyping, we were able to map translocation breakpoints within specific genes, which will ultimately allow for a more complete diagnostic test. While the optical mapping could molecularly localize breakpoints in all of the samples, one sample proved to be unique and complex. In this sample, a t(4;21) translocation was found by karyotype and a deletion found by CMA; however with optical mapping, we could determine that there were three breakpoints with a 107 kbp deletion on chromosome 4 next to one breakpoint and a 57 kbp deletion on chromosome 21 next to another junction. There was an additional breakpoint with no loss of material that resulted in a nested inversion in the der(4) translocation. In summary, optical mapping could find three breakpoints and two deletions, and define genes directly impacted – broken or deleted, while CMA and karyotyping could only find two breakpoints and one deletion together. In addition, optical mapping identified a balanced t(3;20) which broke genes on both chromosomes, a balanced t(4;9) which also broke two genes, and two additional balanced translocations in different samples. In addition to the ability for Bionano optical mapping to detect most or all variants detected by CMA and karyotyping, it can detect smaller structural variants (>500bp) very robustly, making it a great complement to whole genome sequencing to integrate SNP and Indel calling with structural variation calling.

## PgmNr 2570: Rare case of monocentric isochromosome Y with inversion-duplication of p arm in patient presenting with azoospermia.

### Authors:

B. Kar <sup>1</sup>; A. Afreen <sup>1</sup>; S. Kundavi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Studies & Research, The Madras Medical Mission, Chennai, Tamil Nadu, India; 2) Institute of Reproductive Medicine & Women's Health, The Madras Medical Mission, Chennai, Tamil Nadu, India

### Introduction

The occurrence of isochromosome Y along with mosaicism is relatively common in present literature but the presence of inversion duplication of the p arm of the Y chromosome has not been widely reported. The present study describes a 36-year-old male patient with azoospermia who had two cell lines – one cell line with rare isochromosome monocentric derivative Y and the other with a 45,X complement which is reported for the first time to the best of our knowledge

### Methods

Informed consent was taken to perform karyotyping for the proband and his wife as well as to test for Y chromosome microdeletion for the male patient. The cytogenetic analysis involved karyotyping and C banding. Y microdeletion study was performed to test for the presence or absence of selected AZF regions. Fluorescent in situ hybridization (FISH) test was performed using probes for SRY gene, heterochromatin region of the Y chromosome and centromeric region of the X chromosome.

### Results

Karyotyping of the proband revealed two cell lines – mos 46,X,ider(Y)(q10)inv(Y)(p11.3q11.1)/45,X. C banding and FISH revealed inversion-duplication of a segment of the p arm of the Y chromosome. The derivative chromosome contained two SRY genes but only one centromere. Y chromosome microdeletion studies were negative for the regions tested. The proband's wife had 46,XX genetic complement.

### Conclusion

Individuals containing such abnormalities have variable phenotypes. Since such structural chromosomal anomalies are known to be highly associated with increased risk of aneuploidy, the proband and his wife were advised to undergo artificial reproductive techniques using donor sperm in order to achieve conception.



## PgmNr 2571: Emanuel syndrome: A clinical case report.

### Authors:

J. Villalta Fallas; F. Suarez Obando; M. Guerra Gomez; R. Quero Angarita; C. Manotas Ahumada; A. Paredes Grijaldo

[View Session](#) [Add to Schedule](#)

**Affiliation:** Universidad Pontificia Javeriana, Bogota, Colombia

---

Emanuel syndrome (ES) is an unbalanced translocation syndrome, an inheritable chromosomal abnormality also known by the names of: derived syndrome 22, derived syndrome 11:22, partial trisomy 11:22 or supernumerary syndrome or der (22) t (11; 22). It originates generally from an inadequate 3: 1 segregation process in meiosis I during gametogenesis, in a phenotypically normal carrier with a balanced translocation, which causes the presence of a supernumerary derived chromosome 22. Patients with Emmanuel syndrome are characterized by a characteristic phenotype, which consists of facial dysmorphism (broad forehead, epicanal folds, inferior palpebral fissures, wide and flat nasal bridge, filter and long nasolabial groove, micrognathia) microcephaly, mental retardation severe, delay in development milestones and other reported abnormalities especially in men: congenital heart defects, renal and genital anomalies.

The diagnosis should be suspected in those patients with characteristic phenotypic findings and associated symptoms. It is established karyotypically, when it is identified in the proband the duplication of 22q10-22q11 and the duplication of 11q23-qter in a chromosome 22 derived supernumerary. Its incidence is unknown and a little more than 200 cases have been described in the world literature.

We present the case of a male patient of 14 months of age, with Emanuel syndrome, confirmed by aCGH study where a pathogenic duplication detected in the 11q23.3q25 chromosomal region of a 18.21 Mb size and a second duplication of 22q11.1q11.21 with a size of 5.1 Mb and a G banded karyotype of 25 metaphases that presents a marker chromosome, indicating a probable microduplication syndrome



## PgmNr 2572: Frequency of deletions in the iduronate-2-sulfatase gene in patients with Hunter syndrome.

### Authors:

M. Ramirez <sup>1,2</sup>; L. Figuera <sup>2</sup>; L. Rizo <sup>3</sup>; J. Garcia <sup>2</sup>; S. Mendoza <sup>2</sup>; C. Briseño <sup>1,2</sup>; M. Gallegos <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Departamento de Genética, Instituto de Genética, Universidad de Guadalajara, Jalisco, Mexico; 2) División de Genética, Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, Jalisco, México; 3) División de Biología Molecular, Centro de Investigación Biomédica de Occidente, IMSS, Guadalajara, Jalisco, México

---

**INTRODUCTION:** Hunter syndrome (HS) or mucopolysaccharidosis type II (MPS II) (OMIM # 309900) is an X-linked recessive disease caused by deficiency of iduronate-2-sulfate-sulfatase lysosomal enzyme, which leads to progressive accumulation of heparan and dermatan sulfate in different organs and tissues. It has an approximate global incidence of 0.69 to 1.19 / 100,000 live births. The gene involved is IDS were wrong sense mutations, nonsense mutations and deletions have been described. Deletions are classified by size in large and small, being 14-20% and 4-10% of patients respectively. In some patients with more severe phenotypes, total deletions or rearrangements have been described between the IDS gene and the pseudogene IDS2.

**OBJECTIVE:** To determine the deletions frequency in IDS gene and IDS2 pseudogene in patients with HS.

**MATERIAL AND METHODS:** 12 DNA samples from patients with clinical diagnosis of HS between 4 and 12 years old were used. They were processed by multiplex ligation-dependent probe amplification (MLPA) with SALSA probemix P164-B2 specific for HS and finally analyzed by Coffalyser software.

**RESULTS:** 16.6% (2/12) of the samples from patients with HS present a complete deletion of IDS gene and pseudogene IDS2.

**CONCLUSIONS:** The frequency of deletions in IDS gene and IDS2 pseudogene of 16.6% observed in these patients agrees with the previous described in the literature. However, these deletions have been associated with severe phenotypes at the neurological level, which differs in the present study when observed in patients with moderate psychomotor delay.

## PgmNr 2573: Molecular cytogenetic characterization of a karyotype of a female patient with secondary amenorrhea.

### Authors:

A. Okabe <sup>1,2</sup>; A. Nguyen <sup>1,2</sup>; A. Reyes <sup>1,2</sup>; D. Zhao <sup>1,2</sup>; D. Chung <sup>1,2</sup>; F. Laban <sup>1</sup>; L. Thompson <sup>3</sup>; M. Murphy <sup>3</sup>; K. Cunnie <sup>3</sup>; B. Wyatt <sup>3</sup>; C.A. Tirado <sup>1,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The International Circle of Genetic Studies Chapter Los Angeles, CA.; 2) UCLA; 3) Allina Health Inc, Minneapolis , MN; 4) HPA, Minneapolis , MN; 5) The University of Minnesota, Department of Laboratory Medicine and Pathology

---

Disorders of sex development (DSD) are congenital conditions characterized by atypical development of chromosomal, gonadal, and phenotypic sex. Herein we report a case of a 24 female patient who presented with secondary amenorrhea. Cytogenetic studies of her peripheral showed two abnormal cell lines. One cell-line (27 cells) had a normal copy of the X chromosome and an abnormal chromosome that needed further investigation.

FISH analysis was performed using Kallmanns, Xist, and X centromere probes. Only one signal was observed for each of these probes, and metaphase FISH analysis showed these signals were located on the one normal X. So this normal copy of the X chromosome is positive for KAL (ANOS1), DXZ1, and XIST.

Fluorescence in situ hybridization (FISH) analysis using the SRY, Yp11.1-q11 and the Yq12 probes was also performed and revealed two copies of the SRY, two copies of the heterochromatin Yq12, and two copies of the Y centromere Yp11.1-q11.1 on what resulted to be an isodicentric Y chromosome.

Therefore the karyotype was described as

45,X[3]/46,X,idic(Y)(p11.31)[27].ish idic(Y)(SRY+,DYZ3++,DYZ1++,ANOS1-,DXZ1-,XIST-).nuc  
ish(ANOS1,DXZ1)x1,(DXZ1,XIST)x1,(SRY,DYZ1)x2

Females with a 46,XY karyotype have gonadal dysgenesis and most of them are mosaic, with a 45,X cell line. Some of them showed small deletions of the short arm of the Y chromosome. Further studies based on the clinical picture, as well as possible prophylactic gonadectomy due to an increased risk of gonadal malignancy, (gonadoblastoma or dysgerminoma) are suggested. Genetic counseling was recommended.

## PgmNr 2574: PCR-free whole genome sequencing using isolates from dried blood spot.

### Authors:

Y. Ding; J. Le; S. Batalov; S. Chowdhury; L. Van Der Kraan; C. Yamada; Z. Bezares-Orin; N. Veeraraghavan; M. Niedzwiecki; D. Dimmock; C. Hobbs; S. Nahas; S. Kingsmore

[View Session](#) [Add to Schedule](#)

**Affiliation:** Rady Children's Institute for Genomic Medicine, San Diego, California.

---

**Background:** Polymerase Chain Reactions-free rapid Whole Genome Sequencing (PCR-free rWGS) using Illumina sequencing platforms has emerged as a powerful diagnostic tool within hospital intensive care units for children who are acutely ill. Current PCR-free WGS library construction kits utilize genomic DNA (gDNA) isolated from whole blood. Although dried blood spot (DBS) isolates can be used for WGS, whole genome amplification (WGA) or post-ligation PCR is required to ensure sufficient library yield for next generation sequencing. Recent publications have shown that WGA and PCR-based libraries result in downstream sequencing bias such as increased GC content. To demonstrate the application of DBS isolates to generate WGA free and PCR-free WGS, we report feasibility studies and initial results using the DBS isolates and KAPA HyperPlus kits. **Method:** DBS specimens were collected with Whatman FTA paper or Protein saver cards 903 (40 ul EDTA whole blood). Ten disks of 3x3 mm punches were lysed using proteinase K and lysis buffer. After 60 minutes incubation at 56°C, magnet KAPA Pure Beads were applied for purification purposes. DBS isolates were quantified with Picogreen assay and Nanodrop followed by 0.8% agarose electrophoresis to ensure the integrity of the DNA. KAPA HyperPlus PCR-free library construction kits (pair-end) were utilized following the manufacturer's protocol. Final libraries were measured using real time PCR assay and fragment analysis. Lastly five PCR-free libraries were pooled and sequenced on Illumina Novaseq6000 S2 flow cell with read length 101x2 format. **Results:** Around 100 ng to 600 ng gDNA were extracted from DBS without degradation from both Whatman FTA and Protein saver cards. Approximately 3 to 11 nM of PCR free WGS libraries were generated with peak insert size around 450 bp. Greater than 95% of OMIM genes had >10x coverage for all coding bases. All genome build metrics were equivalent to the data from the previously validated clinical workflow. **Discussion and conclusion:** DBS are routinely collected for newborn screening. Our approach allows researchers and clinical laboratories to utilize DBS specimens for PCR-free WGS. Additional validation will be performed to further improve sequencing outcomes at Rady Children's Institute for Genomic Medicine.

## PgmNr 2575: Diagnostic utility of whole genome sequencing in a clinically diverse, exome-negative patient population.

### Authors:

A.Y. Huang <sup>1</sup>; H. Lee <sup>2,3</sup>; L. Wang <sup>1</sup>; E.D. Douine <sup>3</sup>; A.J. Yoon <sup>3</sup>; G. Renteria <sup>3</sup>; C. Eno <sup>2</sup>; T.N. Senaratne <sup>2</sup>; S. Nieves-Rodriguez <sup>3</sup>; S.F. Nelson <sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Precision Health, University of California, Los Angeles, CA.; 2) Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, CA; 3) Human Genetics, University of California, Los Angeles, Los Angeles, CA; 4) Pediatrics, University of California, Los Angeles, Los Angeles, CA

---

Since 2012, UCLA Clinical Genomics Center has performed clinical whole-exome sequencing (WES) on over 2,200 patients with a diagnosis rate of 25%. As sequencing costs continue to decline, a critical question moving forward is to what extent whole-genome sequencing (WGS) can improve diagnostic yield over WES. Here, we present preliminary results from a pilot study to evaluate the added value of performing WGS in a diverse cohort of patients with suspected rare Mendelian diseases. A total of 133 patients have been enrolled at UCLA in the study thus far, all of which have received no molecular diagnosis from WES. A significant portion of these have also received a negative result(s) from clinical microarray (CMA) and/or gene panel testing. 30x, PCR-free WGS sequencing was performed, with 90% of patients receiving trio or quad-based family testing. In addition to analysis of all coding and non-coding SNVs and indels, comprehensive analysis of structural variants (SVs) was also performed using three different detection methods aimed at targeting SVs of different sizes and types. Out of 85 cases that thus far have received in-depth genetic review, 13 cases (15%) have been given a molecular diagnosis from WGS data. Seven (8% of the original cohort) were solved by detection of a single nucleotide variants (SNV) or small insertions/deletions (indels) within coding regions. Although WGS allowed for the complete coverage of a substantially greater number of genes overall compared to WES (97 vs 90%, respectively), no additional diagnoses resulted from the increased gene coverage of WGS. Rather, all were due to the discovery of a novel disease gene since the initial WES and would have been resolved upon reanalysis of the original exome. An additional 6 cases (7% of the original cohort) were diagnosed on the basis of SVs, most of which were single-exon deletions (average size of 1.2kb) and were well below the resolution of traditional CMA testing. In one instance, WGS detected a large 350kb pathogenic deletion that was missed by CMA due to mosaicism. Although the diagnostic yield from variants not assayable by WES is modest (6 of 85 cases, 7%), an additional significant minority of patients have a strong candidate non-coding variant of uncertain significance that may be determined to be pathogenic when augmented by mRNA sequencing. Our findings indicate that WGS could replace CMA and WES as a single genetic test, and further augment SV detection at comparable cost.

## PgmNr 2576: Clinical whole genome sequencing: The New York Genome dxperience.

**Authors:**

A. Thomas-Wilson; S. Guha; A. Rehman; A. Abhyankar; A. Fang; H. Poisner; V. Jobanputra

[View Session](#) [Add to Schedule](#)

**Affiliation:** New York Genome Center, New York, NY.

---

With the introduction of Whole Genome Sequencing (WGS), the ability to detect exonic and intronic single nucleotide, copy number, and complex structural variants in a single diagnostic test has become feasible. As cost decreases, WGS has the potential to become a first-tier test for individuals with suspected complex genetic disorders. Despite the increasing utility of WGS, few reports are available on the detection of variants which are undetectable by standard clinical diagnostic tests such as whole exome sequencing, NGS based panel testing, or microarray. Here we describe New York Genome Center's experiences in identifying these variants in cohorts of individuals with undiagnosed genetic disorders as well as healthy individuals referred for predisposition testing. In the course of our time performing more than 300 clinical WGS, we have identified causal pathogenic or likely pathogenic single nucleotide variants, mosaic variants, single and multi-exon deletions, mobile element insertions, rare deep intronic variants, mitochondrial variants, and copy number variants in disease genes associated with the individual's primary phenotype, as well as in genes recommended by the American College of Medical Genetics & Genomics (ACMG) for reporting medically actionable secondary findings. We discuss the various methods used to validate the variants identified in WGS, the challenges we have encountered in clinical interpretation, as well as our experience with cases previously reported negative by whole exome sequencing or microarray. The power of WGS to detect variants that were previously only detectable in single gene or specific targeted panel testing makes the clinical utility of WGS unparalleled in the current field of clinical diagnostics. Our data illustrates the power of WGS and its potential to redefine how we perform clinical genetic testing and what we know about the genetic basis of disease.

## PgmNr 2577: Analysis of products of conception (POC) by array-CGH and comparison to conventional USG findings.

### Authors:

D.D. Jhala <sup>1</sup>; A.R. Patel <sup>2</sup>; S.S. Chettiar <sup>2</sup>; J. Dattani <sup>3</sup>; R. Uppala <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Zoology, Biomedical Technology and Human Genetics, Gujarat University, Navarangpura, Ahmedabad, Gujarat, India. ddjhala@gmail.com; 2) Department of Genetics, Genexplore Diagnostics and Research Centre Pvt. Ltd. Ellis Bridge, Ahmedabad, Gujarat, India. info@genexplore.co.in.; 3) Consultant, NVBDCP, Ministry of Health and Family Welfare, Ahmedabad, Gujarat, India. jigudattani@gmail.co; 4) Department of Obstetrics and Gynecology, Oakland University- William Beaumont School of Medicine 3811 West 13 Mile Road, Royal Oak, MI, USA, Uppala.Radhakrishna@beaumont.edu

---

Array Comparative Genomic Hybridization (aCGH) is a preferred technique for genetic analysis of abortus samples over conventional cytogenetic analysis, as it provides high resolution detection of the submicroscopic duplications and deletions. It was already known that aCGH showed increased detection rate of copy number variations (CNV) as compared to conventional cytogenetic analysis. Moreover, Ultra sound (USG) findings during fetal development also play a major role in detection of many anomalies. Correlation of USG findings and aCGH data can lead to understanding of the role of CNVs in disease manifestation and also give phenotypic to genotypic association which can be further used as diagnostic tools. In the present study, we categorized USG findings of the Product of Conception (POC) samples in five major classes. They are (A) Cardiac anomalies, (B) Central nervous system anomalies, (C) Skeletal anomalies, (D) Hygroma related anomalies, (E) Other types of anomalies, and we correlated it with aCGH data. Array CGH was performed on oligonucleotide microarray platform (G2565CA, Microarray Scanner System, Agilent, USA). Data of the array was analyzed using cytogenomics software (CytoGenomics v4.0.3, Agilent technologies). In the present study, we analyzed 103 POC samples for the genetic analysis. Among these samples 41.75% samples showed abnormal fetus USG. whereas, 33.01% of the samples had abnormal aCGH results. Results also revealed that 30.23% CNV detection rate in samples with abnormal USG findings. Among these, hygroma related anomalies showed highest (61.5%) CNVs followed by Cardiac anomalies (23.07%). Hence, we can conclude that CNVs detection is higher in POC samples having hygroma related anomalies and cardiac anomalies detected in USG findings.

## PgmNr 2578: Advanced structural analysis of CDH risk loci with optical genome mapping technology.

### Authors:

M. Longoni <sup>1,2</sup>; F.A. High <sup>1,2</sup>; C. Coletti <sup>1</sup>; W.W. He <sup>3</sup>; P.K. Donahoe <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) PSRL, Massachusetts General Hospital, Boston, MA; 2) Harvard Medical School, Boston, MA; 3) Huiying Memorial Foundation, Washington, DC

---

Recurrent microdeletions and microduplications have been reported in over 10% of patients with Congenital Diaphragmatic Hernia (CDH), a genetically heterogeneous condition associated with high mortality and severe long-term morbidity in survivors. We designed a pilot study to compare the output of Bionano optical genome mapping technology in different sample sources and across a variety of genomic events, such as balanced and unbalanced translocations, and Kb- and Mb-size copy number variants (CNVs) in a cohort of 5 trios and 4 singletons affected with CDH.

The expected high-confidence events originally detected by standard karyotyping and/or chromosomal microarrays were validated in all the samples using the Bionano platform, regardless of the type of source material, and additional variants of unknown significance were uncovered. Among the samples assayed:

A) Translocation breakpoints were mapped in one case of an apparently balanced t(1;6)(p22;q15) translocation, which was confirmed in the proband and absent in his parents suggesting a de novo event, and similarly in an unrelated case with a t(1;15)(q44;q26.3) unbalanced translocation.

B) 1.5 Mb *NR2F2* deletion at 15q26.2 was shown to contain the clustering of two separate events: the net deletion/loss of 1.3 Mbp with a nested inversion of 192.9 Kbp.

C) 2.4 Mb gain telomeric to the 16p11.2 duplication syndrome region (OMIM #614671) was previously proposed as a risk factor for CDH. A more complex structure with two smaller duplications collectively spanning approximately 1.8 Mb was identified in a carrier in our cohort.

Optical genome mapping using Direct Label and Stain (DLS) of DNA motifs is a non-sequencing technology allowing the assembly of chromosome-arm length maps for the detection of structural and copy number variants. The results of the pilot study presented here are informative for the design of a larger study cohort of CDH-affected individuals, who have recently undergone NGS and CNV analysis, for further discovery and refinement of their genomic events.



## PgmNr 2579: Copy number variants in the ClinGen Allele Registry.

### Authors:

R.Y. Patel <sup>1</sup>; A.R. Jackson <sup>1</sup>; K. Reihle <sup>1</sup>; S. Paithankar <sup>1</sup>; T. Nelson <sup>2</sup>; L. Babb <sup>3</sup>; E. Riggs <sup>3</sup>; R. Hart <sup>4</sup>; C.L. Martin <sup>3</sup>; S.E. Plon <sup>1</sup>; A. Milosavljevic <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, Texas.; 2) Geisinger's Autism and Developmental Medicine, Lewisburg, Pennsylvania.; 3) Broad Institute, Cambridge, MA; 4) San Francisco, California, US

---

Sharing genomic variant data is critical to improving our knowledge base for both sequence and copy number variant (CNV) interpretation. The ClinGen Allele Registry ([reg.clinicalgenome.org](http://reg.clinicalgenome.org)) supports genomic data sharing by providing stable, unique, and dereferenceable identifiers for over 910 million unique sequence variants, a critical step in aggregating knowledge. Various resources use these identifiers for interoperability. Representing CNVs, however, has been a challenge, due to limitations in the standardized systems of variant nomenclature. This limitation hampers the ability to accurately aggregate information and share knowledge about CNVs. Some of the main challenges in representation of CNV data arise due to the inherent imprecision in the determination of breakpoints, imprecise and inconsistent quantification of the number of copy number gains and copy number differences on the X and Y chromosomes as they relate to X-linked disorders in males and females. To address this problem, we first reviewed existing methods for representing CNV data in key databases, including ClinVar, DGV, dbVAR, and ExAC. We also compared commonly used CNV representations with nomenclatures proposed by ISCN and HGVS. We identified discrepancies or limitations in a few areas including differences between the HGVS expressions and preferred names in ClinVar, absence of exact copy numbers gains in ExAC, lack of nomenclature in DGV, and floating point representations of copy number in dbVAR. Through our review of existing resources, we identified key information to define CNVs and have incorporated updates to the ClinGen Allele Registry to support their registration and query. One of the services offered by the registry is variant registration: for CNVs, this service returns an existing identifier if present, or if the variant is not present in the registry, the service presents overlapping and nested CNVs, helping the user decide whether to use one of the existing identifiers or create a new variant entry for the specific CNV. Other registry services include query and retrieval using various combinations of CNV attributes to provide links to existing CNV entries. We show how these registry services support CNV data sharing by aggregating information about the same CNV across current CNV databases, detecting redundancy, and creating a refined map of recurrent copy number variations based on the combined databases.



## PgmNr 2580: Frequency and characteristics of copy number variants (CNVs) identified by a whole exome sequencing platform.

### Authors:

J. Schleit <sup>1</sup>; I. Saarinen <sup>2</sup>; K. Gall <sup>1</sup>; J. Hathaway <sup>1</sup>; E. Salminen <sup>2</sup>; J. Koskenvuo <sup>2</sup>; T.-P. Alastalo <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Blueprint Genetics, San Francisco; 2) Blueprint Genetics, Helsinki Finland

### Background:

Copy number variants (CNVs) are an important cause of hereditary disease. The prevalence of CNVs across various disease categories is only beginning to be understood as next generation sequencing (NGS) technologies have become more sensitive at detecting these. However, an NGS assay's ability to detect various sized CNVs is highly dependent upon the testing platform and bioinformatics pipeline.

### Objective:

We describe the frequency and characteristics of diagnostic CNVs identified on a whole exome sequencing platform across 15 medical specialties in patients referred for clinical genetic testing. We report on the proportion of diagnostic CNVs that contribute to the overall diagnostic yield in each disease category.

**Methods:** Next-generation sequencing (NGS) was performed using the IDT xGEN Exome Research Panel with added custom probes and the Illumina NovaSeq 6000 platform. We performed a retrospective review of all genetic tests run on this platform from April 2018 to April 2019. The type of CNV (deletion or duplication), size (in bp), genomic coordinates, impacted gene(s), classification and the presence of other variants (reported as primary findings) was extracted from our internal database.

### Results:

Overall, CNVs made up 9.8% of all diagnostic variants (401/4093) and were the primary diagnostic variant in 3.7% of all tests. The majority (91.5%) of diagnostic CNVs were deletions. The median size of all deletions was 10.7kb (range 122bp-154Mb), and 25.9% of deletions were  $\leq 500$ bp. The median size of all duplications was 1.2Mb (range 374bp-72Mb) and 5.9% of duplications were  $\leq 500$ bp. Diagnostic CNVs made up the highest proportion of all diagnostic findings in whole exome tests (45.3%), pulmonology tests (33.3%) and ear-nose-throat tests (20.1%). Diagnostic CNVs made up the smallest proportion of all diagnostic findings in ophthalmology, endocrinology and cardiology (6.9%, 6.9% and 6.0% respectively).

### Conclusions:

Diagnostic CNVs are identified in approximately 10% of all tests performed on this whole exome sequencing platform. CNVs ranging from one exon deletions/duplications to multiple gene deletions/duplications were identified in multiple genes across almost all medical specialties. Importantly, 24.2% of CNVs were less than or equal to two exons in size. These results highlight the importance of a comprehensive genetic testing approach with high sensitivity to detect CNVs less than 2 exons in size across all medical specialties.

## PgmNr 2581: A systematic evaluation of copy number variation detection based on capture sequencing data.

### Authors:

Y. Sun; X. Ye; Y. Fan; L. Wang; H. Liu; X. Gao; Z. Gong; Y. Wang; W. Qiu; H. Zhang; L. Han; J. Ye; L. Liang; H. Ye; B. Xiao; X. Ji; W. Cai; X. Gu; Y. Yu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Shanghai Jiaotong University School of Medicine Xinhua Hospital, Shanghai, China

---

Capture sequencing (CS) is widely applied to detect small variations. Algorithms based on depth of coverage comparison enable CS to detect copy number variations (CNV). However, a systematic evaluation is lacking with a large sample size. We conducted a retrospective analysis of CS data to evaluate the efficiency of CS-based CNV detection in the clinical diagnosis. A total of 3010 samples referred to our diagnostic lab for the CS test were included in this study. They were divided into four subsets for using in different parts of the study: (i) to select the appropriate CNV calling algorithm, (ii) to setup a validated filtration step, and (iii) to select CNV algorithm and optimize filtration parameter to build a CS-CNV pipeline and apply it on the undiagnosed samples to improve the yield.

Concatenate CNV list fromXHMM and CNVKit shows highest sensitivity. Thus they were selected to identify the CNV. The raw sensitivity was high but the raw positive predictive value (PPV) was low. Filtration with in-house reference CNV database could increase the PPV significantly to 49.4%. The yield of CS-CNV pipeline was up to 6.8% for the undiagnosed samples. In conclusion, marked yield of CS-based CNV detection could benefit the molecular diagnosis as it enabled simultaneous evaluation of CNV and small variations.

## PgmNr 2582: Detection of copy number variations in blood typing research and ADME genes using Axiom™ assays.

### Authors:

R. Varma; K. Aull; C. Bruckner; J. Gollub; A. Roter; J. Schmidt

[View Session](#) [Add to Schedule](#)

**Affiliation:** Thermo Fisher Scientific, 3450 Central Expressway, Santa Clara, California 95051.

---

Copy number variations (CNVs) in *CYP2D6* and other genes which are relevant for drug absorption, distribution, metabolism, and excretion (ADME) are difficult to detect because these genes are members of highly homologous gene families, or because of the high prevalence of deletions and duplications in the population. Genotyping blood type variants is complicated by the frequent deletion of the *RHD* gene and its high homology with *RHCE* which together encode for Rh factor antigens. Copy number aware genotyping of markers in these regions significantly improves the accuracy of variant calling.

Results from experiments run on the Applied Biosystems™ Axiom™ PMD Array are presented here. The platform has the dual ability to detect CNVs in targeted genomic regions and to genotype SNPs across the whole genome using a single assay. CNV analysis methods include (a) whole genome de novo analysis for discovery and (b) fixed region analysis when breakpoints of CNV regions of interest are known a priori and there is little breakpoint variability from sample to sample. In CNV discovery analysis, CN states are determined by a Hidden Markov Model implementation in single sample mode. Breakpoints are discovered and CN segments are labeled by states. Fixed region analysis uses an optimized multisample clustering algorithm to assign CN states to each region in each sample.

CNVs in *RHD*, *CYP2D6*, *CYP2A6*, *GSTT1*, *GSTM1*, and *UGT2B17* were identified on replicate runs of more than 90 HapMap samples. Discovery analysis detected CNVs with overall analytical reproducibility greater than 90% for losses and 70% for gains. It was less sensitive in detecting duplications in a small 150 bp region around exon 9 of *CYP2D6* which is important in predicting metabolic capability of individuals. However, using fixed region CNV analysis, duplications and deletions were detected with 100% concordance with known truth in this region in all samples and were reproducible in all but one sample. Across all defined fixed regions of interest, homozygous deletions were detected with 100% concordance, while single copy losses were detected with 100% analytical sensitivity and over 95% analytical specificity. In summary, the platform easily allows performing the analysis with both methods, which results in superior analytical sensitivity and specificity for known small regions, while enabling discovery for larger regions and across the whole genome.

## PgmNr 2583: Analytical validation of clinical whole genome sequencing for germline disease diagnostics: Best practices and performance standards.

### Authors:

C.R. Marshall <sup>1</sup>; N.J. Lennon <sup>2</sup>; S. Chowdhury <sup>3</sup>; R.J. Taft <sup>4</sup>; D.J. Stavropoulos <sup>1</sup>; M.S. Lebo <sup>5</sup>; S.M. Harrison <sup>5</sup>; J.G. Buchan <sup>6</sup>; P. Liu <sup>7</sup>; S. Kulkarni <sup>7</sup>; D. Dimmock <sup>3</sup>; J.W. Belmont <sup>4</sup>; D. Bick <sup>8</sup>; E.A. Worthey <sup>8</sup>; R. Rowsey <sup>9</sup>; E.W. Klee <sup>9</sup>; H.M. Kearney <sup>9</sup>; Medical Genome Initiative

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Diagnostics, The Hospital for Sick Children, Toronto, ON, Canada; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Rady Children's Institute for Genomic Medicine, San Diego, CA, USA; 4) Illumina Inc., San Diego, CA, USA; 5) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA, USA; 6) Stanford Medicine Clinical Genomics Program, Stanford Health Care, Stanford, CA; 7) Baylor Genetics Laboratories, Houston, TX, USA; 8) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA; 9) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA

---

Current standard-of-care for genetically heterogeneous phenotypes often employs serial targeted gene testing, chromosomal microarray analysis and/or whole exome sequencing (WES), resulting in a lengthy time to diagnosis. Whole genome sequencing (WGS) addresses many of the technical limitations of these methodologies and can detect most forms of variation in a single test. Our group and others have demonstrated the diagnostic superiority of WGS compared with conventional testing in pediatric patients and critically ill infants.

Although WGS is poised to become a first-tier test for diagnosis of patients with rare genetic disorders, technical challenges as well as a lack of standards in both defining and deploying a clinical WGS test remain. To address these challenges, the Medical Genome Initiative (<https://medgenomeinitiative.org/>) was formed as a consortium of leading health care and research organizations in the US and Canada with an overarching mission to expand access to high quality clinical WGS. The work presented here focuses on the analytical validation of a clinical WGS test and aims to provide practical recommendations based on the consensus of group members

An analytical validation working group was formed to share common challenges faced by laboratories during the implementation of clinical WGS which led to the identification of practical solutions employed by the group. A series of discussions and surveys yielded consensus on several important topics: 1) Intended use and composition of a clinical WGS test including recommendations for the types of variation that should be reported; 2) the types of controls required for analytical validation of a clinical WGS test including reference standards and either commercially available or laboratory held positive controls for each variant type; 3) measurement of genome completeness (coverage and uniformity) to define WGS performance (callability) that can be tied to variant calling sensitivity; 4) a validation framework that includes accounting for genome complexity and variant type since both will affect test accuracy; 5) use of a comprehensive set of performance metrics for ongoing monitoring of quality and periodic use of positive controls dependent on sample volume.

These consensus recommendations may reduce the burden on new-to-WGS laboratories who wish to introduce WGS into clinical practice, and more importantly may serve to support safe and effective WGS testing for genetic disease.

## PgmNr 2584: LC-MS/MS-based glycosaminoglycan measurement: Expert considerations on methods and use in screening, diagnosis and management of patients with mucopolysaccharidoses.

### Authors:

M. Pasquali <sup>1</sup>; M. Fuller <sup>2</sup>; C. Auray-Blais <sup>3</sup>; D. Burke <sup>4</sup>; A. Burlina <sup>5</sup>; C.-K. Chuang <sup>6</sup>; C. Colón <sup>7</sup>; B. Crawford <sup>8</sup>; J. Fletcher <sup>2</sup>; M. Gelb <sup>9</sup>; S. Heales <sup>4</sup>; E. Izzo <sup>8</sup>; F. Kubaski <sup>10</sup>; G. la Marca <sup>11</sup>; R. Lawrence <sup>8</sup>; D. Matern <sup>12</sup>; C. Merry <sup>13</sup>; T. Wood <sup>14</sup>; P. Harmatz <sup>15</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Utah School of Medicine, Salt Lake City, UT, USA; 2) SA Pathology, Adelaide, Australia; 3) Université de Sherbrooke, Sherbrooke, Quebec, Canada; 4) Great Ormond Street Hospital, London, UK; 5) University Hospital, Padova, Italy; 6) Mackay Memorial Hospital, Taipei, Taiwan; 7) Hospital Clínico Universitario de Santiago de Compostela, A Coruña, Spain; 8) BioMarin Pharmaceutical Inc., Novato, CA, USA; 9) University of Washington, Seattle, WA, USA; 10) Hospital de Clinicas Porto Alegre and UFRGS, Porto Alegre, Brazil; 11) Meyer Children's Hospital, Florence, Italy; 12) Mayo Clinic, Rochester, MN, USA; 13) University of Nottingham, Nottingham, UK; 14) Greenwood Genetic Center, Greenwood, SC, USA; 15) UCSF Benioff Children's Hospital Oakland, Oakland, CA, USA

---

**Background:** Glycosaminoglycans (GAGs) are ubiquitous cell and matrix-resident complex sugars and substrates of specific lysosomal enzymes. Urine GAG excretion is characteristic to mucopolysaccharidoses (MPS) disorders. Liquid chromatography-tandem mass-spectrometry (LC-MS/MS)-based GAG methods are more sensitive and reliable to measure specific GAG elevations for MPS diagnosis, screening and patient management.

**Methods:** An international expert group of lab directors and clinicians gathered to discuss various LC-MS/MS-based GAG assays and best practices for clinical use. Results are reported.

**Results:** Following key considerations were discussed:

- 1) In MPS diagnostic work-up, LC-MS/MS-based GAG detection has greater sensitivity and identifies specific GAG elevation which may be missed by non-MS based assays. Urine is the most characterized sample matrix to detect specific GAGs. Experts recommend enzyme testing as gold standard to confirm an MPS diagnosis.
- 2) LC-MS/MS-based GAG measurement in MPS newborn screening has potential to decrease false positive rates and improve positive predictive value when introduced as a second-tier test after dried blood spot enzyme screening.
- 3) While relative GAG elevation may correlate with disease severity for the same MPS type, correlation of LC-MS/MS-based GAG levels with treatment efficacy and disease progression requires robust comparison to multiple baseline samples and demonstrate clinical utility of specific methods with diverse and substantial patient samples.

4) Current methods differ among laboratories regarding use of internal standards (IS), internal controls, results reporting, and validation sample sets. Standardization of different LC-MS/MS-based GAG methods is desirable, although challenging to achieve.

**Discussion:** To further validate, harmonize and expand clinical utility of LC-MS/MS-based GAG methods it is important to establish inter-laboratory sample exchanges and share information on IS, reagents, normalization protocols and methods comparison.

## **PgmNr 2585: Not all exomes are built the same: Comparison of exome technologies towards achieving whole genome like quality and improving diagnostics in challenging genomic regions.**

### **Authors:**

R. Barnard <sup>1</sup>; J. Yuan <sup>1</sup>; G. Creadon-Swindell <sup>1</sup>; K. Schneider <sup>2</sup>; R. Layer <sup>2</sup>; A. Gaskell <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Precision Diagnostics, Dept of Pathology and Laboratory Medicine, Children's Hospital of Colorado, Aurora, CO.; 2) Department of Computational Sciences, University of Colorado Boulder, Boulder CO.

---

With the decrease in sequencing costs and increased potential for diagnostic yield over small panels, whole exome sequencing (WES) is rapidly being offered in the clinical testing space. However, there has not been a thorough evaluation of how different capture technologies and library preparations affect the quality of WES data. Here we show that not all exomes are the same. We selected the 4 most common exome panels, comprising 3 different technologies and combinatorially tested them with 4 different library preparation technologies. We used 3 clinical samples with challenging variants (a large CNV, an exon deletion, and an exon duplication) along with the gold standard NA12878 genome to produce a total of 96 exomes. As whole genome sequencing (WGS) data has reduced sequencing bias compared to targeted capture like WES, we also generated WGS data on the 3 clinical samples for comparison. For each exome, we considered several metrics concerning sequencing quality and variant discovery potential to determine which combination approached the same levels as WGS of the same sample. We also determined precision and recall comparing against the NIST NA12878 defined variants. To provide additional clinical relevance, we determined the performance of each technology in specific regions of interest with different sequence contexts and clinical features. These genomic loci included sites with high clinical significance, i.e. many known pathogenic variants and patient sample CNVs, technically challenging regions such as high GC content or segmental duplications, and less understood but likely important regions based on evolutionary conservation and constraint. In each of these regions we determined the overall error rate, read mismatch propensity, coverage uniformity, and compared across exome preparations and WGS data. With these results, we provide clinical laboratories and the molecular diagnostic community with information to make the most informed decision. We also highlight the importance of assessing performance in challenging genomic regions for optimizing variant discovery. By improving variant identification, including CNVs, we can enhance our understanding of the clinical significance of these loci and enable more accurate interpretation.



## **PgmNr 2586: Three-years' accomplishment of Initiative on Rare and Undiagnosed Diseases (IRUD): A nation-wide project in Japan for discovering causes, mechanisms and cures.**

### **Authors:**

Y. Takahashi <sup>1</sup>; H. Date <sup>1</sup>; H. Oi <sup>2</sup>; S. Kosugi <sup>3</sup>; N. Matsumoto <sup>4</sup>; K. Kosaki <sup>5</sup>; Y. Matsubara <sup>6</sup>; H. Mizusawa <sup>7</sup>; IRUD. Consortium <sup>8</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan; 2) Translational Medical Center, National Center of Neurology and Psychiatry, Tokyo, Japan; 3) Department of Medical Ethics and Medical Genetics, Kyoto University School of Public Health, Kyoto, Japan; 4) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 5) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 6) National Center for Child Health and Development, Tokyo, Japan; 7) National Center of Neurology and Psychiatry, Tokyo, Japan; 8) Agency for Medical Research and Development. Japan

---

### **Background**

IRUD was launched in 2015 as a nation-wide project in Japan to construct a comprehensive medical and research system for establishing diagnosis, discovering causes and ultimately providing cures for rare and undiagnosed diseases, supported by Agency for Medical Research and Development (AMED). IRUD covers the entire geographic areas and the whole fields of specialties, taking advantage of the health care system in Japan. IRUD has been motivated by a global trend to tackle undiagnosed diseases through international collaborative networks, leveraging the phenotypic and genetic data by extensive data-sharing to solve so-called 'N-of-1' problems.

### **Purpose**

To demonstrate the accomplishment of the three-years' effort of IRUD project illustrating the landscape of rare and undiagnosed diseases in Japan.

### **Methods**

IRUD consists of the four pillars: a coordinating center (IRUD-CC) to orchestrate the whole system and strengthen the governance, clinical centers?IRUD-C?to operate diagnostic committees which take charge of the process from the decision for recruitment to the establishment of final diagnosis, analysis centers?IRUD-AC? to conduct whole exome sequencing analysis and a data center (IRUD-DC) to implement data-sharing platform, IRUD Exchange. Phenotypic data described in Human Phenotype Ontology and candidate genes are registered in IRUD Exchange, which architecture is compatible to international data-sharing platform, MatchMaker Exchange. Resources including DNA samples and cell lines are deposited in a resource center (IRUD-RC) and prepared for further utilization.

### **Results**

Thirty-seven clinical centers, 5 analysis centers, 1 data center and 1 coordinating center have been established. Until the end of July 2018, 3356 pedigrees including 9524 individuals have been enrolled

in IRUD. WES has been completed in 2756 pedigrees and final diagnosis has been established in 1027 pedigrees (36.9%). Novel genes or new disease entities have been discovered in 18 pedigrees through extensive data-sharing.

#### Conclusion

IRUD has realized a nation-wide comprehensive diagnosis network for rare and undiagnosed diseases and succeeded in establishing diagnosis, identifying novel causative genes and new disease entities through extensive data-sharing and accomplishing resource repository. Extensive international collaboration would play an essential role in elucidating causes and ultimately providing cures for such rare and undiagnosed diseases.

## PgmNr 2587: Findings from the Diamond Blackfan Anaemia cohort in the 100K Genomes Project.

**Authors:**

M. Alikian; G.D. Rose; T. Ng; D. Iskander; Y. Harrington; E. Thomas; A. Karadimitris; J. De La Fuente

[View Session](#) [Add to Schedule](#)

**Affiliation:** Imperial College Healthcare Trust, UK

---

**Scope:** DBA is characterized by dominant inheritance, incomplete penetrance, anaemia and multiple congenital abnormalities where single gene tastings is able to diagnose only up to 25% of the cases.

**Methods:** The Hammersmith Hospital in the UK is the referral centre for DBA with a registry of over 200 families tested locally with a custom targeted panel comprising of 83 ribosomal genes. The centre has participated in the 100K Genomes Project with a cohort of 35 DBA families for Whole Genome Sequencing (WGS).

**Results:** together, the targeted NGS and WGS have identified 34 pathogenic variants in known genes, 44 variants of unknown clinical significance (VUS) in known genes, 47 VUSs in novel genes and 26 pathogenic CNVs. In a number of cases through WGS, the diagnosis of DBA was eliminated due to the identification of characteristic pathogenic variants of other disease leading to case referral to the appropriate clinic.

**Conclusion:** Using Whole Genome Sequencing (WGS) or at least Whole Exome Sequencing (WES) we were able to identify additional cases of DBA that have otherwise been difficult. These techniques are therefore the best way to diagnose heterogeneous disease such as DBA where the outcome of the test can not only explain the basis for the disease, but also assist in the accurate classification facilitating the appropriate treatment.

## PgmNr 2588: Considering potential secondary/incidental findings associated with off-target disease from targeted NGS panels.

### Authors:

M. Luo <sup>1,2</sup>; M. Dulik <sup>1,2</sup>; J. Murrell <sup>1,2</sup>; L. Conlin <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

---

Targeted NGS panels can either focus on a well-defined phenotype or be quite broad in scope. While recommendations exist on the use of secondary findings detected by genome-wide testing, a different type of secondary or incidental finding can arise in panel testing, as some panel genes can be associated with additional, unrelated disorders outside of the scope of the panel test. In this study, we evaluated our panel-based tests for potential unrelated findings due to an off-target disease. We performed an evidence-based curation of 592 genes present on 39 panels offered by our laboratory. These panels cover 19 unique disease groups ranging from well-defined disorders with few genes, to broader phenotype categories, such as hearing loss. We categorized each gene based on their associated disorders, inheritance patterns, and the relation of the associated phenotypes to the intended purpose of the panel. The genes were divided into three categories: Category 1 (n=461) are associated with a single disorder or phenotypic spectrum with consistent inheritance patterns; Category 2 (n=77) are associated with multiple disorders with at least one additional phenotype outside of the scope of the panel (e.g. *GJB2*); Category 3 (n=54) are associated with multiple disorders, with at least one disorder completely unrelated to the scope of the panel (e.g. *MARS*). Category 3 genes are the most worrisome, as the associated phenotypes can be completely unrelated to the reason for testing, and some conditions can be adult onset. We identified category 3 genes in 26 panels, covering 13 disease groups, with hearing loss panel having the most genes (16). So far, in >1800 panel tests, we have not identified an additional finding, suggesting the incidence is rare; however, as the panel test design and analysis strategy takes into consideration disease-specific variant types, inheritance, and prevalence, the off-target use of genes on panels is likely to miss pathogenic variants associated with secondary/incidental findings. Given the potential for unrelated findings, we propose that all potential outcomes should be considered when designing targeted NGS panels. Testing limitations should be included in the report, and upfront decisions should be made on how to deal with additional findings, and how to report them. Informed consent may be warranted and off-target use of the panel is generally not recommended due to the design limitations.

## PgmNr 2589: Development and validation of a 2-step PCR library preparation method for the detection of *GBA* variants by next generation sequencing.

### Authors:

R. Owen; S.H. Rosenthal; A. Gerasimova; A. Smolgovsky; A. Buller-Burckle; F. Lacbawan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Quest Diagnostics, San Juan Capistrano, CA

---

**Background:** The *GBA* gene encodes glucocerebrosidase, a lysosomal enzyme that breaks down glucocerebroside into ceramide and glucose. Genetic variants in *GBA* can cause the autosomal recessive disorder Gaucher disease, the most common lysosomal storage disorder, and are strong risk factors for Parkinson's disease. *GBA* analysis is complicated by the presence of a pseudogene, *GBAP1*, located 16 kb downstream of the *GBA* gene. Sanger sequencing of gene specific PCR amplicons is the most commonly used method for variant detection. However, the method is not suitable to consolidation with other analytical methods, such as next generation sequencing (NGS) when used in carrier screening for inherited disorders. Here, we aimed to design and validate a new 2-step PCR NGS library preparation method for the sequencing of *GBA* using an Illumina Miseq.

**Method:** The 8 most common *GBA* variants included for detection were: p.N409S (N370S), c.84dupG (84GG), p.R535H (R496H), p.L483P (L444P), c.115+1G>A (IVS2+1G>A), p.V433L (V394L), p.D448H (D409H), and c.1263\_1317del (del55bp). The first PCR used conventional long-range PCR to specifically amplify targets in *GBA* but not in *GBAP1*. The second PCR used nested PCR with primers containing Illumina sequencing primer binding sites which enabled addition of specimen specific barcoded sequencing adapters via concomitant fusion PCR. Sequencing ultimately was performed on the Illumina MiSeq platform. Compatibility of the method within an expanded NGS carrier screening panel was evaluated in combination with 18 other inherited diseases. **Results:** Method validation using 159 samples (94 positives [17 *GBA* positives], 65 negatives) showed 100% sensitivity and specificity. Subsequent testing using 11,778 consecutive clinical samples submitted for the validated carrier screening test showed that 2.0% (232/11,778) of cases harbored a *GBA* variant: 174 cases positive for N370S, 29 cases with L444P, 9 cases with R496H, 6 cases with 84GG, 6 cases with D409H, 4 cases with V394L, 3 cases with IVS2+1G>A, and 1 case with del55bp. **Conclusion:** This study demonstrated that the 2-step PCR NGS library construction method successfully screens for common *GBA* variants and can be readily consolidated for use in a comprehensive carrier screening panel.

## PgmNr 2590: Clinical validation and implementation of whole genome sequencing in molecular diagnostics of Mendelian disorders.

### Authors:

P. Liu <sup>1,2</sup>; B. Yuan <sup>1,2</sup>; J. Sinson <sup>2</sup>; V. Nguyen <sup>2</sup>; N. Dhaliwal <sup>2</sup>; C.J. Williams <sup>2</sup>; B. Halen <sup>2</sup>; M. Santana <sup>2</sup>; C.-Y. Chen <sup>2</sup>; R. Dominguez-Vidana <sup>1,2</sup>; P. Ward <sup>1,2</sup>; P. Thaduri <sup>2</sup>; M. Kaiparettu <sup>2</sup>; V. Perez <sup>2</sup>; Y. Feng <sup>2</sup>; S. Leake <sup>2</sup>; C.-C. Yeh <sup>2</sup>; V. Patel <sup>2</sup>; C. Qu <sup>2</sup>; I. Machol <sup>1</sup>; F. Gibellini <sup>2</sup>; C. Shaw <sup>1,2</sup>; C.M. Eng <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Mol & Hum Gen, Baylor College Med, Houston, Texas.; 2) Baylor Genetics, Houston, Texas.

---

Whole genome sequencing (WGS) has been widely used in research studies, and evidence is accumulating to support its use in clinical diagnostics. A comprehensive validation is warranted to assess the standards, efficacy, and limitations of WGS in the clinical setting. Such an effort requires performing clinical sequencing and analysis in a CLIA and CAP certified environment using a large number of samples with diverse relevant mutations independently identified and confirmed with a separate technology.

We have performed a validation study to evaluate the detection of SNV/INDEL, CNV, SV, and UPD using a unique collection of samples with various types of mutations validated by assays such as WES, CMA and cytogenetics. Our goal was to characterize WGS in terms of variant detection of SNV, SV, CNV and other variant types. An additional goal was to characterize the impact of genome build on results.

Samples are sequenced by Illumina NovaSeq 6000 with a PCR-free protocol with minimum of 40X average coverage and 2X150bp read length with 550bp average insert size. Data are processed with the Illumina Dragen pipeline mapping to the GRCh38 reference, with haplotype, depth based CNV and Manta split read SV callers.

For our evaluations we selected the GIAB sample NA12878 as well as a large set of internal clinical samples with known pathogenic variants. High analytical precision (>99.97%) and recall (>99.84%) were achieved for SNV. INDEL detection also demonstrates high quality metrics with correlation to the INDEL sizes. For CNV, we selected >70 samples carrying clinically reported CNVs as well as manually curated benign or unknown significance CNVs, adding up to >200 individual events spanning losses or gains of hundreds of base pairs to megabases. The read depth method alone demonstrated 100% sensitivity for events larger than 3kb. For smaller events, joint analysis of the read depth and the Manta results increased the sensitivity. For SV and AOH/UPD, the analysis detected clinically reported inversions, translocations, insertions, hetero- or iso- disomies, and AOH in >25 samples. We also examined mosaic variations with positive results. In terms of regions of low sequencing coverage and sequencing depth uniformity, WGS provided data of higher quality than WES as expected.

This validation demonstrates that WGS applied to clinical molecular diagnostics can detect diverse mutation types with a comparable or superior robustness to traditional methods.

## PgmNr 2591: Targeted long-read sequencing of complex human duplicated regions.

### Authors:

A. Sekar<sup>1,2</sup>; D. Soto<sup>1,2</sup>; R. Sahasrabudhe<sup>3</sup>; G. Kaya<sup>1</sup>; M. Dennis<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Center, University of California, Davis, Davis, CA.; 2) Integrative Genetics and Genomics Graduate Group, University of California, Davis, CA, USA; 3) DNA Technologies Sequencing Core Facility, University of California, Davis, CA 95616

---

Human-specific segmental duplications (“HSDs”; genomic regions >1 kbp in size with >98% identity), or loci existing exclusively in the human lineage, have the potential to contribute to traits and diseases unique to our species. Unfortunately, due to the high identity of these regions, they are technically challenging to study. Recent comparisons of great ape whole-genome shotgun sequences identified over 30 HSD gene families uniquely duplicated, including two genes previously characterized as potential drivers of human neocortex expansion (*SRGAP2C* and *ARHGAP11B* ).

To evaluate the potential evolutionary/disease impact of HSD genes, we surveyed genetic variants from human population controls using existing datasets and found less than 2% and 5% of HSD regions contain confidently mapping reads (1000 Genomes Project and Genomes in a Bottle (GIAB), respectively). As such, 78% of HSDs lack any common variants (dbSNP). To address this gap in variant discovery, we have established a high-confidence benchmark variant callset using BAC sequences from a haploid hydatidiform mole cell line (CHM1) across nearly 30 Mbp of the genome, including 22% of all annotated HSDs. We confirmed that Illumina CHM1 sequencing (>30X) underperforms in variant calling within HSDs due to inaccurate mapping of short reads. Sequencing CHM1 using Oxford Nanopore Promethion (>30X ) found that long reads map accurately within HSDs but high sequence error rates result in faulty variant calling. To address this, we combined target enrichment of select HSDs (>2 Mbp, 48 genes) using the new PacBio Sequel II sequencing platform. This approach uses single-molecule circular consensus correction to generate accurate long reads useful for variant calling. As proof of principle, we performed targeted sequencing of eight individuals (two haploid hydatidiform mole cell lines and two GIAB trios) generating, on average, 903 Mbp of sequences per sample with median polymerase read length 41.7 kbp and median circular consensus length 3.7 kb (14X coverage per 3 kbp fragment). Work is ongoing to confirm accurate variant calling. Once established, we will expand this approach to sequence HSDs in a larger cohort of modern humans to confidently survey genetic variants and shed light on the evolutionary impact of these complex regions within the human genome.

## PgmNr 2592: Assessment of sensitivity and specificity in a population genetic screen to identify medically relevant variation.

### Authors:

K.M. Bowling <sup>1</sup>; M.L. Thompson <sup>1</sup>; D.E. Gray <sup>1</sup>; K.T. Williams <sup>1</sup>; K.M. East <sup>1</sup>; A. Cannon <sup>2</sup>; W.V. Kelley <sup>1</sup>; I.P. Moss <sup>2</sup>; J.C. Edberg <sup>2</sup>; A.C.E. Hurst <sup>2</sup>; E.C. Partridge <sup>1</sup>; D.M. Absher <sup>1</sup>; G.S. Barsh <sup>1</sup>; B.R. Korf <sup>2</sup>; G.M. Cooper <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama; 2) University of Alabama at Birmingham, Birmingham, Alabama

---

The Alabama Genomic Health Initiative (AGHI) is a state-wide effort to conduct genetic testing for ~10,000 individuals not ascertained for any phenotype. We are using an Illumina Global Screening Array (GSA) to detect pathogenic/likely pathogenic (P/LP) variation in medically actionable disease genes; predominantly associated with cancer/cardiac risk, that overlap largely the 59 actionable genes defined by the American College of Medical Genetics and Genomics.

To date, 4,355 participants have been analyzed. 63 (1.4%) harbor a P/LP variant deemed suitable for return, comprising 48 unique variants across 17 genes. P/LP variants were validated by CAP/CLIA-certified Sanger before return to participants.

Array-based genotyping is more efficient and cost-effective than sequencing, but does have limitations in terms of sensitivity and specificity. We found that ~60% of secondary P/LP variants found from sequencing ~1,000 healthy individuals are present on the array, indicative of a sensitivity upper limit. To assess sensitivity of P/LP variants on the array, we tested 24 individuals known to harbor a medically actionable variant targeted by the GSA; all expected variants were identified, suggesting high sensitivity to array-targeted variants.

We also assessed specificity. Hundreds of potential P/LP variants on the array are seen more frequently in AGHI than reported in databases, or is plausible given associated disease prevalence, suggesting false positives (FPs). After filtration to remove obvious FPs, the remaining P/LP variants in actionable genes were manually curated and Sanger tested. We found that 41 unique P/LP variants, in 39 individuals, detected by GSA were FPs; added to the 48 validated P/LP variants (above), these data indicate a false discovery rate among array-identified P/LP variants of ~46%, even after extensive automated filtration. In 28 of 39 individuals, Sanger detected benign missense/synonymous variation at, or nearby, the tested position, indicating that the array detected a non-targeted variant. For 11 individuals, no nearby alteration was identified by Sanger; reason for array detection is unknown.

Through AGHI, we have implemented an array-based process to screen a population of Alabamians to cost-effectively identify highly penetrant genetic variants in actionable disease genes. We demonstrate the need for clinical follow-up of array-identified variants, especially when considered in the context of direct-to-consumer genetic testing.



## **PgmNr 2593: Assessing clinical validity of genes of unknown significance in patients with undiagnosed diseases utilizing ClinGen's Clinical Validity Framework: Experience in the NIH Undiagnosed Diseases Program.**

### **Authors:**

C.C. Lau; E. MacNamara; B.N. Pusey; N. Balanda; A. Brown; S.M. Jafri; M.C. Malicdan; C. Toro; C.J. Tifft; W.A. Gahl; D.R. Adams

[View Session](#) [Add to Schedule](#)

**Affiliation:** NIH Undiagnosed Disease Program, National Institutes of Health, Bethesda, Maryland.

---

The NIH Undiagnosed Diseases Program (UDP), a member site of the Undiagnosed Diseases Network (UDN), enrolls patients with diseases that remain undiagnosed despite extensive diagnostic evaluation and clinical testing. Clinical exome or genome sequence analysis is frequently non-diagnostic for our patients, therefore, the major work in the diagnostic odyssey for our patients is the assessment of candidate variants in genes of unknown significance (GUS). The major hurdle in this process is evaluating the level of evidence and establishing the clinical validity of these GUS in the context of our patients' clinical phenotypes. The UDP invests extensive resources to tackle this problem; one approach included reanalysis of 234 probands and their nuclear families, which was carried out over the span of 3 years from May 2015 to May 2018. Their variants of unknown significance (VUS) were triaged based on 3 principles: segregation with disease phenotypes, low population frequency consistent with a rare disease, and multifactorial prediction of molecular impact(s). The GUS are then further prioritized based on 3 broad evidence categories that either support or disqualify gene-disease relationship: clinical, genetics and experimental. The major outcome of this process is the prioritization of variants most likely to be the underpinning cause of the patients' diseases, and more importantly, designation of these variants to: (i) those deemed to be the diagnosis; (ii) those likely to be the diagnosis pending functional validation and/or accumulation of experimental evidence; and (iii) those needing additional evidence such as additional cases with similar genotypes and phenotypes for further actions. Over 750 candidate variants in 251 GUS were prioritized using this method, and to date, with the accumulation of evidence, 34 cases have reached a definitive diagnosis. To objectively evaluate the level of evidence supporting the gene-disease associations in these cases, a study was carried out using the ClinGen evidence based clinical validity framework. Gene-disease associations were scored through intervals before and after the diagnoses were made. Overall, the utilization of ClinGen's framework provides an objective assessment for the degree of evidence needed to reach a diagnosis for our patients, and how well the clinical validity of gene-disease associations has stood over time.

## PgmNr 2594: X chromosome haplotypization in the group of Turner syndrome patients.

### Authors:

P. Vrtel <sup>1</sup>; R. Vrtel <sup>1</sup>; K. Adamova <sup>1</sup>; D. Vrbicka <sup>1</sup>; J. Stellmachova <sup>1</sup>; E. Klaskova <sup>2</sup>; J. Pavlicek <sup>3</sup>; V. Hana <sup>4</sup>; V. Stara <sup>5</sup>; O. Soucek <sup>5</sup>; J. Lebl <sup>5</sup>; M. Prochazka <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Clinical Genetics, Palacky University, Faculty of Medicine and Dentistry, University Hospital Olomouc, Olomouc, Czech Republic; 2) Department of Pediatrics, Palacky University Olomouc, Faculty of Medicine and Dentistry, University Hospital Olomouc; 3) Clinic of Pediatrics, The University Hospital in Ostrava; 4) 3rd Department of Medicine - Department of Endocrinology and Metabolism, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague; 5) Department of Pediatrics 2nd Faculty of Medicine, Charles University in Prague and University Hospital Motol

---

Monosomy of chromosome X - Turner syndrome is the only monosomy postnatally compatible with life. The incidence of this chromosomal abnormality is 1 : 2 500 live female births. These patients are phenotypically characterized by lower stature compared with the general population, fertility disorders and other deviations from the physiological female genotype. Life expectancy could be affected by heart disorders that are present in almost 50 % of these females.

The origin of heart abnormalities and other phenotypical defects of these patients may correlate with maternal or paternal haplotype of chromosome X. We used molecular genetic analysis for detection microsatellite DNA markers on chromosome X for reveal of origin of these gonosome in Turner's syndrome patient's cases.

In the initial part of our study we selected cohort of 20 Turner syndrome patients with cytogenetically confirmed X monosomy and their parents. Results of another 22 families will be presented on the congress.

The biological material for DNA isolation was the peripheral blood of the patients and the cells of the buccal mucosa of the parents. We used the Argus X-12QS forensic kit for the X-chromosome analysis, which was optimized for our genetic testing. The method is based on detection of 12 STR markers from chromosome X. The variability of these STR markers allow to establish parental origin of haplotype.

Preliminary results suggest that 65 % of X chromosomes are of maternal origin and 35 % cases of paternal origin. We plan to extend the group of tested families and the next part of our project will be a specification of phenotypic deviations that will be compared with the origin of gonosome.

**Supported by MH CZ - DRO (FNOI, 00098892), grant nr. 87-26 and grant nr. 17-29111A**

## PgmNr 2595: Copy number variation and chromosomal aneuploidy in embryo biopsy of infertile couple.

### Authors:

J. Dattani <sup>1</sup>; S. Chettiar <sup>2</sup>; A. Patel <sup>2</sup>; D. Jhala <sup>3</sup>; B. Dey <sup>4</sup>; C. Dey <sup>1</sup>; U. Radhakrishna <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) RoHFW, Ministry of Health and Family Welfare, Ahmedabad, Gujarat, India; 2) Genexplore Diagnostics and Research Centre, Ahmedabad, Gujarat; 3) Department of Zoology, Gujarat University, Ahmedabad; 4) Shri D M Patil Medical College, Vijayapura, Karanataka, Inidia; 5) Departmetn of Obstetrics and Gynaecology, William Beaumont School of Medicine, Detroit, USA

---

**Introduction:** Assisted reproductive technology (ART) *techniques* have become common practice in many countries today and has supported numerous infertile couples worldwide. Although, ART has evolved technologically in last few years resulting in increased success rate and low cost. Of all aspects of ART viz, gamete retrieval, in vitro fertilization, selection of embryo, cryopreservation, vitrification and embryo transfer, selection of embryo has significant impact on ART outcome. Advance genomic technologies like next generation sequencing and microarray can aid clinicians in selecting embryos with normal euploid and genomic content. Here, we present our experience with 1200 embryos from 280 females.

**Material and Method:** Subjects included in the study were both women and men who were then sub-grouped based on four major reasons viz. 1) advanced maternal age, 2) more than three IVF failures, 3) low sperm count and 4) genetic indications. Blastomere or Trophectoderm biopsy were performed in 1200 embryos from 280 females undergoing IVF treatment. These single cell (blastomere) or four to six cells (trophectoderm) biopsy were subjected to whole genome amplification (WGA) using a picoplex WGA kit (Takara). Amplified product was then taken up for customized microarray and NGS based chromosomal aneuploidy analysis where a lower mark of 10MB of deletion duplication was taken for reporting and a mosacisim of minimum 25% was reported.

**Results:** Combined NGS and microarray data revealed 272 (22.66%) embryo had segmental or whole chromosomal aneuploidy. From the pooled data of all aneuploid embryos, it is evident that chromosome 16 was involved in 4.33% of events and chromosome 8 showed the least (1.26%) incidence of deletion and duplication. Among the incidence of aneuploidy heterozygous deletion was a significantly frequent (18.7%) event as compared to homozygous deletion (8.42%), homozygous duplication (9.65%) and heterozygous duplication (12.4%).

**Conclusion:** A significant number of embryos showed chromosomal and/or segmental aneuploidy, thus requiring the genetic screening of potential embryos in couples undergoing IVF treatment. Chromosome 16 has maximum risk of deletion or duplication as compared to other chromosomes.

## PgmNr 2596: Unique dual indexing primers and universal blockers enable multiplexed target enrichment applications.

### Authors:

K. Butcher; L. Arbiza; S. Chen; H. Chilton; R. Gantt; S. Gude; Y. Hasin-Brumshtein; C. Thompson; R. Zeitoun

[View Session](#) [Add to Schedule](#)

**Affiliation:** Research and Development - Molecular Biology, Twist Bioscience, San Francisco, CA

---

Target enrichment has been an enabling approach to sequencing specific regions of the genome and calling variants while reducing sequencing cost per sample. The ability to create customizable panels has been a key driver for increased utilization of target enrichment within various workflows and applications.

However, two important components, (1) unique dual-indexed (UDI) Y-adapters for library creation and (2) specific adapter blockers used to improve on-target rates during target enrichment, have seen few recent advancements and are fast becoming limiting factors in multiplexed target enrichment applications. Synthesis of UDI Y-adapters is challenging due to their relative long length (>65bp), 5'-phosphate chemical modification, and purity requirements that demand HPLC purification of two oligonucleotides for every desired index set. They also commonly generate significant amounts of undesired adapter dimer during ligation reactions. Some 'universal' adapter blockers are commercially available, but their application is limited to subsets of index lengths and artificially constrain experimental design.

To address these limitations, fundamental changes to (1) installation of UDI sequences during library creation and (2) adapter blockers were investigated and provided numerous benefits. Application of the Twist Universal Adapter System reduced undesired adapter dimer formation below the lower limit of detection under common reaction conditions, increased total gDNA library yield 2X with a 20% reduction of PCR amplification cycles, and enabled facile switching between index designs. In addition, facile synthesis allows for access to thousands of UDI pairs so the number of samples and amount of data generated during sequencing can be maximized. To complement this robust adapter system, Twist Universal Blockers eliminate any requirement for specific adapter blockers. This design functioned independently of index length, increased capture of on-target reads when compared to sequence-specific adapter blockers on a molar basis, was agnostic to target enrichment panel size, and remained functionality equivalent in singleplex and multiplex workflows.

Taken together, these advancements in the application of shared adapters for UDI sequence installation and universal adapter blockers eliminate numerous bottlenecks for target enrichment multiplexing applications.

# PgmNr 2597: Analytical validation of SNP arrays for clinical diagnostics: Multi-platform framework.

## Authors:

P.F. Cherukuri<sup>1,2,3</sup>; D.E. Condon<sup>1,2</sup>; M. Soe<sup>1</sup>; L. Fricke<sup>1</sup>; S. Bartaria<sup>1</sup>; K. Meis<sup>1</sup>; J. Lubieniecka<sup>1,2,3</sup>; K. Lubieniecki<sup>1,2,3</sup>; S. Shaaban<sup>1,2</sup>; D. Lyalin<sup>1,2</sup>; R. Pyatt<sup>1,2,3</sup>; C.F. Boerkoel<sup>1</sup>; L. Carmichael<sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Sanford Imagenetics, Sioux Falls, SD; 2) Sanford School of Medicine, University of South Dakota, Sioux Falls, SD; 3) Sanford Research Center, Sioux Falls, SD

## Background

Clinical interpretation of genotype data requires high positive predictive value (PPV) and understanding of genotyping platform performance to minimize genotyping error. Single nucleotide variant (SNV) genotyping arrays, like the Global Screening Array (GSA), offer high-throughput, low-cost genotyping for actionable single nucleotide variants. Consequently, we hypothesize that analytical validation of GSA enables precision medicine.

## Methods

To validate the GSA, we generated whole genome sequence (WGS) data (30x, GenomeNext Churchill pipeline) for 263 1000 Genomes (1KG) project samples from Coriell. 236 of these were genotyped in triplicate with the GSA. Publicly available genotype and WGS data were downloaded and paired with GSA data. All WGS data were archived on Amazon S3. Following exclusion of low quality sample data and following stratification of the data by nucleotide type, change, allele frequency (AF), and DNA complexity at locus, we analyzed or are analyzing the GSA data performance metrics relative to the 1KG and the WGS data using two approaches: (1) GSA vs. 1KG, and (2) GSA vs. WGS. A random subset of discordant variants discovered by these analyses will be assessed using Sanger sequencing and performance of GSA relative to WGS and Sanger sequencing will be calculated.

## Results

Based on above analyses, we observed high concordance with the 1KG data across 612,341 GSA assays in the 236 sample triplicate dataset (avg. concordance range [0.991 – 0.994]; std-dev range [0 - 0.001]). Additionally, we observed the following for a single replicate dataset: PPV (0.989 ±0.0009), sensitivity (0.993 ±0.0006) and specificity (0.996 ±0.0003). Limiting the data to high-quality samples (call rate > 0.98) and high genotype call rate assays (assay call rate >0.997; 602,371 assays) improved the average overall concordance (0.9936 vs. 0.9937) and sensitivity (0.9929 vs. 0.993). Stratifying SNVs by various genomic features showed a reduced PPV if the alternate allele is rare (AF < 1%) and a reduction in all performance metrics if the SNV mapped to low-complexity regions. We also present additional performance metrics and discuss their role in analytical validation.

## Conclusion

Our systematic evaluation shows that SNV genotyping arrays, such as the GSA, can achieve analytical validity. This framework for analytical validation of genotyping arrays is a step toward their clinical use.

## **PgmNr 2598: Bionano Prep SP isolates high quality ultra-high molecular weight (UHMW) genomic DNA to improve research of cancer and undiagnosed disorders.**

### **Authors:**

H. Sadowski; C. Proskow; A. Files; K. Pham; D. Carroll; H. Barnes; Y. Zhang; G. Pljevaljcic; M. Borodkin

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bionano Genomics, San Diego, California, USA

---

Optical mapping of genomic DNA on the Bionano Genomics Saphyr® system for genome assembly or structural variation detection relies on starting with UHMW DNA. To achieve this, protocols using agarose plugs have been the gold standard. The “plug lysis” method is extremely robust, but it is also labor intensive, difficult to automate, lengthy and expensive. To address these shortcomings, we coupled solution-based lysis with a purification step that leverages a novel process to bind, wash and elute UHMW genomic DNA. This entire protocol can be conducted in less than 4 hours on a batch of 6 samples, allowing 12 samples to be processed in one day. The eluted material is ready to use by day 2 and contains high quality DNA that is clean enough for the direct label and stain (DLS) protocol. The resulting single molecule quality metrics of this labeled DNA on a Saphyr Chip® are comparable or exceed the metrics for labeled DNA isolated by the traditional plug lysis protocol. We have previously validated protocols for fresh/frozen human blood (EDTA) and cultured cells and released the Bionano Prep SP kit that provides virtually all of the required reagents and consumables. For many samples, the addition of DNA stabilizer has allowed us to achieve average labeled DNA centers of mass (COM) N50s of  $\geq 280$  kbp, with a significant proportion of Mbp-sized molecules. Unlike plug lysis, the SP protocols are clearly automatable, providing a solution for researchers needing to purify DNA from hundreds to thousands of individuals per year.

We have continued to expand the number of validated SP protocols for new sample types including fresh/frozen heparinized human blood and bone marrow aspirates (BMA)s. The new protocols also include; the addition of DNA stabilizer to the heparinized bloods/BMAs before they are frozen, to increase the size of isolated DNA (COM N50s  $\geq 300$  kbp), and cell straining upon thawing, to avoid clots and insoluble aggregates before cell counting. We have also developed working SP protocols for small amounts (5-15 mg) of animal and human tissues including tumor biopsies. These SP protocols are significantly shorter and less labor intensive than our previously released plug lysis protocols, while QC metrics of the SP isolated genomic DNA are comparable or exceed the metrics of genomic DNA isolated by the plug lysis protocols.

## PgmNr 2599: Monitoring the impact of pre-analytical parameters on cfDNA quality.

### Authors:

L. Friedmann <sup>1</sup>; A. Berger <sup>1</sup>; E. Graf <sup>1</sup>; S. Siembieda <sup>2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) R&D and Marketing GmbH & Co. KG, Agilent Technologies, Waldbronn, Germany; 2) Agilent Technologies Inc, 2450 SE Oak Tree Ct, Ankeny, IA 50021, USA

---

Cell-free DNA (cfDNA) has become an important source for potential new biomarkers. However, it represents a challenging sample type due to its low yield and complex fragment size distribution. Pre-analytical parameters have a major impact on the quality of cfDNA samples. Controlling and monitoring pre-analytical parameters, including cfDNA extraction, is crucial for experimental success. A new automated electrophoresis assay was used to compare nine different kits tailored for cfDNA extraction. Three independent plasma samples were processed with each kit in parallel and subsequently analyzed with the new electrophoretic assay in six replicates. The assay offers total DNA quantification, and automatically assigns a region for cfDNA separately from high molecular weight DNA by size. Using region analysis, the concentration of cfDNA apart from high molecular weight (HMW) DNA is evaluated by the software. The quality score %cfDNA is provided as additional quality parameter, determining the percent cfDNA of total DNA in presence of HWM. This set of quality parameters adjusted to the needs of cfDNA characterization was applied to compare the kits. The extraction kits varied in rate of yield and their size selection capacity to remove HMW DNA. In general, each kit afforded either superior yield or high selectivity towards cfDNA. High yield may be crucial for plasma with low cfDNA content, yet if the plasma is expected to contain considerable amounts of HMW DNA, the selectivity towards cfDNA might be favored. Therefore, the choice for the most suitable cfDNA extraction kit depends on the requirements for downstream analyses as well as the expected cfDNA concentration and occurrence of HMW DNA. Even if no preference for a single kit could be established, the results present a straightforward way to monitor the impact of pre-analytical parameters on cfDNA concentration and quality.



# PgmNr 2600: Uncovering the precise genomic location of inversion events using optical mapping.

## Authors:

C.M. Grochowski<sup>1</sup>; C. Gonzaga-Jauregui<sup>2</sup>; J. Eisfeldt<sup>3</sup>; A. Lindstrand<sup>3</sup>; P. Stankiewicz<sup>1</sup>; J.R. Lupski<sup>1,4,5,6</sup>; C.M.B Carvalho<sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Molecular and Human Genetics, Baylor College Medicine, Houston, TX, United States; 2) Regeneron Genetics Center, Tarrytown, NY, United States; 3) Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, United States; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, United States; 6) Texas Children's Hospital, Houston, TX, United States.

---

Genomic inversions are a class of DNA structural variation (SV) that usually presents with two breakpoints that occur in *cis* and a subsequent 180-degree longitudinal turn of DNA. Copy-number neutral inversions are challenging to resolve using classical methods for SV detection as there may be no immediately obvious genomic alteration to infer orientation of the copy number neutral event. Inversions are often flanked by inverted repeats and have breakpoints embedded within repeat regions further complicating analysis. Detection methods for inversions, such as fosmid clone sequencing (PMID: 18451855, 17901297) and paired-end (PE) sequencing (15895083), have limited ability to resolve/map breakpoint junctions of inversions larger than 100 kb in size. Optical mapping (OM) involves tagging sequence motifs along unbroken DNA strands >100 kb in length preserving the architecture of SVs harboring more than one breakpoint junction *incis*. Given the mounting evidence for inversion rearrangements contributing to disease pathogenesis we sought to compare different SV detection methods on a well characterized control genome.

We performed OM on a female lymphoblastoid cell line (NA15510) using BioNano direct label stain (DSL) technology. Given that inversions can be mediated by repeats, we incorporated a custom database, Highly Similar Intrachromosomal Repeats (HSIRs), allowing for the visualization of inverted repeat regions genome-wide. To allow for a direct comparison of OM versus both fosmid and PE sequencing the genomic location of inversions detected by these methods were also considered as part of our analysis. In total, 90 inversion events were detected by OM and/or other methods. Of the 46 inversion events detected by OM, 36% were not detected by other methods while 15% were detected by at least one other genomic sequencing-based technique. Only 1% of inversions were detected by both OM and all other methods analyzed. Only 19% of inversions studied involved HSIRs while 35% involved segmental duplications. The precise genomic location of all previously documented inversions was further refined due to the uncertainty inherent to previous methods versus that of OM.

The visualization of long lengths of unbroken DNA through OM is critical in properly studying an SV with two breakpoints in *cis*. As the role of SVs in human disease becomes more clear new methods like OM may be required to study a structural aberration in its totality.



## PgmNr 2601: Phenotype-driven prioritization of exome data in a clinical laboratory.

### Authors:

K. Dumas <sup>1</sup>; D. Fisk <sup>1</sup>; M. Grove <sup>1</sup>; C. Reavey <sup>1</sup>; S. White <sup>1</sup>; J. Buchan <sup>1,2</sup>; E. Spiteri <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanford Health Care, Stanford, California; 2) Stanford University, Stanford, California

---

Next generation sequencing technologies provide a means to detect vast quantities of rare variation in clinical subjects. With this opportunity comes the challenge of distinguishing reportable, clinically-relevant variation from thousands of non-reportable variants. Clinical laboratories have developed a variety of methods for prioritizing variants for manual review, including the use of patient phenotype information to create targeted gene lists that prioritize variants within genes potentially related to patient disease phenotypes. Using patient medical records and HPO terms listed by ordering providers, Stanford Medicine Clinical Genomics Program (CGP) generates a unique, phenotype-driven, target gene list (TGL) for patients referred to CGP for clinical exome analysis. This phenotype-driven prioritization approach is run upstream of broader variant prioritization strategies to maximize the chances of identifying clinically-relevant variation. Here, we retrospectively investigate the value of a phenotype-driven variant prioritization strategy on a cohort of proband-only and trio samples referred to CGP for clinical exome sequencing and analysis.

Gene, variant and prioritization data were collected from each case, including: number of genes on each patient's TGL, number of variants prioritized through the TGL, total number of prioritized variants, report result, and reported variant info. We investigated whether reported variants were identified via target gene strategies or broader strategies, whether variants identified by TGLs would have been identified by broader strategies had TGLs not been used, and which filters were most useful to prioritize variants when TGLs were not used.

Of the variants reported to clinicians, a majority were identified via the TGL, despite comprising a small quantity of the total prioritized variant list. For positive cases, about half of the reported variants were identified via the TGL. We observed that nearly all of our reported variants would have been identified by broader variant prioritization strategies had we not created a TGL. When TGLs were not used, inheritance and canonical loss of function filters were most useful in identifying reportable variants. Variants not detected in the absence of a TGL included rare, not previously reported, missense variation. These results demonstrate additional disease-gene association resources may help to improve phenotype-driven analysis.

## PgmNr 2602: Clinically ambiguous cases screened by NGS: VUS interpretation in the light of the clinical manifestation.

### Authors:

M. Glushkova<sup>1</sup>; S. Ategin<sup>1,2</sup>; T. Todorov<sup>1</sup>; A. Kirov<sup>1</sup>; P. Dimova<sup>3</sup>; H. Kathom<sup>4</sup>; L. Pacheva<sup>5</sup>; A. Maver<sup>6</sup>; B. Peterlin<sup>6</sup>; A. Todorova<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human genetics, GMDL Genica1. Genetic Medico-Diagnostic Laboratory Genica, Sofia, Bulgaria, Sofia, Bulgaria; 2) Department of Medical Chemistry and Biochemistry, Medical University, Sofia, Bulgaria; 3) Department of Neurosurgery, University Hospital "Saint Ivan Rilski", Sofia, Bulgaria; 4) Department of Clinical Genetics, University Pediatric Hospital Sofia, Sofia, Bulgaria; 5) Department of Pediatrics and Medical Genetics, Medical University - Plovdiv, Plovdiv, Bulgaria; 6) Clinical Institute of Medical Genetics, UMC Ljubljana, Šljamerjeva 4, SI-1000 Ljubljana, Slovenia

Here we report 3 clinically ambiguous Bulgarian cases screened for genetic verification by NGS. Case 1: 2 years old boy with horizontal nystagmus, muscular hypotonia, retinopathy, neck hyperkinesia without teeth and epilepsy at the moment of genetic testing. Compound heterozygous variants in the *ROGDI* gene were detected: c.432+1\_432+2insG, likely pathogenic and c.727C>G, VUS. Pathogenic homozygous and compound heterozygous variants in this gene represent an established cause of Kohlschütter-Tonz syndrome, epilepsy, psychomotor regression and amelogenesis imperfecta, however development may be normal until the onset of seizures in infancy. Both variants are not detected among the healthy control population. Based on the previous reports of other pathogenic loss-of-function variants and the compatibility with the symptoms the variants are classified as probably pathogenic and disease causing.

Case 2: 1 year old girl with cerebellar atrophy, neonatal hypotonia and seizures. The NGS data indicated two heterozygous missense variants in the *GFM1* gene: c.881C>T and c.6607A>G, classified as VUS. Pathogenic biallelic variants in this gene represent an established cause for Combined oxidative phosphorylation deficiency 1 with generalized brain atrophy, hypotonia and seizures as it is in our patient. The mutations in this gene are associated with neonatal liver failure and lactic acidosis which are not manifested in our case. The variants are novel and we cannot conclude in favor of their pathogenicity.

Case 3: 3 years old boy with autism, neonatal epilepsy, encephalopathy and muscle hypotonia. A heterozygous variant was detected in the *CLCN1* gene c.1437\_1450del, reported to be associated with Myotonia congenita AR. The variant was detected in asymptomatic father. Later, the NGS data were reevaluated and two heterozygous missense variants were detected in the *SYNE1* gene: c.24569A>C and c.17650C>T, classified as VUS. Pathogenic variants in this gene were reported in association with Emery-Dreifuss muscular dystrophy 4 AD and Spinocerebellar ataxia AR8. Genetic variants in the *SYNE1* gene have been reported in patients with autism.

The detected variants in these 3 families segregated in terms of the expected mode of inheritance. The question is how positive we could be that the genetic diagnosis is clarified in these cases? Can we consider these genes as the only and definite cause of the diseases in the tested patients?

## PgmNr 2603: Broad genomic screening of active-duty Air Force members: Findings from the MilSeq study.

### Authors:

K. Machini <sup>1, 2, 3</sup>; H. Mason-Suares <sup>1, 2, 3</sup>; M.D. Maxwell <sup>3, 4, 5, 6</sup>; S. Bandyadka <sup>1</sup>; S. Pereira <sup>7</sup>; C. Blout <sup>4</sup>; J. Krier <sup>3, 4</sup>; B. Zettler <sup>4</sup>; C. Gardner <sup>6</sup>; J.M. Killian <sup>6</sup>; A.L. McGuire <sup>7</sup>; M. De Castro <sup>8</sup>; R.C. Green <sup>1, 3, 4, 9</sup>; M.S. Lebo <sup>1, 2, 3, 9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Partners Healthcare Personalized Medicine, Cambridge, MA; 2) Department of Pathology, Brigham & Women's Hospital, Boston, MA; 3) Harvard Medical School, Boston, MA; 4) Department of Medicine, Brigham and Women's Hospital, Boston, MA; 5) University Health System, San Antonio, TX; 6) Joint Base San Antonio - Lackland Air Force Base, San Antonio, TX; 7) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 8) Air Force Medical Genetics Center, 81 MDG. Keesler AFB, MS; 9) The Broad Institute of MIT and Harvard, Cambridge, MA

### Introduction:

Within the framework of the MilSeq Project, a pilot study examining the process of incorporating genomic testing into the United States Air Force Military Health System, 75 active-duty participants underwent genomic screening via exome sequencing.

### Methods:

Five main categories of genomic findings were returned to participants: 1) monogenic and 2) carrier status findings from >5000 genes, 3) 18 established or likely risk alleles with odds ratios >3 for 8 common diseases, 4) 230 pharmacogenomic (PGx) variants with CPIC and PharmGKB recommendations affecting the metabolism of >12 drugs, and 5) blood and platelet antigen typing results. Variants from categories 1 and 2 were assessed according to the 2015 ACMG/AMP guidelines, and pathogenic or likely pathogenic variants in genes with a strong level of evidence for association with disease were reported. Risk variants were assessed according to a novel framework developed by the laboratory and were an expansion in report content from our previous MedSeq and BabySeq studies.

### Results:

We identified monogenic findings in 12% (9/75) of participants and carrier status in 87% (65/75; average of 2 and range of 0-6 variants), consistent with previous studies. In 9 individuals, no variants associated with monogenic or carrier status findings were detected. We also identified risk alleles for 4 diseases in 37 participants, 29 of which had 1 reportable risk allele and 8 of which had 2 risk alleles. The most commonly identified risk allele was heterozygosity for the *APOE* e4 variant for Alzheimer's disease (22 participants), followed by heterozygosity for *MUC5B* c.-3133G>T for pulmonary fibrosis variant (17 participants). Only 1 participant was heterozygous for the frequent Factor V Leiden variant, which is associated with a risk for venous thromboembolism. Warfarin genotyping was provided for every individual due to incorporation of non-genetic factors in dosing guidelines. Additional PGx diplotypes prompting a potential change in drug management were identified in 92% of participants (average of 4 and range of 0-8 drugs). The most frequently affected drugs were PEG-IFN-alpha/ribavirin and tamoxifen.

#### Conclusions:

The findings from the genomic screening within the MilSeq project highlight the broad range – and potentially high impact – of genetic information that can be attained through systematic analysis of genomic sequencing data, particularly in a more diverse population.

## PgmNr 2604: Affordable trio-genome sequencing as first tier genetic test in critically ill infants: Practice in China.

### Authors:

H. Wang <sup>1</sup>; Y. Lu <sup>1</sup>; X. Dong <sup>1</sup>; G. Lu <sup>2</sup>; G. Cheng <sup>3</sup>; Y. Qian <sup>1</sup>; Q. Ni <sup>1</sup>; P. Zhang <sup>1</sup>; L. Yang <sup>1</sup>; B. Wu. <sup>1</sup>; W. Zhou <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Molecular Medicine, Children's Hospital of Fudan University, Shanghai, Shanghai, China; 2) Pediatric intensive care unit, Children's Hospital of Fudan University, Shanghai 201102, China; 3) Department of Neonates, Key Laboratory of Neonatal Diseases, Ministry of Health, Children's Hospital of Fudan University, Shanghai 201102, China

---

**Background** Genome sequencing was feasible to provide genetic diagnoses in critically ill infants with rapid turn around time (TAT). Herein, to delineate the value of genetic diagnosis, we provide the results from 130 pediatric patients in the hospital in China.

**Methods** This study was performed with an affordable trio genome sequencing (ATGS) test. The sequencing depth of patients were 40-50 X, and of their parents were 8-10 X. Patients from PICU/NICU with complicated clinical features were enrolled from June 2018 to Dec. 2018, each with phenotypes suggesting of an underlying genetic disorder.

**Results** The ATGS testing identified diagnostic variants in 47.7% (62/130) individuals. The TAT was 3-5 work days and the cost was 2500\$ per family. Of the sixty-two infants with diagnoses, 48 (77.4%) patients were found with pathogenic single-nucleotide variants (SNVs), 12 (19.4%) patients were detected with pathogenic copy number variations (CNVs) or structure variants (SVs), 2 (3.2%) patients were found small deletions plus pathogenic variants in another allele of autosomal recessive gene. Therapeutic strategies of 48.4% (30/62) diagnosed patients were modified, including transplantation, dietary recommendation or change of drugs, which obviously avoided morbidity and improved prognosis.

**Conclusions** This study provided a powerful capacity of ATGS testing in detecting of SNVs and chromosome abnormalities with fast response, higher diagnostic yield and lower cost. ATGS shows the potential to be the first tier genetic test in critically ill infants in developmental country.

## PgmNr 2605: Detection of structural variations in monozygotic twins using Nanopore long-read sequencing technology.

### Authors:

H. Mishima <sup>1</sup>; S. Yoshida <sup>2</sup>; Y. Morimoto <sup>2</sup>; S. Ono <sup>3</sup>; A. Kinoshita <sup>1</sup>; A. Imamura <sup>2</sup>; H. Ozawa <sup>2</sup>; K.-I. Yoshiura <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 2) Department of Neuropsychiatry, Unit of Translation Medicine, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; 3) Eijinkai Aino Ariake Hospital, Nagasaki, Japan

---

**Purpose:** Recently Oxford Nanopore Technologies (ONT) released the MinION sequencer and its multiplex model the GridION sequencer. Although ONT's single molecule long-read sequencing technology has disadvantages in its lower base-call accuracy and high false-positive ratio in short insertion-deletions (indels) detection compared to Illumina's short-read technology, it is expected to have advantages in detecting genomic structural variations (SVs), including large indels, inversion, translocation, copy-number variation and tandem-repeat variation. The aim of this report is to assess and evaluate ONT whole genome sequencing data quality to detect SVs.

**Materials and methods:** We collected a pair of monozygotic twins and an unrelated control. Genomes from twins were previously analyzed serially by using G-banding, DNA microarrays, and whole-exome sequencing and concluded having only small genetic difference. All participants provided informed consent to undergo genetic research being approved by concerned ethics committees.

**Results:** We performed genome-wide sequencing by GridION on these three participants. Mean coverage depth was >15x, >15x, and >10x, respectively. To detect SVs for following validation, we integrated combinations of long-read genome mapping and SV detection software packages: NGM-LR + Sniffles, LAST + NanoSV, and LAST + Picky. We picked up called SVs which are shared in the twins and not shared in the control. We successfully and validated polymorphic SVs which had never been detected in our previous studies.

**Conclusions:** We performed genome-wide sequencing of human genomes using ONT long-read sequencing. Integrating SV detection workflows successfully found polymorphic SVs. Our report discussed only specificity of the method. Thus, future directions of study could discuss sensitivity of the method.

## PgmNr 2606: Complexities of exome reporting practices: Balancing the emphasis of phenotype and genotype.

### Authors:

L.B. Lulis<sup>1</sup>; E.T. DeChene<sup>1</sup>; I.D. Krantz<sup>2,3</sup>; S.E. Raible<sup>2,3</sup>; K. Izumi<sup>2,3</sup>; C.M. Skraban<sup>2,3</sup>; C. Gray<sup>2</sup>; H.D. Hartung<sup>4</sup>; J.M. Kalish<sup>3</sup>; E.R. Hathaway<sup>3</sup>; E.H. Zackai<sup>3</sup>; B. Keena<sup>3</sup>; B.L. Krock<sup>5</sup>; J.R. Murrell<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genomic Diagnostics, Department of Pathology & Laboratory Medicine, The Children's Hospital of Philadelphia; 2) Roberts Individualized Medical Genetics Center, Department of Pediatrics, The Children's Hospital of Philadelphia; 3) Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia; 4) Division of Hematology, Department of Pediatrics, The Children's Hospital of Philadelphia; 5) Caris Life Sciences, 4610 South 44th Place, Phoenix, AZ 85040

---

Clinical diagnostic criteria have been used as a gold standard for clinical diagnosis in order to guide medical management and reproductive counseling. However, this approach may artificially define the phenotypes for a clinical entity based on the limited ascertainment of patient cohorts and a bias towards including patients who exhibit the most consistent phenotypic pattern. With the advent of robust molecular diagnostic tools, including exome sequencing, the definition of various disorders is now evolving, sometimes contracting, but often times expanding to include additional findings and a range of severity. It can be particularly demanding to strike a balance between phenotype driven and genotype driven analysis, as sometimes the molecular findings, patient features, and the understanding of the disease spectrum are not consistent. Approximately 8-10% of reports issued by our lab disclose a novel variant with strong molecular evidence for pathogenicity in an individual with limited or absent features of the characteristic phenotype. These may be due to 1) age-dependent penetrance, 2) gender-specific presentation, 3) nonsyndromic presentations of classically-reported syndromic disorders, or 4) mosaicism. For example, we detected a *de novo* *NSD1* missense variant of uncertain clinical significance. Heterozygous pathogenic variants in *NSD1* cause autosomal dominant Sotos syndrome with most pathogenic variants occurring within the SET domain. Despite the variant location within the SET domain, the proband did not exhibit the typical overgrowth but did express expected facial features and developmental delay. While the *NSD1* case *may* represent the potential to report phenotypic expansion or early presentation of a disorder, we must consider the risk to cause undue worry or harm from reporting a variant that may *not* cause the expected disorder, such as those associated with specific screening protocols or later-onset conditions. For example, an inherited loss-of-function variant in *NF2* was initially reported as pathogenic. After detailed evaluation and imaging of the proband and carrier parent revealed normal results, the classification of this variant was downgraded. By extrapolating from the lessons learned in these cases, we propose the implementation of an evidence-based approach to the reporting of these variants in an attempt to increase consistency, reduce time spent in decision-making, and improve the patient's and family's experience.

## PgmNr 2607: Challenges in returning actionable genomic results: The Mayo Return of Actionable Variants Empirical (RAVE) study.

### Authors:

D. Kochan <sup>1</sup>; E. Winkler <sup>2</sup>; N. Lindor <sup>3</sup>; G. Shaibi <sup>4</sup>; J. Olson <sup>5</sup>; P. Caraballo <sup>6</sup>; R. Freimuth <sup>7</sup>; J. Pacyna <sup>8</sup>; C. Radecki-Breitkopf <sup>8</sup>; R. Sharp <sup>8</sup>; I. Kullo <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cardiovascular Medicine, Mayo Clinic, Rochester, Minnesota.; 2) Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota.; 3) Department of Health Sciences Research, Mayo Clinic, Scottsdale, Arizona.; 4) Center for Health Promotion and Disease Prevention, Arizona State University, Phoenix, Arizona.; 5) Department of Health Sciences Research - Epidemiology, Mayo Clinic, Rochester, Minnesota.; 6) Department of Internal Medicine, Mayo Clinic, Rochester, Minnesota.; 7) Department of Health Sciences Research - Digital Health Sciences, Mayo Clinic, Rochester, Minnesota.; 8) Department of Health Sciences Research - Biomedical Ethics Program, Mayo Clinic, Rochester, MN.

---

**Background:** To inform best practices for returning actionable results to participants in genome sequencing studies, we describe challenges encountered and lessons learned from the Mayo Return of Actionable Variants Empiric (RAVE) study.

**Methods:** Participants (n = 2535, mean age 63±7, 57% female) ascertained based on hypercholesterolemia and/or colon polyps were sequenced for actionable genes (n=68) and single nucleotide variants (n=14). To participate, individuals had to consent to receive results related to hypercholesterolemia and colorectal cancer, but could opt out of receiving secondary findings. Participants with actionable results (n=121) were invited to schedule an appointment with a genetic counselor (GC). Neutral results (n=2414) were disclosed via mail. Results were placed in the electronic health record (EHR) following disclosure.

**Results:** Of those with actionable results, two opted out of receiving secondary findings and one participant died; 84 of 118 returnable results were returned in-person by a GC and 12 were returned via phone. Challenges in RoR could be categorized as: a) *sequencing report-phenotype mismatch*. These included three reports that were revised after review of the clinical record, and two reports requiring corrections. No gender mismatch was detected in those with actionable results (four were detected in those with neutral results). Three actionable results were probably mosaic requiring alternative DNA samples for confirmation. b) *contacting participants*: 16 participants were non-responders after four mailed letters. Of these five had previously identified clinically and were already documented in the EHR. In all 23 actionable results were already documented in the EHR. c) *EHR integration of results*. Ten results were disclosed to care providers prior to inviting the participant for a GC appointment, following an EHR workflow error. Five reports were reclassified due to modified informatics pipelines, four of which required participant re-contact and correction of the sequencing result in the EHR.

**Conclusion:** With the increasing number of large scale genome sequencing projects, there is a need



for greater awareness of potential challenges in RoR. In our study, such challenges were encountered in a significant proportion (24%) of participants.

## PgmNr 2608: Game of exomes: Comparing sequencing of commercial laboratories in the Undiagnosed Diseases Network.

### Authors:

H. Dai <sup>1,2</sup>; D. Murdock <sup>1,2</sup>; L. Burrage <sup>1,2</sup>; J. Rosenfeld <sup>1,2</sup>; B. Lee <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, Texas.; 2) Undiagnosed Diseases Network

---

The Undiagnosed Diseases Network (UDN) uses diagnostic tools including exome and genome sequencing (ES/GS) to solve complex medical cases. With the greater uptake of genetic testing, many patients have already undergone clinical ES before UDN evaluation. We sought to compare ES quality metrics among three major commercial laboratories alongside GS obtained through the UDN.

For patients with prior ES accepted to the Baylor College of Medicine UDN clinical site, we obtained and reprocessed sequencing data using a common analysis pipeline. We generated coverage and uniformity statistics for different gene classes including RefSeq, OMIM disease, and ACMG Secondary Findings (SF) genes. We paid close attention to first coding exons as well as non-coding regions such as introns and UTRs. We examined coverage at known pathogenic non-coding variant sites from ClinVar and the UDN. For comparison, we analyzed 50X GS data obtained through the UDN alongside these cases using the same methods. Importantly, we only addressed the quality of the sequencing, not the downstream processing or interpretation for clinical reporting.

We analyzed a total of 112 ES and 20 GS cases sequenced between 2013 and 2018. Among the three laboratories there was a significant difference in average coverage across all RefSeq genes (143X vs 116X vs 113X,  $P=0.00015$ ), a disparity that persisted ( $P=0.00049$ ) even in recent years (2016-2018). This difference was even greater for OMIM disease and ACMG SF genes, suggesting an enrichment for these clinically-relevant genes in some ES designs. Uniformity of coverage differed as well; some laboratories provided better sequencing homogeneity overall and in specific genes or regions (e.g., first exons, UTRs). While GS provided lower average coverage ( $\sim 50X$ ) compared to ES, it resulted in superior uniformity across all genes and regions. On average, 30.6/140 (21.8%) known pathogenic non-coding variant sites were missed by all laboratories due to low coverage ( $\leq 5X$ ), in contrast to GS where all sites were well covered.

Clinical ES remains an important tool in the diagnosis of patients with complex genetic cases. Our analysis suggests there are measurable differences in the sequencing quality of different commercial laboratories, in part related to the underlying ES design. Furthermore, with improving understanding of the role of non-coding variation in disease mechanisms, GS should be increasingly considered in the context of gene discovery and investigation.

# PgmNr 2610: Robust targeted sequencing approach for low input and variable quality DNA.

## Authors:

F. Mafra <sup>1</sup>; J. Garifallou <sup>1</sup>; L. Kulikoski <sup>2</sup>; A. Dias <sup>2</sup>; D. Li <sup>1</sup>; J. Smiler <sup>1</sup>; N. Williams <sup>1</sup>; A. Caro <sup>1</sup>; C. Vaccaro <sup>1</sup>; H. Hakonarson <sup>1,3</sup>; R. Pellegrino <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Center for Applied Genomics, The Childrens Hospital of Philadelphia, Philadelphia, Pennsylvania.; 2) USP, Universidade de Sao Paulo, Pathology Division.; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

**Introduction:** Sequencing DNA extracted from dry blood spots and formalin-fixed and paraffin-embedded specimens (FFPE) can frequently provide limited amount of degraded, poor quality DNA. The inconsistent input volume and DNA damage, such as inter and intra strand crosslinks, accumulation of single-stranded DNA, and single-strand DNA breaks, prevalent in these types of samples can both negatively affect sequencing results. The effects can range from simple library failures to generating libraries that produce unauthentic data, leading to misinterpretation of the results. To overcome these issues, we evaluated the performance of the Twist Library Preparation Kit in poor quality samples.

**Methods:** We tested a total of fifty-two challenging samples. Amongst the cohort, there were two poor quality gDNA (saliva), four FFPE, and forty-six dry blood spots. Samples were manually processed to test the Twist Library Preparation Kit that used enzymatic fragmentation and Y-shaped adapters to generate dual indexed libraries, followed by target enrichment using Twist capture probes and hybridization reagents. Sequencing was performed on an Illumina HiSeq 2500 sequencer in Rapid Run mode and the bioinformatics analysis was completed using Illumina DRAGEN Bio-IT Software. We targeted 60X mean coverage for the two poor quality gDNA and four FFPE samples using the Twist Human Core Exome kit. In addition, we targeted 200X mean coverage for the forty-six dry blood spots using an in-house designed Twist custom epilepsy panel.

**Results:** Captured libraries were successfully generated according to the Center for Applied Genomics (CAG) quality standards/metrics for all the samples in question. The assay performance was evaluated by checking the quality of sequencing (passing filters), duplicate rates, on target, and coverage at bases on 20x; particularly by looking at the number of reads and the amount of data generated and by determining the presence of good coverage and high uniformity across the curated genes of interest.

**Conclusions:** The kit tested was able to overcome challenges in DNA purification, including low yields, poor quality, and difficult sample types. Following the initial tests, the Center for Applied Genomics NGS unit established the Twist Core Exome pipeline, and we are now optimizing the procedures on PerkinElmer Sciclone instruments for high-throughput use of the kit. Further improvements are still in development.

## PgmNr 2611: Is likely pathogenic really 90% likely? A look at the data.

### Authors:

S. Harrison <sup>1</sup>; H. Rehm <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA

---

The 2015 ACMG/AMP guideline for variant interpretation proposed that the term likely pathogenic (LP) be used to mean greater than 90% certainty of a variant being pathogenic. However, no study has analyzed the outcomes of LP variants to determine if the 90% certainty threshold is valid. Understanding LP classification confidence is necessary as many clinicians treat LP and P classifications equally, meaning reclassifications to VUS/LB/B are likely unanticipated. We sought to determine the true certainty threshold by calculating the reclassification rates of variants submitted to ClinVar, choosing variants assessed after Jan 2016 in hopes of restricting to variants classified with the 2015 ACMG/AMP guideline.

Between Jan 2016 and May 2019, there were 552,134 classifications submitted to ClinVar. By May 2019, 4,445 of these classifications had been reclassified, of which 92.5% moved to a classification category of more certainty and only 7.5% moved to a less certain (7.2%) or opposing (0.3%) category.

Of the five classification terms, LP variants had the highest reclassification rate (2.1%; 739/35731); comparatively, benign classifications had the lowest reclassification rate (0.03%; 17/56273). Of the 739 LP reclassifications, 5 were reclassified to LB/B, 145 reclassified to VUS, and 589 reclassified to P. If only including LP reclassifications to a more definitive category (LP to P or LP to LB/B), LP reclassification rates suggest a 99% (589/594) certainty of being P compared to B. However, 0.4% (145/35731) of LPs dropped to VUS suggesting that some may eventually move to LB/B and the rate may be lower than 99%.

Variants were further interrogated to determine if certain variant types were more likely to be upgraded (LP to P) or downgraded (LP to VUS/LB/B). We found that 48% (281) of LP variants reclassified to P were LOFs and 52% (308) were missense or intronic. In contrast, 18% (18) of LP variants downgraded to VUS/LB/B were LOFs versus 82% (81) were missense or intronic.

In summary, current reclassification data from ClinVar shows that 99% of LP classifications move to P compared to LB/B, suggesting that application of the term is consistent with the intended confidence level. However, the vast majority of LPs still remain as LP within a three year window and a small subset (0.4%) dropped to VUS suggesting that more data and a longer time of analysis will be needed to more robustly evaluate the rate of LP reclassification.

## PgmNr 2612: Whole genome sequencing versus standard-of-care genetic testing in patients with a suspected genetic disorder.

### Authors:

D. Brockman <sup>1</sup>; C. Austin-Tse <sup>1</sup>; R. Pelletier <sup>1</sup>; C. Harley <sup>1</sup>; H. Head <sup>1</sup>; C. Patterson <sup>2</sup>; P. Natarajan <sup>1</sup>; K. Aragam <sup>1</sup>; M. Udler <sup>1</sup>; S. Kathiresan <sup>1,2</sup>; H. Rehm <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 2) Broad Institute of MIT and Harvard, Boston, MA

---

Current standard-of-care (SOC) diagnostic genetic testing typically relies upon high-coverage sequencing of a disease-specific set of candidate genes. This approach has two key limitations: first, it requires that a set of genes be prespecified for each disease area, limiting the ability to reanalyze after new genetic discovery data becomes available; and second, it requires provider awareness and commercial availability of numerous testing options. Clinical whole genome sequencing (cWGS) has the potential to allow use of a single genetic test regardless of disease area. Here, we set out to explore two key unanswered questions: first, does cWGS -- which provides lower coverage across the entire genome -- have adequate technical sensitivity to identify pathogenic variants across a spectrum of disease areas? Second, does the increased scope of cWGS permit identification of variants in genes not on the SOC gene panel that might nonetheless contribute to the patient's observed phenotype? We explored these questions among patients in whom a SOC diagnostic genetic test was being ordered within genetic clinics at Massachusetts General Hospital (MGH). Of 76 patients who underwent both SOC and cWGS, a pathogenic or likely pathogenic variant felt to explain the patient's phenotype was identified by SOC testing in 9 (12%) of the enrolled patients. Each of these 9 variants was also identified by cWGS (technical sensitivity 100%). Moreover, cWGS identified a potentially causal variant in 2 additional patients -- one in whom a variant detected by both SOC and cWGS was classified as likely pathogenic by our team but a variant of uncertain significance by the SOC test, and a second in whom a likely pathogenic variant was identified in a gene not covered in the SOC candidate gene panel. cWGS did not identify pathogenic or likely pathogenic secondary findings in the 59 genes recommended for return by the American College of Medical Genetics and Genomics (ACMG) in any of the 76 participants. In summary, these results suggest that cWGS may enable genetic diagnoses across a full range of patient phenotypes using a single test -- providing adequate technical sensitivity and allowing for analysis beyond a prespecified set of genes on a SOC testing panel.

## PgmNr 2613: Assessing for extreme polygenic risk burden in patients with presumed monogenic disease.

### Authors:

J. Krier <sup>1</sup>; D. Hui <sup>2</sup>; N. Patsopoulos <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine, Brigham & Women's Hospital, Boston, Massachusetts.; 2) Neurology, Brigham & Women's Hospital, Boston, Massachusetts.

---

**Introduction.** Many research initiatives, such as the Undiagnosed Disease Network (UDN), undertake exploratory analyses of genomic sequencing data to identify novel monogenic etiologies for rare disease. Key questions typically unexplored by these analyses, however, include whether the patient's phenotype can be explained by an extreme polygenic risk burden of a related common disease and whether extreme polygenic risk burden to a common disorder can impact the expressivity of an underlying monogenic disorder. Here, we utilize genetic findings of well-characterized common diseases to identify patients with extreme polygenic risk burdens in a cohort of patients with suspected monogenic disease. **Methods.** Patients enrolled in Brigham Genomic Medicine/UDN Harvard Clinical Site received whole-genome sequencing for the affected proband and unaffected relatives. We generated genetic risk scores (GRSs) across multiple phenotypes and specifically focused on each proband's most analogous common phenotype (e.g. multiple sclerosis for CNS white matter disease) based on statistically significant variants from recent GWAS meta-analyses. We further leveraged genome-wide data to calculate polygenic risk scores (PRS). GRS and PRS were corrected for the underlying ancestry leveraging reference population data. **Results.** Generally, the affected probands and their unaffected relatives did not have high values either GRS or PRS for the respective common diseases (z-score < 2 for any comparison), suggesting that picking analogous diseases is error-prone in such complex patients and/or that there may be an undiscovered monogenic or oligogenic etiology for probands. However, samples with high GRS or PRS for common diseases often had phenotypic characteristics suggestive of the respective disease. For example, two probands and families with early onset multiple sclerosis (MS) had GRS and PRS near the underlying ancestral population mean, z-scores < |1|. When we surveyed the entire cohort for MS polygenic risk burden, the two patients with the highest burden for MS (z-score >2) had MRI-evident delayed myelination and T cell mediated central nervous disorder. **Conclusion.** We outline here a nascent methodology and report on initial results for evaluating extremes of polygenic risk in rare disease cohorts. Additional data will be presented from our broader rare disease cohort (>200 affected individuals) including summary statistics and illustrative examples.

## PgmNr 2614: Assessment of pathogenic structural variants in medically actionable genes in 982 healthy individuals.

### Authors:

J. Morales-Rosado <sup>1,2</sup>; F. Pinto e Vairo <sup>1,3</sup>; E. Jessen <sup>2</sup>; R. Moore <sup>2</sup>; Y. Liu <sup>2</sup>; C. Wang <sup>2</sup>; E.W. Klee <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Individualized Medicine, Mayo Clinic, Rochester MN 55905; 2) Department of Health Sciences, Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN 55905; 3) Department of Clinical Genomics, Mayo Clinic, Rochester MN 55905

---

**Background:** The clinical utility of whole-genome sequencing (WGS) has been primarily studied in the context of suspected heritable disorders and cancer. However, there is an increasing interest of its use for preventative medicine. Currently, limited statistics exist on the prevalence and penetrance of pathogenic alterations in actionable genetic conditions in the general population. Additionally, most studies have focused on single nucleotide variants (SNVs) and indels, with data lacking on the contribution of pathogenic structural variation (SV), including copy number variants (CNV) in the general population.

**Methods / Results:** We processed WGS data (median of 20X) generated for 982 unrelated individuals of primarily European ancestry in the Mayo Clinic Biobank. Subjects had no personal history of hematological malignancy or bone marrow transplant at the time of recruitment and an average length of electronic health record (EHR) data of 19 years. Sequencing data was analyzed using GenomeGPS v4.0.1 with BWA-MEM alignment to GRCh38 reference. SV calling was done with 4 algorithms: Lumpy, Delly, Cnvnator and Wandy (an internally developed segmentation algorithm for CNV events). Context and technical annotation was used to filter findings, resulting in a total of 8,870 deletions, 2,633 inversions and 238 duplications identified across the 982 samples. On average, 242 events per sample overlapping one of the 59 genes deemed as actionable by the ACMG were called.

**Discussion/Conclusion:** We found a frequency of 0.2% of known pathogenic deletions in our cohort. Two 11kb germline heterozygous deletion encompassing the BRCA1 gene were identified. One deletion belonged to a female with strong personal and family history of breast and colon cancer; while the other was found on a male with no personal or family history of cancer. Our experience demonstrates the utility, feasibility and challenges in technical and context annotation associated with interrogating a WGS dataset for SV. Replication of our SV calling and filtering strategy in larger sample sizes may inform the clinical advantage of our approach for comprehensive interrogation of WGS in “healthy” cohorts.

## PgmNr 2615: Clinical evaluation of rapid whole-genome sequencing in the intensive care unit.

### Authors:

B. Wu; P. Zhang; Y. Lu; H. Wang; W. Zhou

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of pediatrics, children's hospital of Fudan University, shanghai, China

---

**Background and aim:** Rapid diagnosis of genetic disease in critically ill newborns is vital to the swift initiation of targeted treatment. The constellation of symptoms that constitute the clinical diagnosis of a genetic disorder may be non-specific and difficult to recognize in the newborn period. Whereas, rWGS (P) technology has the potential interrogate a variety of genetic etiologies in a single test with the shortest time. In this study we aim to establish an rWGS workflow and evaluate its role in the diagnosis of patients suspected with genetic diseases in ICU.

**Methods:** Critical ill infants in NICU/PICU were enrolled according to the inclusion criteria. Clinical manifestations, laboratory results were ascertained comprehensively by physicians. DNA was isolated from the blood samples of trio-family numbers. Trio WGS sequencing were performed on Illumina NovaSeq 6000 sequencing platform. Rapid alignment, variant calling and interpretation will be conducted by Center for Molecular Medicine of Children's hospital of Fudan University. Variants are categorized according to American College of Medical Genetics (ACMG) recommendations for reporting sequence.

**Results:** Forty nine critical ill infants, including 26 males were enrolled and received the Trio-rWGS test, 71% of whom were less than 1 month of age. TAT time from sample collection to oral report was ranging from 7-12days. Twenty of 49 cases received diagnosis by rWGS. Pathogenic or likely Pathogenic variants were detected in 16 genes (6 AD, 8 AR and 2 XLR). In addition, pathogenic microdeletion and/or duplication were detected in 4 cases by using copy number variant calling on the WGS data. Case 28, who had an unbalanced chromosome aberration with a duplication in the long arm of chromosome 21 and a distal deletion in the long arm of chromosome 18, derived from a maternal balanced translocation. The results of genetic testing in 6 cases have a regulatory effect on clinical treatment, and 4 cases have turned to palliative care guidance.

**Conclusions:** Rapid WGS provide a timely diagnosis of genetic diseases, and may improve the outcomes of the critical ill patients in NICU and PICU.



## PgmNr 2616: Implementation of sex identification and sample tracking into clinical genomic analysis workflows.

### Authors:

J. Hu <sup>1</sup>; A. Balasubramanian <sup>1</sup>; L. Zhang <sup>1</sup>; V. Korchina <sup>1</sup>; C. Kovar <sup>1</sup>; E. Venner <sup>1</sup>; M. Murugan <sup>1</sup>; D. Murdock <sup>1,2</sup>; H. Zouk <sup>3</sup>; M. Harden <sup>4</sup>; A. Macbeth <sup>4</sup>; S. Harrison <sup>4</sup>; N. Lennon <sup>4</sup>; R. Rowley <sup>5</sup>; M. Smith <sup>6</sup>; A. Gordon <sup>6</sup>; G. Jarvik <sup>7</sup>; P. Sleiman <sup>8</sup>; M. Kelly <sup>9</sup>; S. Bland <sup>10</sup>; E. Boerwinkle <sup>1,11</sup>; C. Prows <sup>12</sup>; D. Crosslin <sup>13</sup>; I. Stanaway <sup>14</sup>; L. Mahanta <sup>3</sup>; H. Rehm <sup>3</sup>; R. Gibbs <sup>1</sup>; D. Muzny <sup>1</sup>; the eMERGE III consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Partners Healthcare Laboratory for Molecular Medicine (LMM), The Broad Institute of MIT and Harvard, Cambridge, MA; 4) The Broad Institute of MIT and Harvard, Cambridge, MA; 5) Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD; 6) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 7) Departments of Medicine (Medical Genetics) and Genome Sciences, University of Washington Medical Center, Seattle, WA; 8) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 9) Geisinger Genomic Medicine Institute, Danville, PA; 10) Vanderbilt University Medical Center, Nashville, TN; 11) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 12) Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio; 13) Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, Seattle, WA; 14) Kaiser Permanente Washington Health Research Institute, Seattle, WA

---

Next generation sequencing (NGS) has been rapidly adopted in clinical laboratories. Stringent quality control of sample identity for each patient-derived specimen is critical, however the large number of steps and complexity of the clinical sequencing workflow create the potential for sample identification errors at all phases, from sample collection to the sequencing and return of results. Here, we demonstrate the power of monitoring assigned sample sex assignment, to assure process quality. Methods were developed for the NIH Electronic Medical Records and Genomics (eMERGE) Network, that has collected >25,000 samples from 11 clinical sites and sequenced each using a 'gene panel', at two sequencing core sites, the Human Genome Sequencing Center Clinical Laboratory (HGSC-CL) at Baylor College of Medicine (BCM) and Partners Healthcare Laboratory for Molecular Medicine (LMM) in partnership with the Clinical Research Sequencing Platform (CRSP) at the Broad Institute. Following sample transfer to the sequencing sites, an aliquot of each was immediately tested with a Fluidigm SNP array panel, providing a unique genotype based upon 96 single base sites, which can be later compared with the gene panel sequence data, to ensure no sample swaps occurred while at the sequencing centers. As six of these SNP markers are assigned to sex chromosomes, they also were used to identify putative sample sex-assignment discrepancies that were subsequently confirmed by either (or both) the gene panel data, or hybridization to an Illumina exome array with 14k sex chromosome sites. The two sequencing sites have collectively confirmed 110/25,015 samples (0.44%) with inconsistencies between data at the genome centers versus test submission-provided sex assignment. Further investigation in collaboration with eMERGE clinical sites resolved most inconsistencies as the samples were from transgender participants, stem cell or bone marrow

transplant patients, or else mis-assignments by traceable clerical errors. As a result, process improvements were made at each collection site. The SNP array assay was demonstrated to be a cost-effective and robust tool that can be implemented in the clinical laboratory for validating process integrity. Although checking sex assignment detects just 50% of possible sample swaps and a similar proportion of transplant recipients, it will detect all transgender participants and identify requirements for process improvements in new pipelines.

## PgmNr 2617: Standardising single nucleotide variant classification through external quality assessment.

### Authors:

S. Deans; F. Khawaja; R. Treacy

[View Session](#) [Add to Schedule](#)

**Affiliation:** GenQA/UK NEQAS, Edinburgh, United Kingdom

---

The aim of this external quality assessment (EQA) was to provide an online competency assessment for individuals routinely performing interpretation and classification of variants. Participants were required to apply their usual processes to classify five single nucleotide variants (SNVs) and to submit evidence supporting the assigned classification. Participation was delivered using the Genomics Training, Assessment and Competency Tool (G-TACT). This EQA has been developed for genomic centres to demonstrate the competency of individuals, both clinical and laboratory based, performing variant classification and to identify any unmet training needs.

The variants ranged from class 3, uncertain significance to class 5, pathogenic and included variants in the following genes: *COL4A4*, *ETFDH*, *GK*, *SHOC2* and *SPG7*.

The submissions were scored, reported to each participant and a summary report was issued. The report detailed the expected variant classifications and provided an overview of all submitted results. A follow up webinar summarised the evidence assessed, the expected variant classifications and addressed participant queries.

This trial scenario was used by 142 individuals, with 82 participants completing the classification of all variants. As G-TACT is web-based then participation is global and 34 countries were represented worldwide. None of the participants provided the expected classification for all five variants, with only 10% of participants providing the expected classification for four variants. This clearly demonstrates the challenges associated with variant classification and the need for education, competency assessment and standardisation worldwide.

## PgmNr 2618: The genetic architecture of pediatric cardiomyopathy.

### Authors:

S.M. Ware<sup>1</sup>; S. Bhatnagar<sup>2</sup>; P.J. Dexheimer<sup>2</sup>; J.D. Wilkinson<sup>3</sup>; A. Sridhar<sup>1</sup>; E.M. Miller<sup>2</sup>; J.A. Schubert<sup>4</sup>; S.D. Colan<sup>5</sup>; L. Shi<sup>6</sup>; C.E. Canter<sup>7</sup>; D. Hsu<sup>8</sup>; S.A. Webber<sup>3</sup>; D.A. Dodd<sup>3</sup>; M.D. Everitt<sup>9</sup>; P.F. Kantor<sup>10</sup>; L.J. Addonizio<sup>11</sup>; J.L. Jefferies<sup>2</sup>; J.W. Rossano<sup>4</sup>; E. Pahl<sup>12</sup>; P. Rusconi<sup>13</sup>; W.K. Chung<sup>11</sup>; T. Lee<sup>11</sup>; J.A. Towbin<sup>14</sup>; L.A. Sleeper<sup>5</sup>; A.K. Lal<sup>15</sup>; H. Razoky<sup>16</sup>; J. Czachor<sup>16</sup>; B. Aronow<sup>2</sup>; L.J. Martin<sup>2</sup>; S. Lipshultz<sup>17</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics and Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana.; 2) University of Cincinnati School of Medicine and Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, TN; 4) Children's Hospital of Philadelphia, Philadelphia, PA; 5) Cardiology, Boston Children's Hospital and Harvard Medical School, Boston, MA; 6) New England Research Institutes, Watertown, MA; 7) Washington University, St. Louis, MO; 8) Albert Einstein College of Medicine and Children's Hospital of Montefiore, Bronx, NY; 9) Children's Hospital Colorado, Denver, CO; 10) Stollery Children's Hospital, Edmonton, AB, Canada; 11) Columbia University Medical Center, New York, NY; 12) Ann and Robert H. Lurie Children's Hospital, Chicago, IL; 13) University of Miami Miller School of Medicine, Miami, FL; 14) Le Bonheur Children's Hospital, Memphis, TN; 15) Primary Children's Hospital, Salt Lake City, UT; 16) Wayne State University School of Medicine, Detroit MI; 17) Pediatrics, John R. Oishei Children's Hospital, Buffalo, NY

---

Cardiomyopathy (CM) in childhood is a rare disease with substantial morbidity and economic burden. Identifying the underlying genetic etiology in a child with CM has the potential to guide management, improve risk stratification, and decrease risk of sudden death, but the genetic basis of early onset CM is not well understood. To study this, we performed exome sequencing in 528 unrelated children with CM. ACMG guideline-based clinical interpretation of rare variants in 37 known CM genes gave an overall diagnostic yield of 32%, including 19% in patients with dilated CM, 51% in hypertrophic CM, 50% in restrictive CM, 31% in left ventricular noncompaction (LVNC)/mixed, and 33% in non-LVNC mixed. Each phenotype exhibited differences with respect to age at diagnosis, family history, and ancestry. Infants were less likely to have diagnostic results than older children ( $p = 0.0009$ ). To identify new causal genes, we created a discovery set of 2188 genes for bioinformatics evaluation of rare ( $MAF < 0.01$ ) deleterious variants using MetaSVM and CADD. A prioritized candidate gene list was identified by ranking observed intolerance to variation in 1000 Genomes. Interestingly, ClinVar CM genes represented  $< 5\%$  of the prioritized genes, with genes derived from ontologies for abnormal heart morphology in mouse models and human heart development showing strong contribution ( $p < 0.0001$ ). Analysis of variant burden by ancestry, CM phenotype, and variant type (loss of function, nonsynonymous) identifies important patterns. Overall, our CM cohort is enriched for rare deleterious variants as compared to 1000 Genomes controls ( $p < 0.0001$ ). However, the rates of deleterious burden in cases versus controls for African and Admixed American as compared to European ancestries indicate that possible ancestry-specific genetic etiologies remain to be identified, especially for hypertrophic CM. There was phenotype-specific enrichment of patients with multiple rare deleterious variants and the greatest burden was seen in ClinVar CM genes. Loss of function variation accounted for only  $\sim 12.3\%$  of deleterious variant burden. Overall, this bioinformatics

approach identifies 63% of the pediatric CM cases with deleterious variants, an increase of 31% over clinical interpretation. This study represents a large pediatric cardiomyopathy cohort with comprehensive genetic evaluation and supports both shared and unique phenotype-, ancestry-, and age of diagnosis-specific contributions to CM.

## PgmNr 2619: Identification of rare de novo variants in transposition of the great arteries families using whole exome sequencing.

**Authors:**

Y. Zeng; W. Chen; Y. Liu; Z. Zhou

[View Session](#) [Add to Schedule](#)

**Affiliation:** Fuwai Hospital, Beijing, China

---

**Introduction:**

Transposition of the great arteries (TGA), one of the most common and severe types of congenital heart disease (CHD), is the leading cause of cardiac death in neonates and infants if not treated. The etiology of TGA is largely unknown, but increasing evidence supports that genetic factors play an important role in the pathogenesis of transposition of the great arteries. We used whole exome sequencing to assess the prevalence of rare deleterious and de novo variants in the cohort of familial TGA.

**Methods and Results:**

For this study, 66 TGA trios (proband and their parients) were recruited and screened together with a control population of 182 obesity patients using whole exome sequencing. Variants calling and filtering were performed according to GATK best practice pipeline. The screening criteria for rare variants were MAF < 0.001 in public cohorts (ExAC EAS; esp6500; 1000 Genome EAS; gnomAD exome EAS) and for damaging variants were LOF (frameshift; stop gain; splicing) or missense variants with CADD phred score > 15(DNS).The screening criteria for de novo variation were as follows: hitting by at least one de novo LOF/DNS allele in TGA, hitting by at least two LOF/DNS allele in TGA, no more than 1 LOF/DNS allele in control. All variations were verified using Sanger. We identified 11 genes carrying rare de novo variants including 1 loss-of-function variants and 41 missense variants among nine probands. Top candidate genes (4 in 12) with de novo damaging variants were FOXH1, TMEM260, DYNC2LI1 and USP9X.

**Conclusions:**

Our study revealed that a great portion of variants were de novo, suggesting that de novo variants played important roles in CHD pathogenesis. This may help stratify patients for guiding the therapeutic management.

## PgmNr 2620: Characterization of thoracic aortic disease in vascular Ehlers-Danlos syndrome: Defining the phenotype.

### Authors:

A.C. Cecchi <sup>1</sup>; E.S. Regalado <sup>1</sup>; D. Guo <sup>1</sup>; M. Azouz <sup>1</sup>; A.L. Estera <sup>2</sup>; H.J. Safi <sup>2</sup>; S.K. Prakash <sup>1</sup>; D.R. Deyle <sup>3</sup>; K.E. Agre <sup>3</sup>; L. Jayakumartie <sup>4</sup>; D.M. Milewicz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, Texas, USA; 2) Department of Cardiothoracic and Vascular Surgery, University of Texas Health Science Center at Houston, Houston, Texas, USA; 3) Department of Clinical Genomics, Mayo Clinic, Rochester, Minnesota, USA; 4) University of Texas Health Science Center at San Antonio, Division of Vascular and Endovascular Surgery, San Antonio TX

---

The arterial pathology, namely spontaneous dissection and rupture, is the primary cause of fatality in individuals with vascular Ehlers-Danlos syndrome (vEDS). In addition to the abdominal aorta, medium sized visceral, cervical, and peripheral arteries are most often affected. Challenges characterizing the phenotype in vEDS, including misdiagnosis, underdiagnosis, and a paucity of longitudinal data, still exist, but have lessened as more individuals are diagnosed via genetic testing. With newly generated data describing surgical outcomes and the impact of medical therapies, such as  $\beta$  blockers, the need for management guidelines has become more pressing. Unlike other heritable arteriopathies, such as Marfan syndrome, which is known to be associated with aortic root aneurysms and thoracic aortic dissections (TAAD), the thoracic aortic phenotype in vEDS remains poorly defined. We describe the thoracic aortic pathology in 12 individuals (7 female, 5 male) with pathogenic *COL3A1* variants ranging from 19 to 65 years at time of TAAD presentation; of which, 9 did not have a prior vEDS diagnosis. Missense variants (Gly substitutions) accounted for 8 cases, followed by 2 haploinsufficiency, and 2 splice-site. Four individuals were diagnosed with aneurysms of the aortic root or ascending aorta and 3 underwent successful elective repairs (4.7-5.7 cm). Acute type A dissections occurred in 5 cases, and type B dissections in 3. All 7 females had history of prior pregnancies, but only one had a major pregnancy-related complication--type B dissection with descending thoracic aortic rupture--immediately postpartum. Two individuals died from acute type A dissections; neither underwent surgical intervention. Surgery (elective and emergency) was pursued for 9 individuals and included 8 open repairs for ascending or aortic root aneurysm (3), type A dissection (3), and thoracoabdominal aneurysm associated with type B dissection (2). Thoracic endovascular repairs were performed for 3 cases (type A and type B dissections, descending thoracic aortic rupture). No fatalities occurred during surgery or within the 6 month post-operative period. These data confirm an increased risk for TAAD with associated mortality in patients with vEDS, and establish that elective thoracic aortic repairs can successfully be performed in experienced surgical centers. Thoracic aortic pathology is important to recognize as we consider optimal screening and interventional approaches for this population.

## PgmNr 2621: A rare splice-site variant in *TNNT2*: Cardiomyopathy and sudden death associated variant or common Polynesian polymorphism?

### Authors:

J. Ingles<sup>1</sup>; K. Thomson<sup>2</sup>; F. Harrington<sup>3</sup>; N. Henden<sup>3</sup>; K. McGuire<sup>2</sup>; C. Caleshu<sup>4</sup>; J.M. Bos<sup>5</sup>; M.J. Ackerman<sup>5</sup>; N. Nowak<sup>1</sup>; J. Juang<sup>6</sup>; M. Tiemensma<sup>7</sup>; A. Ronan<sup>8</sup>; C. Semsarian<sup>1</sup>; I. Hayes<sup>3</sup>; J. Skinner<sup>9</sup>; R.D. Bagnall<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Cardiology Program, Centenary Institute and The University of Sydney, Sydney, NSW, Australia; 2) Oxford University Hospitals NHS Trust, Oxford UK; 3) Diagnostic Genetics, Auckland Hospital, Auckland New Zealand; 4) Stanford Center for Inherited Cardiovascular Disease, Stanford School of Medicine, USA; 5) Windland Smith Rice Sudden Death Genomics Laboratory, Mayo Clinic, Rochester USA; 6) Cardiovascular Center and Division of Cardiology, National Taiwan University Hospital, Taipei Taiwan; 7) Forensic Pathology Unit, Royal Darwin Hospital, Darwin, Australia; 8) Hunter Genetics, Hunter New England Health Service, Newcastle, Australia; 9) Greenlane Paediatric and Congenital Cardiac Service, Starship Children's Hospital, Auckland New Zealand

---

A rare variant in cardiac troponin T (*TNNT2*; NM\_001001430.2: c.571-1G>A) was found via exome sequencing in an Aboriginal Australian who had a sudden unexplained cardiac death aged 23 years, and in a 2 year old of Maori ancestry with restrictive cardiomyopathy who died from rapidly progressive heart failure. Cardiac troponin T is an integral protein in the cardiac sarcomere. Missense variants in *TNNT2* are an established cause of hypertrophic cardiomyopathy (HCM).

Variant curation was performed. A previous functional assay of the variant demonstrated altered splicing of *TNNT2* exon 12, resulting in an in-frame deletion of one amino acid. Review of cases in ClinVar and the literature revealed >20 probands with a cardiac phenotype, including 12 with HCM, 3 with DCM including presentation <3 years in two children. Seven probands died suddenly with the variant identified on postmortem genetic testing. At least 5 probands report a family history of sudden cardiac death in a close relative. Two segregations were reported.

Given the related ancestry of the two cases, we queried groups/laboratories previously reporting the variant about ancestry of their cases. Amongst the 20 cases, >15 were reported as being of Polynesian ancestry, including New Zealand Maori, Samoan, Tongan and Hawaiian. Only 1 patient was reported to have European ancestry. Allele frequencies of a large publicly available Polynesian reference population do not exist, therefore determining whether the variant is a common Polynesian polymorphism is not possible. Polynesians share origins with Southeast Asia and Taiwan, with a widely accepted theory suggesting migration of Taiwanese Aborigines between 3000-1000 BC to Pacific Islands. The variant is absent in the Taiwanese Biobank (WGS), though this comprises primarily Han Chinese n=1517 participants. In the Genome Aggregation Database it is seen in 5/249,022 (0.002%) alleles, with the highest sub-population frequency in South Asian alleles (5/30,444; 0.016%). We classified this as a variant of uncertain significance. Whether it is over-represented in cases with cardiac disease and sudden cardiac death remains to be determined.



We demonstrate the challenges of variant curation in ethnic minorities, currently under-represented in reference datasets. Further, reporting of ancestry should be routinely included in variant assertions. In our experience this was not commonly reported and required additional specific contact with the laboratories.

## PgmNr 2622: *De novo* mutation in a MAP kinase leading to cardiac malformation and loss of vascular integrity.

### Authors:

M.S. Kane <sup>1</sup>; D.L. Bodian <sup>2</sup>; X.J. Liu <sup>2</sup>; M. Provenzano <sup>2</sup>; K.B. Tokarz <sup>2</sup>; B.D. Solomon <sup>2</sup>; N.S. Hauser <sup>2</sup>; T. Vilboux <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Functional Genetics and Genomics Lab, Inova Translational Medicine Institute, Falls Church, VA; 2) Inova Translational Medicine Institute, Falls Church, VA

---

Mitogen activated protein (MAP) kinases are critical signaling proteins involved in transmitting extracellular signals to regulate cellular proliferation, survival, apoptosis and gene expression. Not surprisingly, variants in the MAP kinases have been implicated in multiple types of cancer and these signaling cascades have been shown to be activated in a wide range of physiologic and disease states. We have identified a *de novo* variant in a MAP kinase component in a female neonate who presented with cardiovascular malformations including hypoplastic left heart and fatal vascular permeability. Based on the phenotypic similarities with some mouse model knockouts for MAP kinase cascade components and targets, we have investigated the potential pathogenicity of this variant in causing the congenital malformations observed in the proband.

Primary dermal fibroblasts from the proband demonstrate no loss of mRNA or protein but have significantly reduced phosphorylation of the protein in response to multiple MAP kinase-activating stimuli (interleukin, growth factor and cellular stress). Additionally, we observe altered transcriptional profiles in the patient cells and exaggerated cellular stress responses. These cell responses correlated with an increase susceptibility to activation of caspases and decreased cellular viability. Further analyses of the variant in an isogenic cell model were performed to identify transcriptional targets with aberrant expression in the context of the MAP kinase mutant protein. Collectively, these data indicate that the variant in the patient acts in a dominant negative fashion to deregulate the MAP kinase cascade, resulting in aberrant activity of the MAP kinase protein and its targets. These cellular changes alter key developmental pathways that likely contributed to the cardiac malformation and significantly impaired vascular integrity in the proband.

## PgmNr 2623: An Amish neonate with fatal cardiomyopathy due to homozygous *MYBPC3* mutation.

### Authors:

D. Karahalios; S. Baluta; J. Anderson; F. Smith; R. Lebel

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pediatrics, SUNY Upstate Medical University, Syracuse, NY.

---

**Introduction:** The Amish community is composed of genetically and geographically isolated, large, multigenerational families with similar environmental influences. The genetic isolation of these families increases the risk of homozygosity for rare autosomal traits due to founder effect.

**Case Report:** A 12 day-old Amish male presented to the emergency department with a history of respiratory distress, poor feeding, and cardiogenic shock. Term delivery was at home after an uncomplicated pregnancy. He was admitted to the pediatric intensive care unit for inotropic support and mechanical ventilation. An echocardiogram showed biventricular enlargement, hypertrophy, and moderate global hypokinesis, consistent with hypertrophic and dilated cardiomyopathy. After initiation of digoxin and an ACE inhibitor, the patient clinically improved, was extubated, and was discharged home on anticongestive medications. Two weeks later, he was readmitted to the pediatric intensive care unit with worsening congestive heart failure exacerbated by a viral upper respiratory infection. An echocardiogram showed worsening hypokinesis and an ejection fraction of 29% (normal >55%). Due to his poor prognosis, his parents declined further interventions. He was discharged home with palliative care and died the next day. Before his second discharge, genetic testing was performed and revealed that he was homozygous for a pathogenic variant in *MYBPC3* (c.3330+2T>G), which produces a truncated protein through frame shift and translation of a premature stop codon. Heterozygosity and homozygosity of this variant has been associated with both dilated and hypertrophic cardiomyopathy in the Amish population.

**Discussion:** Cardiomyopathy presenting in the neonatal period carries a poor prognosis that is even worse when associated with a genetic component. This patient's pathogenic variants in *MYBPC3* were inherited in an autosomal dominant fashion, and homozygosity profoundly worsens the prognosis. The parents of this patient are not significantly symptomatic, but as heterozygotes, they are likely to develop cardiomyopathy in the future. Since early medical treatment may reduce the progression of disease, cardiac assessment is indicated.

## PgmNr 2624: Rare association of mucopolidosis III alpha/beta with dilated cardiomyopathy.

### Authors:

H. Lee <sup>1</sup>; M. Kwack <sup>2</sup>; G. Song <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cardiology Dept, Pusan National University Hospital, Busan, Korea, Republic of; 2) Pediatrics Dept, Pusan National University Hospital, Pusan National University School of Medicine, Busan, Korea, Republic of; 3) Department of Computer Science, Pusan National University, Busan, Korea, Republic of

---

Mucopolidosis III alpha/beta (ML III alpha/beta) is an autosomal recessive lysosomal storage disorder caused by N-acetylglucosamine-1-phosphotransferase (GlcNAc-phosphotransferase) deficiency. It is characterized by coarse facial features, developmental delay, short stature, and skeletal deformities. Its cardiovascular symptoms include valvular thickening or hypertrophic cardiomyopathy. A 32-year-old female patient received heart transplantation due to end-stage heart failure caused by dilated cardiomyopathy (DCM). We performed whole exome sequencing to determine the etiology of DCM and/or skeletal deformities. The test revealed c.2715+1G>A and c.3173C>G mutations in the GNPTAB gene encoding the alpha and beta subunits of GlcNAc-phosphotransferase. Finally, she was diagnosed with ML III alpha/beta. This report describes a rare case of ML III alpha/beta associated with DCM. Our findings expand the clinical spectrum of ML III alpha/beta.

## PgmNr 2625: Novel *SMAD3* in-frame indel associated with thoracic aortic aneurysm and dissection: Potential loss of function mechanisms.

### Authors:

M.L. Murray <sup>1</sup>; C. Wilson <sup>2</sup>; W. Allen <sup>2</sup>; J. Tusi <sup>1</sup>; D. Macaya <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) GeneDx, Gaithersburg, MD; 2) Fullerton Genetics Center, Mission Health, Asheville, NC

---

Pathogenic variants in *SMAD3* are associated with a risk to develop aortic and arterial aneurysms/dissections with variable systemic features, referred to as aneurysm-osteoarthritis syndrome or Loeys-Dietz syndrome 3. While loss of function due to nonsense, frameshift, or canonical splice site variants is an established mechanism of disease, missense variants represent over 60% of those reported to date, most located in the MH2 domain. In clinical testing, rare missense and in-frame insertion/deletion (indel) variants can be difficult to interpret, especially in protein domains that are less well-characterized or in which few similar variants have been observed. We report a novel in-frame indel variant that segregates with disease in a three generation family with thoracic aortic aneurysm and dissection. Clinical features present in genotype-positive family members include thoracic aortic aneurysm/dissection, precocious osteoarthritis, pectus deformity, pelvic organ prolapse, and joint hypermobility. The variant, c.275\_281delinsC, p.Trp92\_Trp94delinsSer, is located in the MH1 domain of the *SMAD3*-encoded protein. Missense substitution of these residues, Arg93Gln and Trp94Leu, have been published as variants of uncertain significance, supporting the difficulty of clinical interpretation of variants in this region of the protein. The involved residues, Trp92-Arg93-Trp94, are highly conserved across species and in homologous SMAD proteins, and participate in hydrogen bonding critical for local protein structure/function. One important function of the MH1 domain is to bind DNA as a transcriptional factor; however, these residues do not directly participate in DNA binding. Published functional analysis of variants in the residues equivalent to Arg93 in homologous proteins (*SMAD2* Arg133Cys, *SMAD4* Arg100Thr), suggest multiple possible loss-of-function mechanisms, including increased autoinhibition, decreased DNA binding, and/or increased protein degradation. Association of p.Trp92\_Trp94delinsSer with an aortopathy and connective tissue phenotype in this family establishes the clinical importance of these residues, which will aid in interpreting similar variants encountered in clinical testing. Furthermore, review of similar variants in homologous proteins suggest loss-of-function by various mechanisms not readily predicted based upon variant type and location. Variants in other SMAD proteins may offer insight to novel *SMAD3* variants detected in clinical testing.

## PgmNr 2626: Ventricular septum defect can be a key for the early diagnosis of Brugada syndrome in young patients with *SCN5A* mutations.

### Authors:

S. Ohno <sup>1</sup>; J. Ozawa <sup>2</sup>; K. Sonoda <sup>1</sup>; T. Makiyama <sup>3</sup>; M. Horie <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Bioscience and Genetics, National Cerebral and Cardiovascular Ctr, Suita, Japan; 2) Department of Pediatrics, Niigata University Graduate School of Medical and Dental Sciences, Niigata, Japan; 3) Department of Cardiovascular Medicine, Kyoto University of Graduate School of Medicine, Kyoto, Japan; 4) Center for epidemiologic research in Asia, Shiga University of medical Science, Otsu, Japan

Background: Brugada syndrome (BrS) is an inherited disease characterized by the coved type ST elevation in right precordial leads and sudden cardiac death caused by ventricular arrhythmias (VA). Although BrS is frequent in middle aged men, young patients sometimes suffer VA due to BrS. Therefore, early diagnosis of BrS is indispensable to prevent sudden cardiac death. *SCN5A* encodes cardiac sodium channel, NaV1.5, and nearly 20% of the adult BrS patients carry mutations in *SCN5A*. BrS patients with *SCN5A* mutations frequently shows cardiac conduction block (CCB). CCB is also accompanied with ventricular septum defect (VSD). Therefore, VSD might be a key for the diagnosis of BrS.

Purpose: This study aimed to clarify the clinical phenotype of young BrS patients with VSD.

Methods: The study cohort is 44 (9 females) BrS probands who were diagnosed at age <20. We performed clinical evaluation including cardiac echography and genetic analysis for *SCN5A*.

Results: The mean age of diagnosis for BrS is  $11.0 \pm 5.4$  years old. Three (6.8%) patients accompanied with VSD. We detected *SCN5A* mutations in 18 (41%) patients, and all the patients with VSD carried the mutations. The first patient was an asymptomatic 4-year-old boy who received surgical repair of VSD at 8-month-old. His ECG 5 months after the operation showed complete right bundle branch block (CRBBB) without ST elevation. At the age of 4, ST elevations in V1-2 leads were detected, and pilsicainide loading test provoked coved type ST elevation. He carried a missense mutation, c.4035G>T, p.W1345C. The second patient was an asymptomatic 7-year-old boy who were diagnosed membranous VSD at the age of 2, and it was closed spontaneously. His ECG at 2 y.o. revealed CRBBB and coved type ST elevation, and CRBBB pattern diminished after spontaneous closure. His mutation was c.2677C>T, p.R893C. The last patient was an asymptomatic 2-year-old boy who received surgical repair for membranous VSD. His ECG before operation showed CRBBB and coved type ST elevation in V1-V3. He carried c.1099C>A, p.R367S. In NaV1.5, the first mutation, p.W1345C, locates in the segment 5 of domain 3, and p.R893C and p.R367S locate in the pore regions of domain 2 and 1, respectively.

Conclusions: Young VSD patients sometimes accompanied with BrS and carried *SCN5A* mutations. Careful ECG recording for VSD patients would be helpful for the early diagnosis of BrS with *SCN5A* mutations, leading to the prevention of sudden cardiac death.

## PgmNr 2627: *De novo* rare variants clustered in a limited region of *RNF213* lead to early onset and aggressive Moyamoya disease.

### Authors:

A. Pinard <sup>1</sup>; A.C. Cecchi <sup>1</sup>; A.L. Rideout <sup>2</sup>; M.D.J. Fiander <sup>3</sup>; D. Guo <sup>1</sup>; S. Parkash <sup>5</sup>; S. Walling <sup>4</sup>; M.J. Bamshad <sup>6,7</sup>; D.A. Nickerson <sup>7</sup>; A.M. Vandersteen <sup>5</sup>; D.M. Milewicz <sup>1</sup>; University of Washington Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Internal Medicine - Division Medical Genetics, The University of Texas Health Science Center at Houston, Houston, Texas; 2) Maritime Medical Genetics Service, IWK Health Centre Halifax, NS Canada; 3) Division of Neurosurgery, Dalhousie University, Halifax, Nova Scotia; 4) Division of Neurosurgery, Dalhousie University, IWK Health Centre Halifax, NS Canada; 5) Department of Pediatrics, Division of Medical Genetics, Dalhousie University, Medical Geneticist, Maritime Medical Genetics Service, IWK Health Centre Halifax, NS Canada; 6) Division of Genetic Medicine, Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; 7) Department of Genome Sciences University of Washington, Seattle, WA 98195, USA

Moyamoya disease (MMD) is a rare cerebrovascular disease, characterized by progressive stenosis and occlusion of the distal internal carotid arteries with the formation of a compensatory network of collateral vessels. MMD predisposes to ischemic and hemorrhagic strokes with two incidence peaks of onset, one in children around 10 years of age and another in adults in their 30-40 years. The high incidence of MMD in Japan is due to *RNF213* founder variant, p.Arg4810Lys, present in 90% of Japanese MMD cases but only 2.5% in healthy controls with the same ethnic background. We identified other *RNF213* rare variants in European descent MMD cases that segregated with disease in families and these variants clustered in and around the RING domain spanning amino acids 3997-4035.

Exome sequencing of 145 individuals from 80 families, including 28 trios were analyzed. We identified 3 *de novo* variants in *RNF213* in 3 independent European descent cases in a highly conserved region (4114-4120). The first case carried a one amino acid deletion, p.Lys4115del, had his first stroke at 10 months, and died at 1 year old following revascularization surgery. The second case carried a missense variant, p.Thr4114Arg, and is a 10-month-old girl diagnosed with severe bilateral MMD requiring neurosurgical procedure. The third case carried the variant p.Phe4120Leu, had his first stroke at 1 year old, and was diagnosed with bilateral MMD at 3 years old. After revascularization surgery, he had no further strokes but did develop bilateral renal stenosis and right nephrectomy. He is currently 37 years old. Further, there is a published case of a Hispanic girl with a *RNF213 de novo* variant, p.Ser4118Phe, diagnosed with MMD at 3 months as well, and also had narrowing of the inferior abdominal aorta.

These results strongly support the conclusion that *de novo* missense or small amino acid deletion within the 4114-4120 amino acid region lead to early onset and aggressive MMD. Interestingly, one case presented with early onset and aggressive MMD that did not progress with age, but rather developed bilateral renal artery stenosis. Future studies will be focused on defining the role of this region in the *RNF213* protein to elucidate why alterations in this region lead to such early and severe

MMD.



## PgmNr 2628: The Cardiac Genome Clinic: Genome sequencing in families with congenital heart disease.

### Authors:

M.S. Reuter<sup>1,2</sup>; R.R. Chaturvedi<sup>2,3</sup>; E. Liston<sup>2,4,5</sup>; R. Manshaei<sup>2</sup>; S.M. Hosseini<sup>2</sup>; M. Curtis<sup>2</sup>; R. Khan<sup>2</sup>; J. Okello<sup>2</sup>; I. Cohn<sup>2,6</sup>; R. Hayeems<sup>2,7,8</sup>; R. Cohn<sup>2,4,5</sup>; S. Bowdin<sup>2,4,5</sup>; R. Aul<sup>2,4</sup>; S.W. Scherer<sup>1,5,9</sup>; R.H. Kim<sup>2,4,5,10</sup>; R. Jobling<sup>2,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 2) Ted Rogers Centre for Heart Research, The Hospital for Sick Children, Toronto, ON, Canada; 3) Division of Cardiology, The Hospital for Sick Children, Toronto, ON, Canada; 4) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 6) Division of Pharmacology, The Hospital for Sick Children, Toronto, ON, Canada; 7) Child Health Evaluative Sciences Program, Hospital for Sick Children Research Institute, Toronto, ON, Canada; 8) Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, ON, Canada; 9) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 10) Fred Litwin Family Centre in Genetic Medicine, United Health Network, Toronto, ON, Canada

---

Genome sequencing has the potential to characterize the molecular architecture of congenital heart disease, thereby contributing to improved clinical outcomes through personalized management and treatment plans. The Cardiac Genome Clinic (CGC) was established as a prospective multi-site project to investigate the optimal use of novel genome-based technologies in pediatric and adult cardiology settings.

As part of a pilot study, we obtained genome sequencing data of 88 families, and systematically analyzed for rare disease-causing sequence and copy number variants. In 10 families (11%), we identified pathogenic or likely pathogenic variants associated with cardiac disease (in *ANKRD11*, *FLT4*, *KMT2D*, *NEXMIF*, *NIPBL*, *NR2F2*, *PTPN11* (2x), *PTEN*, *PTEN/PURA*). Promising variants with insufficient evidence for disease causality were identified in *FLNC*, *DNAH8*, *DNAH9*, as well as microdeletions at 3p11.2-3p12.3, 2p13.1-2p12, and 3p26.1-3pter. We also present potential novel gene-disease associations for tetralogy of Fallot, hypoplastic left heart syndrome, congenitally corrected transposition of the great arteries, and early-onset cardiomyopathy. The identified candidate genes have critical functions in heart development, such as angiogenesis, mechanotransduction, regulation of heart size, chromatin remodeling, and ciliogenesis.

Findings from this dataset demonstrate the value of unbiased genomic testing and comprehensive data analysis in patients with congenital heart disease. With regard to the genetic heterogeneity, the delineation of novel gene-disease associations will require large-scale genomic initiatives.

## PgmNr 2629: The role of glycogenin in cardiac function using a glycogen storage disease XV mouse model.

### Authors:

T. Rolle <sup>1</sup>; T. Rasmussen <sup>2</sup>; J. Lang <sup>3</sup>; N. Li <sup>3</sup>; M. Dickinson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human genetics, Baylor College of Medicine, Houston , TEXAS.; 2) Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, TEXAS; 3) Department of Medicine, Baylor College of Medicine, Houston, TEXAS

---

Glycogen storage diseases (GSD) are the result of mutations in genes encoding enzymes involved in glycogenesis or glycogenolysis. Glycogenin (GYG) encodes a muscle glycosyltransferase that catalyzes glycogen synthesis by nucleating the first glucose polymer before elongation and branching. Deficiencies in *GYG 1* results in GSD XV (OMIM: 613507) and Polyglucosan Body Myopathy 2 (OMIM: 616199). GSD XV is an autosomal recessive condition that has been characterized in 22 patients that have both skeletal and cardiomyopathies due to abnormal storage of glycogen in muscle tissue. To date, four males have been reported with pathogenic variants causing arrhythmias, heart dilatation and reduced ejection fraction. Two patients ultimately required heart transplants. While diagnosed cases of cardiomyopathy in *GYG 1* deficient patients are rare, further understanding of the role GYG plays in myocardial metabolism is insightful to understanding cardiac energy homeostasis and diseases pathology.

The mouse homolog for *GYG 1* is *Gyg*. *Gyg* deficient mouse knockouts have been reported to accumulate excess glycogen in striated muscle. However, unlike in skeletal muscle, the primary source of fuel in the heart is fatty acid. To explore the role of glycogen synthesis in cardiac function, we generated a *Gyg*<sup>Tm1b</sup> mouse allele in which *Gyg* was replaced with a LacZ reporter. *Gyg*<sup>Tm1b/Tm1b</sup> mice exhibit postnatal lethality by day 2. To visualize *Gyg* expression, we performed LacZ staining in the *Gyg*<sup>Tm1b/+</sup> mice and observed broad cardiac expression as early as embryonic day 8.5 through late fetal stages. In the adult animals, LacZ expression is restricted to the atria. Analysis of myocardial morphology in which mice will be analyzed histologically. In addition, the glycogen content of the hearts will be examined by Periodic Acid Schiff –Diastase staining and by quantifying glycogen content. To assess the functional effect of *Gyg* deficiency, electrocardiograms and echocardiograms will be used to measure the electrical activity of the heart. This study is essential in elucidating the mechanisms that link glycogen metabolism to proper cardiac function and morphology. Furthermore, it will provide novel insights into the pathophysiology of GSD XV patients who lack glycogenin-1 and show high glycogen accumulation.

## PgmNr 2630: De novo RYR2 mutations are associated with severe phenotype of CPVT more strongly than inherited ones.

### Authors:

K. Sonoda <sup>1</sup>; T. Hattori <sup>1</sup>; M. Horie <sup>2</sup>; S. Ohno <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Bioscience and Genetics, National Cerebral and Cardiovascular Center, Osaka, Japan; 2) Center for Epidemiologic Research in Asia, Shiga University of Medical Science, Otsu, Japan

### Background:

Catecholaminergic polymorphic ventricular tachycardia (CPVT) is one of the causes of sudden cardiac death in the young people. RYR2 is a major causative gene for CPVT, and the disease follows an autosomal dominant trait. However, de novo mutations are often identified during familial genetic analysis, and phenotype differences between patients with de novo and other mutations have not been elucidated yet.

### Purpose:

We aimed to characterize the patients carrying de novo RYR2 mutations and compared with those carrying inherited ones.

### Methods and Results:

In 50 CPVT probands with RYR2 mutations, we genotyped their parents and confirmed the origin of the each mutation. We identified 24 probands with de novo mutations and 26 with inherited mutations. The comparison of clinical characteristics was summarized in the table. Median age at first symptom was younger in the probands with de novo mutations than those with inherited one ( $P = 0.01$ ). The detection rate of bidirectional VT and polymorphic ventricular premature beats or VT was significantly higher in the probands with de novo mutations ( $P = 0.02$ ).

### Conclusion(s):

In our study, the phenotype of probands with de novo RYR2 mutations was severer than those with inherited ones. The patients with de novo RYR2 mutations frequently suffer fatal ventricular arrhythmia, and we should treat them more carefully.

	De novo N = 24	Inherited N = 26	P value
Male	12 (50%)	16 (61.5%)	0.41
Median age at first symptom, y	8.0 [6.0 – 10.0]	10.0 [8.5 – 13.0]	0.01
Median age at diagnosis, y	10.0 [8.0 – 15.0]	12.0 [10.0 – 14.0]	0.25
Syncope	19 (79.2%)	19 (73.1%)	0.79
Cardiopulmonary arrest and/or Ventricular fibrillation	14 (58.3%)	12 (46.2%)	0.37
Bidirectional VT	17 (70.8%)	9 (34.6%)	0.02

Polymorphic ventricular premature beats or VT	21 (87.5%)	14 (53.8%)	0.02
Atrial arrhythmia	2 (8.3%)	3 (11.5%)	1.00

## PgmNr 2631: Changes in variant classification between *KCNQ1*-modified criteria and standard ACMG 2015 guidelines.

### Authors:

M. Care <sup>1</sup>; V. Novelli <sup>2</sup>; A.C. Sturm <sup>3</sup>; C.L. Kurtz <sup>4</sup>; E. Abiusi <sup>2</sup>; H. Bikker <sup>5</sup>; J. Semotok <sup>6</sup>; R.E. Hershberger <sup>7</sup>; A.A. Willde <sup>8</sup>; M.J. Ackerman <sup>9</sup>; M. Perez <sup>10</sup>; W. Zareba <sup>11</sup>; J. Ware <sup>12</sup>; J. Garcia <sup>13</sup>; M.H. Gollob <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cardiology, University Health Network, Toronto General Hospital and University of Toronto, ON, Canada; 2) Fondazione Policlinico Universitario A. Gemelli I.R.C.C.S., and Istituto di Medicina Genomica, Università Cattolica del Sacro Cuore, Rome, Italy; 3) Geisinger Genomic Medicine Institute, Danville, PA, USA; 4) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 5) Department of Clinical Genetics, Amsterdam UMC, University of Amsterdam, Heart Center; Amsterdam, The Netherlands; 6) GeneDx, 207 Perry Parkway Gaithersburg, MD, USA; 7) Divisions of Human Genetics and Cardiovascular Medicine in the Department of Internal Medicine, Ohio State University, Columbus, OH, USA; 8) Department of Clinical and Experimental Cardiology, Amsterdam Cardiovascular Sciences, Amsterdam UMC, University of Amsterdam, Heart Center; Amsterdam, The Netherlands; 9) Departments of Cardiovascular Diseases, Pediatrics, and Molecular Pharmacology and Experimental Therapeutics, Divisions of Heart Rhythm Services and Pediatric Cardiology, Windland Smith Rice Sudden Death Genomics Laboratory, Rochester, MN, USA; 10) Division of Cardiovascular Medicine, Department of Medicine, Stanford University, Stanford, CA, USA; 11) Cardiology Unit of the Department of Medicine, University of Rochester Medical Center, Rochester, NY, USA; 12) National Heart & Lung Institute and MRC London Institute of Medical Sciences, Imperial College London and Royal Brompton and Harefield Hospitals NHS Trust, London, UK; 13) Invitae Corporation, San Francisco, CA, USA

**Background:** The 2015 ACMG/AMP Variant Interpretation Guidelines (ACMG<sub>2015</sub>) have become a cornerstone of genetic test interpretation. As these guidelines were written generally for all Mendelian disorders, the application of individual criteria may be subjective resulting in variant classification discrepancies. In an effort to provide more gene- and disease-specific guidance, the ClinGen Channelopathy Working Group modified ACMG<sub>2015</sub> for application to *KCNQ1* in the context of congenital Long QT syndrome (LQTS).

**Objective:** To compare the classification of *KCNQ1* variants using ACMG<sub>2015</sub> and *KCNQ1*-modified guidelines (*KCNQ1*-M).

**Methods:** Fifty *KCNQ1* variants were independently reviewed and classified by three curators, using both ACMG<sub>2015</sub> and *KCNQ1*-M. Final classifications were compared for concordance between curators and overall distribution of individual variant classifications.

**Results:** Completely concordant inter-curator classification was higher using *KCNQ1*-M than ACMG<sub>2015</sub> (34/50[68%] vs 19/50[38%];  $p < 0.01$ ), while clinically significant discordance occurred less frequently using *KCNQ1*-M than ACMG<sub>2015</sub> (12/50[24%] vs 20/50[40%], NS). Overall *KCNQ1*-M resulted in a lower number of combined likely pathogenic (LP)/pathogenic (P) classifications than ACMG<sub>2015</sub> ( $17 \pm 3.1$ ; 13-20 vs  $23 \pm 3.2$ ; 19-27) and a higher number of variants of uncertain significance (VUS) ( $28 \pm 2.6$ ; 26-32 vs  $23 \pm 4.7$ ; 20-30). The increase in VUS using *KCNQ1*-M resulted primarily from the requirement

of stringent scrutiny of the medical literature with respect to the diagnosis of LQTS in the context of specific variants. It was also affected by changes to the use of functional evidence, with the introduction of more specific requirements to activate functional studies as evidence for pathogenicity.

**Conclusions:** *KCNQ1*-modified guidelines resulted in a higher rate of concordance in final variant classifications than ACMG<sub>2015</sub> suggesting that gene-specific guidelines may be useful in decreasing discrepancies in variant interpretation. Although clinically relevant discordance approached 25%, this is expected to decrease with additional *KCNQ1*-M refinements. The increased number of VUS and decreased LP/P classifications requires further investigation as to whether the modified guidelines are too stringent resulting in overly conservative variant classifications or whether the lack of specificity of ACMG<sub>2015</sub> may contribute to inappropriate attribution of disease causality to an identified variant.

## PgmNr 2632: Variable clinical presentation of Congenital Thrombotic Thrombocytopenic Purpura in a large cohort of patients carrying a novel mutation in ADAMTS13 gene.

### Authors:

R. Parvari <sup>1,2</sup>; M. Arafat <sup>1,2</sup>; O. Pikovsky <sup>3,4</sup>; A. Rabinovich <sup>4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Shrager Segal Dept of microbiology, immunology and Genetics, Ben Gurion Univ of the Negev, Beer Sheva, Israel; 2) The National Institute for Biotechnology in the Negev, Ben Gurion University of the Negev, Beer Sheva, Israel; 3) Transfusion Medicine Institute, Soroka University Medical Center, Faculty of Health Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel; 4) Hematology Institute, Soroka University Medical Center, Faculty of Health Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel; 5) Thrombosis and Hemostasis Unit, Soroka University Medical Center, Faculty of Health Sciences, Ben Gurion University of the Negev, Beer Sheva, Israel.

---

**Introduction:** Congenital Thrombotic Thrombocytopenic Purpura (cTTP) is a life-threatening disorder, known to be caused by nearly 150 different mutations in ADAMTS13 gene. Its clinical manifestations are microangiopathic hemolytic anemia, thrombocytopenia and disseminated thrombosis.

**Materials and Methods:** PCR amplification of the exons of ADAMTS13 (NM\_139026) was performed after DNA extraction, followed by direct sequencing of the PCR products. We analyzed 17 patients; all belong to three closely related families with high rate of consanguinity. Presenting symptoms and disease severity among cohort members are markedly heterogeneous.

**Results:** We identified a novel mutation in ADAMTS13, causing a frameshift near the C-terminus of the protein. The mutation will result in the loss of the highly conserved amino acids of the full-length protein at the C-terminus of the protein, that contain CUB2 domain. We validated that all 17 patients of our study were homozygous for the mutation, while healthy individuals had the normal or heterozygous alleles. This cohort is one of the largest cTTP cohorts described in the literature and it is unique in the fact that all members have the same ADAMTS13 gene mutation. The mutation was not present in 254 Bedouin controls.

**Conclusions:** The novel frameshift mutation in ADAMTS13 further expands the spectrum of mutations seen in cTTP. We show that although all patients share the same mutation they vary significantly in the severity of the disease. Early identification of the mutation provides early preventive care instructions to the families in handling these patients and contributes for further family planning.

## PgmNr 2633: *TGFB3*-related Loews-Dietz syndrome causes arterial tortuosity in two patients and concurrent Klinefelter syndrome in one prevents aortic dilation.

### Authors:

J.R. Schweber <sup>1</sup>; J. Fleischer <sup>2</sup>; D. Groepper <sup>2</sup>; A.C. Braverman <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Southern Illinois University School of Medicine, Springfield, IL; 2) Pediatrics/Genetics, Southern Illinois University School of Medicine, Springfield, IL; 3) Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, MO

Loews Dietz syndrome (LDS) is a connective tissue disease that has been associated with vasculopathies, including aortic dissection and arterial tortuosity. Transforming growth factor - beta (TGF- $\beta$ ) signaling is an important contributor to the pathogenesis of LDS. Recently, TGF- $\beta$  ligand, *TGFB3*, has been shown to cause aortic aneurysms and dissections if mutated. Unlike other genes in the TGF pathway associated with LDS (*TGFB1*, *TGFB2*, *SMAD3*, *TGFB2*), previous studies of mutations in *TGFB3* have not found significant association with arterial tortuosity. We describe 2 patients with LDS5, a truncating *TGFB3* mutation (c.1020T>A, p.Y340X), and radiographic evidence of arterial tortuosity. Patient 1 presented at the age of 10 years with pectus excavatum, positive thumb and wrist signs, hind foot deformity, flat feet, scoliosis, high arched palate with bifid uvula, and a dilated aorta (z score 2.4). CT angiography illustrated tortuosity of the internal carotid artery (ICA), vertebral arteries and proximal portions of the anterior and middle cerebral arteries. Patient 2 presented at the age of 23 years with joint hypermobility, pectus excavatum, scoliosis, dysmorphic facial features, and mosaic Klinefelter Syndrome (KS). He did not have aortic dilatation, but was found to have tortuosity of the ICA and the vertebrobasilar confluence. This is the first patient in literature that presents with both LDS and KS, with the latter possibly being protective against aortic dilation based on a prior study showing small vessel sizes in patients with KS. In conclusion, we report two patients with a truncating *TGFB3* mutation and arterial tortuosity, despite prior studies suggesting absence of arterial tortuosity in patients with *TGFB3* mutations.

### Findings of current and previously reported *TGFB3* patients

Physical sign	Patient 1	Patient 2	<i>TGFB3</i> review - Scheper, 2018
Pectus carinatum/excavatum	excavatum	carinatum	13/20 (65%)
Flat feet	Y	Y	19/23 (83%)
Scoliosis or thoracolumbar kyphosis	Y	Y	20/29 (69%)
Retrognathia	micrognathia	Y	10/20 (50%)
Malar hypoplasia	Y	Y	3/10 (30%)
Skin Striae	N	Y	4/18 (22%)
Myopia	Y	N	6/15 (40%)



Mitral valve prolapse	regurgitation	N	6/15 (40%)
Marfan systemic score	8	7	
Arachnodactyly	Y	Y	18/45 (40%)
Aortic dilation/aneurysm	Y	N	8/24 (33%)
Aortic dissection/rupture	N	N	5/16 (31%)
Tall stature - Height (cm)	>3SD	6' 5'	
BMI	15.58	17.64	
Wide/bifid uvula	bifid	wide	19/24 (79%)
Cleft palate	N	N	7/13 (54%)
Hypermobility	Y	Y	21/26 (80%)
Arterial tortuosity	Y	Y	1/8 (13%)

## PgmNr 2634: Whole exome sequencing identifies a homozygous nine base pair (9bp) deletion in ARSB associated with familial dilated cardiomyopathy.

### Authors:

K. Al Harbi <sup>1</sup>; A.M. Abdallah <sup>2</sup>; J.C. Silas <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics Department, College of Medicine, Taibah University, Al-Madinah, Saudi Arabia; 2) West Midlands Regional Genetics Laboratory, Birmingham Women's NHS Foundation Trust , Birmingham , UK

---

Inherited cardiomyopathy is the major cause of sudden cardiac death (SCD) and heart failure (HF). The genetics of familial DCM are complex and accounting for 50% known disease-causing genes, but many causative genes are unknown. Here we aimed to identify the causative gene for DCM in a consanguineous family. The DCM case, unaffected sibs and parents had tested negative on a 181 candidate DCM genes. Whole-exome sequencing (WES) was performed on the affected DCM case, unaffected sibs and parents. *In silico* structural and functional analyses including protein modeling, structure prediction, and dynamic simulations were performed. Bioinformatics analysis identified a homozygous 9 base pair deletion in exon 1 of the *ARSB* gene that results in an in-frame deletion of three amino acids (Glycine and Alanine) at codon 38 to 40 in the affected DCM case. The variant is heterozygous in the parental samples, in the unaffected brother and was not detected in the unaffected sister. The mutation was predicted to cause a significant and deleterious change in the *ARSB* protein structure. The present study demonstrated *ARSB* mutation associated with life-threatening/pathogenic and recessively inherited DCM. This study highlights a limitation of the targeted gene sequencing and shows the potential for whole exome sequencing to identify the mutation(s) and novel gene(s) responsible for familial DCM of unknown genetic cause.

## PgmNr 2635: Rare variation in *SYNE1* is implicated in the development of childhood-onset essential hypertension.

### Authors:

I. Copeland<sup>1</sup>; M. Gupta<sup>2</sup>; J. Posey<sup>1</sup>; S.S. Hashmi<sup>3</sup>; A. Jajoo<sup>1</sup>; J. Rosenfeld<sup>1</sup>; D.A. Scott<sup>1</sup>; K. Schoch<sup>5</sup>; H.T. Chao<sup>1,4</sup>; V. Shashi<sup>5</sup>; J.R. Lupski<sup>1</sup>; M.S. Barratt<sup>3</sup>; D. Milewicz<sup>3</sup>; S. Shete<sup>3</sup>; N.A. Hanchard<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston 77030, Texas, USA; 2) Department of Cardiology, John Hopkins All Children's Hospital, St Petersburg 33701, Florida, USA; 3) McGovern Medical School, UTHealth, Houston, 77030, Texas, USA; 4) Department of Pediatrics, Division of Neurology and Developmental Neuroscience, Baylor College of Medicine, Houston 77030, Texas, USA; 5) Department of Pediatrics, Duke University, Durham 27708, North Carolina, USA

---

Childhood-Onset Essential Hypertension (COEH) affects 1-2% of children, most often underrepresented minorities, and leads to an increased later-life risk for cardiovascular and renal complications. Although difficult to diagnose, the heritability of COEH is estimated to be 60-80%. In contrast to adult-onset essential hypertension and syndromic forms of childhood secondary hypertension, genetic contributors to COEH remain unknown; however, the high heritability and low prevalence suggest a role for rare genetic variation.

We performed whole exome sequencing on an extensively phenotyped cohort of 68 COEH probands and available family members. After QC and annotation using ATLAS tools, we focused on rare (minor allele frequency < 1%), predicted protein-damaging variants, particularly those in any of 832 known hypertension genes compiled from the literature. We then prioritized variants whose family segregation was consistent with pedigree-implied modes of inheritance. This analysis identified 18 rare, putatively-damaging, biallelic missense variants *in trans* in *SYNE1* in 10 probands (14.7%). This proportion was significantly higher than among 860 trios from the Baylor-Hopkins Center for Mendelian Genomics, (Fisher-Exact P=0.037) and differed significantly from expectations derived from Hardy-Weinberg proportions of rare *SYNE1* variation in gnomAD (biallelic probability =0.004), indicating that biallelic *SYNE1* rare variation is generally uncommon. In five other non-cohort children with compound heterozygous *SYNE1* variation, available blood pressures were above the median despite normal Body Mass Indices. Lastly, using multi-dimensional scaling, COEH individuals with biallelic *SYNE1* variation clustered with individuals of African ancestry, consistent with rare *SYNE1* missense variants in gnomAD having higher frequencies among individuals of African ancestry.

*SYNE1* encodes Nesprin - a part of the cytoskeletal LINC-complex, in which genetic variation has been linked to other childhood-onset cardiovascular phenotypes. Missense variants identified in our cohort were all within the spectrin repeat domain of Nesprin, which is highly expressed in smooth muscle cells. In addition, *SYNE1* has been implicated in two large GWAS of hypertension, and rare *SYNE1* variants have been associated with cardiomyopathy. Our results implicate rare variation in *SYNE1* as a predisposing factor in COEH that might also contribute to the increased incidence of COEH among US minority groups.



## PgmNr 2636: Next-generation sequencing identifies a homozygous mutation in *ALMS1* associated with familial dilated cardiomyopathy.

### Authors:

J.C. Silas <sup>1</sup>; A.M. Abdallah <sup>2</sup>; K. Al Harbi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, College of Medicine, Taibah University, Al Madinah, Saudi Arabia; 2) West Midlands Regional Genetics Laboratory, Birmingham Women's NHS Foundation Trust , Birmingham , UK.

---

**Objective:** Cardiomyopathies represent the second largest group of patients in sudden cardiac death (SCD) victims, next to coronary artery disease. The genetics of familial DCM are complex and accounting for 50% known disease-causing genes, but many causative genes are unknown. Here we aimed to identify the causative gene for DCM in a consanguineous Saudi Arabian family.

**Patients and Methods:** The affected two children and their unaffected parents had tested negative on a 181 candidate DCM genes. Whole-exome sequencing (WES) was performed on the affected two children and their unaffected parents. The identified variant was screened in 50 clinically annotated controls and 50 DCM cases to evaluate its frequency in the Saudi Arabian general population and individuals with DCM. *In silico* structural and functional analyses including protein modeling, structure prediction, and dynamic simulations were performed.

**Results:** Bioinformatics analysis detected a homozygous nonsense variation in exon 8 of the Alstrom syndrome 1 gene (*ALMS1*) that results in a stop codon and premature truncation of the protein. The variant was heterozygous in the unaffected parents. The mutation was absent in 50 clinically annotated controls and 50 pediatric DCM cases. The mutation was predicted to cause a significant and deleterious change in the *ALMS1* protein structure.

**Conclusions:** The present study demonstrated *ALMS1* mutation associated with life-threatening/pathogenic and recessively inherited DCM. This study highlights a limitation of the targeted gene sequencing and shows the potential for whole exome sequencing to identify the mutation(s) and novel gene(s) responsible for familial DCM of unknown genetic cause.

## PgmNr 2637: Mutation spectrum of dilated cardiomyopathy from Korean patients.

**Authors:**

J. Lee; B. Kim; S. Park; M. Seong

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea

---

Dilated cardiomyopathy (DCM) is defined as a myocardial disorder with left ventricular dilatation and contractile dysfunction. Among idiopathic dilated cardiomyopathy, hereditary cardiomyopathy accounts for up to 50% of it. More than 40 genes are known to cause dilated myopathy and next-generation sequencing made it possible to exam these genes at once. The aim of the study was to identify the mutational spectrum of DCM causing genes from Korean patients. We enrolled 109 patients who were diagnosed as DCM and requested for DCM panel testing. A targeted panel of 48 genes was done with next-generation sequencing on genomic DNA extracted from venous puncture. Low coverage region was covered with Sanger sequencing.

Thirty-two patients (29%) have more than one pathogenic or likely pathogenic variants of DCM causing genes and no variant was detected from 17 patients. *MYH7* and *LMNA* were the most common causative genes. Among the pathogenic or likely pathogenic variants, there are 25 missense (78.1%), 4 frameshift (12.5%), 6 nonsense (18.8%), and 1 in-frame deletion (3.1%) variants. Our study showed that the proportion of hereditary cardiomyopathy was significant. Also, it provides insights into the pathophysiology of dilated cardiomyopathy.

## PgmNr 2638: Aggregation of homologous protein domains identifies constrained positions in *TTN* and prioritises missense variants for evaluation of a role in dilated cardiomyopathy.

### Authors:

Zhang<sup>1,2</sup>; N. Whiffin<sup>1,2</sup>; J. Ware<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) 1.National Heart and Lung Institute, Imperial College London, London, United Kingdom; 2) Cardiovascular Research Centre, Royal Brompton and Harefield NHS, Foundation Trust London, London, United Kingdom; 3) MRC London Institute of Medical Sciences, Imperial College London, London, United Kingdom

**INTRODUCTION:** The giant protein titin (encoded by *TTN*) is a crucial component of all striated muscle. *TTN*-truncating variants have been recognized as the most prevalent genetic cause of dilated cardiomyopathy (DCM). However, the assessment and interpretation of *TTN* missense variants in DCM has presented a profound challenge in clinical genome interpretation since rare missense variants in *TTN* are prevalent in the population, and indistinguishable from those in patients with DCM. Previous studies assessing the relevance of *TTN* missense variants, using sequence conservation across species or structural features of proteins, have been inconclusive.

**HYPOTHESIS:** The deleteriousness of missense variants depends on its molecular location within *TTN*. Since ~70% of the titin protein sequence is composed of ~300 Fibronectin type-III (fn3) and Immunoglobulin I-set (I-set) domains, the effect of missense variants in these two domains could be inferred from the genetic intolerance of the amino acid positions in the domain families to variation.

**METHODS:** To increase the power in prioritization of domain positions with large effect, we aggregated equivalent amino acid positions for each domain in the *TTN* canonical transcript. Using the variant data of 141,456 individuals in gnomAD, we then calculated the degree of depletion of neutral variation within the human genome at each position, based on a sequence context model.

**RESULTS:** We aligned 83 and 89 positions with high-coverage of residues across the fn3 and I-set domains respectively. In the fn3 domain, 57% of aligned positions are depleted of missense variants, four of them significantly ( $P$ -value<0.05). In the I-set domain, 42% of aligned positions are under constraint. To validate this approach, we evaluated the distribution of well-characterised myopathy-associated ClinVar variants between constrained and unconstrained positions. The top 5% most-constrained domain positions had a 1.8-fold increased disease association compared with the remaining positions, using a cohort of 990 patients with DCM and a control set of 2,090 healthy volunteers.

**CONCLUSIONS:** The aggregation of constraint within the human population on homologous domains represents complementary information to assess the relevance of domain positions, and highlights an

effective approach to prioritize missense *TTN* variants with large deleterious impact.



## PgmNr 2639: Common genetic variants associated with immune and inflammatory traits and disease severity phenotypes in hereditary hemorrhagic telangiectasia.

### Authors:

L. Pawlikowska<sup>1,2</sup>; J. Nelson<sup>1</sup>; C.E. McCulloch<sup>3</sup>; M.T. Lawton<sup>4</sup>; D.A. Marchuk<sup>5</sup>; M.E. Faughnan<sup>6,7</sup>; H. Kim<sup>1,2,3</sup>; Brain Vascular Malformation Consortium HHT Investigator Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Cerebrovascular Research, Dept Anesthesia and Perioperative Care, Univ California San Francisco, San Francisco, California; 2) Institute for Human Genetics, Univ California San Francisco, San Francisco, California; 3) Dept Epidemiology and Biostatistics, Univ California San Francisco, San Francisco, California; 4) Dept Neurosurgery, Barrow Neurological Institute, Phoenix, Arizona; 5) Dept Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina; 6) Toronto HHT Centre, St. Michael's Hospital and Li Ka Shing Knowledge Institute, Toronto, Canada; 7) Division of Respiriology, Department of Medicine, University of Toronto, Toronto, Canada

**Background:** Hereditary hemorrhagic telangiectasia (HHT) is an autosomal dominant Mendelian disorder caused by mutations in several TGF $\beta$ /BMP9 signaling pathway genes. HHT patients have variable phenotypes including skin and mucosal telangiectases, which cause severe and/or chronic nose and gastrointestinal bleeding and anemia, and organ vascular malformations (VM) and arteriovenous malformations (AVM), which can lead to stroke. Genetic modifier effects and inflammatory processes may contribute to HHT phenotypic heterogeneity. We hypothesized that common genetic variants associated with immune or inflammatory traits are associated with HHT phenotypes.

**Methods:** We genotyped 760 Caucasian HHT patients from the Brain Vascular Malformation Consortium (BVMC) on the Affymetrix Axiom UK Biobank array. We selected 5,081 common (MAF>1%) variants (or their LD proxies,  $r^2>0.8$ ) associated with 354 immune and inflammatory traits from the NHGRI-EBI Genome-Wide Association Study (GWAS) Catalog (version e93, October 2018). We evaluated 6 HHT severity phenotypes: any VM, lung AVM, brain VM, liver VM, anemia from HHT-related bleeding, and stroke (ischemic stroke from lung AVM or intracerebral hemorrhage from brain VM). Association of variant genotype with phenotype was evaluated using additive models adjusted for age, sex and 20 principal components.

**Results:** Six variants were associated with any VM, 6 with lung AVM, 6 with brain VM, 3 with liver VM, 6 with anemia and 10 with stroke at  $p<0.001$ , however no associations achieved FDR-corrected statistical significance. The strongest associations were for: any VM with a *MAZ* variant, previously associated with multiple sclerosis (OR=2.36,  $p=0.000054$ ), brain VM with a *MSI2* variant, associated with coronary artery calcification (OR=1.88,  $p=0.000095$ ), and lung VM with an *IGF1R* variant previously associated with endometriosis (OR=1.57,  $p=0.00034$ ). *MAZ* is a transcription factor involved in TGF $\beta$  signaling, cell proliferation and carcinogenesis. *MSI2* is an RNA-binding protein that modulates cell proliferation and IL6 signaling. IGF signaling has been implicated in HHT VMs. We will replicate top associations in a second cohort of 800 HHT patients from the BVMC.

**Conclusions:** GWAS catalog variants associated with immune or inflammatory traits may also be associated with disease severity phenotypes in HHT.

## PgmNr 2640: Identifying the genetic determinants of spontaneous coronary artery dissection: An analysis of whole genome sequences.

### Authors:

H.A. Burr <sup>1,2</sup>; L. Pilote <sup>2,3</sup>; A. Cheema <sup>4</sup>; L. Dufresne <sup>2</sup>; H.Y. Chen <sup>1,2</sup>; G. Thanassoulis <sup>1,2</sup>; J.C. Engert <sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Experimental Medicine, McGill University, Montreal, QC; 2) Research Institute of the McGill University Health Centre, Montreal, QC; 3) Division of General Internal Medicine, McGill University, Montreal, QC; 4) Department of Medicine, University of Toronto, Toronto, ON; 5) Department of Human Genetics, McGill University, Montreal, QC

---

Spontaneous Coronary Artery Dissection (SCAD) is a rare type of Acute Coronary Syndrome (ACS) that primarily affects younger individuals who do not exhibit “traditional” cardiovascular risk factors. Previous genetic and epidemiologic studies indicate that connective tissue disorders (CTDs) such as Marfan syndrome, Loeys-Dietz syndrome, and Ehlers-Danlos syndrome with known genetic causes are more prevalent among SCAD than other ACS patients, however, these conditions combined are thought to explain less than 0.5% of SCAD cases. The arteriopathy fibromuscular dysplasia (FMD) is also enriched in SCAD patients, and common variants at the *PHACTR1/EDN1* locus previously associated to FMD are also associated with SCAD. While many patients report extreme emotional episodes or physical exertion at SCAD onset, these environmental factors may serve as a trigger in patients with an underlying genetic cause of vascular weakness.

Previous genetic studies of SCAD have used CTD and familial thoracic aortic aneurysm and dissection (TAAD) gene panels to screen for causative mutations, however this approach achieved only an 8% rate of genetic diagnosis. In addition, due to the potential phenotypic heterogeneity of vascular weakness and low penetrance (only 1.2% of cases show familial inheritance) the use of trios, a common practice for Mendelian diseases, may not be useful. New approaches are needed to facilitate improved screening, diagnosis, treatment, and genetic counselling for SCAD patients.

We have Whole Genome Sequenced 45 SCAD patients in the GENESIS-PRAXY cohort, which consists of individuals who had an ACS event at or under age 55. SCAD individuals were significantly more likely to be female ( $p < .001$ ) and have an earlier age of onset ( $p = .006$ ) and less likely to smoke ( $p = 0.15$ ), be obese ( $p = .02$ ), or have hypertension ( $p = .15$ ), dyslipidemia ( $p = .06$ ), or diabetes ( $p = .35$ ). We constructed a pipeline to identify likely causal variants using previous reports of pathogenicity in ClinVar, computational predictions of variant deleteriousness, and allele frequencies in gnomAD. A pilot analysis of exomes revealed possible disease causing variants in the *FBN1* and *LEMD3* genes, which have known connective tissue functions.

## PgmNr 2641: Pathological variants in PLEKHO2 predispose individuals to heritable thoracic aortic disease.

### Authors:

D. Guo <sup>1</sup>; X. Duan <sup>1</sup>; A. Pinard <sup>1</sup>; E. Regalado <sup>1</sup>; L. Gong <sup>1</sup>; E. Hostetler <sup>1</sup>; M. Bamshad <sup>2</sup>; D. Nickerson <sup>2</sup>; D. Milewicz <sup>1</sup>; University of Washington Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Internal Medicine, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, Texas; 2) Department of Genome Sciences, University of Washington, Seattle, Washington

---

Up to 25% of individuals with thoracic aortic aneurysms and dissections (TAD) have multiple family members affected, termed heritable TAD (HTAD). Pathogenic variants in 11 genes are confirmed to predispose to HTAD, including genes that encode proteins involved in canonical TGF $\beta$  signaling (e.g., *TGFBR1* and *TGFBR2* encoding the type I and II TGF $\beta$  receptors). To identify additional HTAD genes, exome sequencing on affected individuals in HTAD families was analyzed and a rare heterozygous variant, p.Gly392Glu, in *PLEKHO2* was identified that segregated with TAD in a large HTAD family with a LOD score of 4.2. This variant is absent in the gnomAD v2.1.1 exome controls and has a combined annotation dependent depletion (CADD) score of 25.2. Analysis of exome sequencing data from 577 individuals from 367 HTAD families and 374 unrelated individuals with early-onset ( $\leq$  56 years of age) sporadic thoracic aortic dissections identified 7 additional rare *PLEKHO2* missense variants with CADD score  $>20$  and minor allele frequency less than 0.005 in the gnomAD v2.1.1 exome controls. *PLEKHO2* encodes a protein with no known function and is homologous to CKIP-1, which increases ubiquitination of bone morphometric signaling proteins (Smad1 and 6). When *PLEKHO2* was expressed in HELA with ubiquitin and various components of the canonical TGF $\beta$  signaling pathway, *PLEKHO2* was found to dramatically increase ubiquitylation of activated TGFBR1. *PLEKHO2* shRNAs were used to knockdown expression in smooth muscle cells (SMCs), which resulted in increased the expression of SMC differentiation genes by immunoblot analyses. Analysis of signaling pathways altered in these cells found increased TGFBR1 levels, but also increased levels of phosphorylated p38 MAPK levels. P38 MAPK levels are also present in SMCs from a Marfan mouse model (mgN/mgN mice). In summary, heterozygous rare variants in *PLEKHO2* cause HTAD, and loss of *PLEKHO2* in SMC alters the same signaling pathways identified in the SMCs of MFS mice, specifically activation of p38 MAPK.

## PgmNr 2642: Sex-limited penetrance of lymphedema to females with CELSR1 haploinsufficiency: A second family.

### Authors:

L. Lai <sup>1</sup>; D.J. Mustacich <sup>2</sup>; M.J. Bernas <sup>3</sup>; P.H. Kuo <sup>4</sup>; M.H. Witte <sup>2</sup>; R.P. Erickson <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pathology, University of Arizona College of Medicine, Tucson, Arizona.; 2) Surgery, University of Arizona College of Medicine, Tucson, Arizona.; 3) TCU and UNTHSC School of Medicine, Fort Worth, Texas.; 4) Medical Imaging (Nuclear Medicine), University of Arizona College of Medicine, Tucson, Arizona.; 5) Pediatrics, University of Arizona College of Medicine, Tucson, Arizona.

---

Development of the lymphatic system is complex and involves products of numerous genes. We report here a multi-generation family with hereditary lymphedema secondary to a novel frame shift variant (NM\_014246.1, c.5121dupC, p.1708fs\*44) in the planar polarity gene, *CELSR1*, a non-classic cadherin known to influence vascular endothelial cell migration and proliferation. This variant has not been reported in the publicly available human genomics databases: gnomAD, 1000 Genomes, the Exome Aggregation Consortium, or dbSNP and is highly conserved across multiple species. The dominant inheritance of the variant was discovered using whole exome sequencing and confirmed by Sanger sequencing. 22 family members were interviewed, examined, and blood drawn. Five females, and no males, exhibited lymphedema as determined by history and physical examination, and staged by International Society of Lymphology (ISL) staging criteria (I mild, II moderate, III severe). No other phenotypes such as kidney diseases or hearing loss were identified. Whole exome sequencing was performed in an affected female, her obligate carrier father, and her affected paternal aunt. All three subjects were shown to be heterozygous for *CELSR1* (p.1708fs\*44) variant, which was further confirmed by Sanger sequencing. Totally, 17 family members were subjected to Sanger sequencing and the heterozygous variant was shown in all 5 affected females and 4 obligate carrier males, but not in the unaffected females or non-carrier males, confirming the segregation of *CELSR1* (p.1708fs\*44) variant with lymphedema in females and carrier status in males in this family. Lymphoscintigraphy in affected females showed previously undescribed lymphatic abnormalities consistent with lymphangiectasia, valve dysfunction, and thoracic duct reflux. Interestingly, the first family with lymphedema with a variant (p.W1957X) in *CELSR1* gene also exhibits sex-limited penetrance to females (Gonzalez-Garay et al., 2016). The sex-limited manifestation of lymphedema is unique for hereditary lymphedema. The underlying mechanism is unknown but one possible explanation is hormonal regulation of lymphatic development consistent with sex variation in some other primary lymphedemas. In summary, the discovery of two multi-generation families with hereditary lymphedema, each with a novel loss of function *CELSR1* gene variant co-segregating with the disease, strongly suggests a role for *CELSR1* in human lymphatic development and function.

## PgmNr 2643: Development of a novel mouse genetic model to study vascular malformations and its use in understanding hereditary hemorrhagic telangiectasia.

### Authors:

D. Belefond <sup>1</sup>; O. Mamai <sup>2</sup>; S. Basu <sup>2</sup>; R. Akhurst <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Division of Medical Genetics, Univ California San Francisco, San Francisco, California.; 2) Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, California; 3) Department of Anatomy, University of California, San Francisco, San Francisco, California

---

Studying the molecular etiology of rare genetic conditions is important to the understanding of select disease processes and can provide crucial insights into the molecular etiology of more common conditions. Hereditary Hemorrhagic Telangiectasia (HHT) is a rare Mendelian condition characterized by multiple arteriovenous malformations (AVMs). Small AVMs (telangiectases) are apparent on mucocutaneous surfaces while larger AVMs may form in the lung (PAVM), brain, or liver. Patients often present with spontaneous and recurrent epistaxis, though hemorrhage from visceral AVMs can be sudden and catastrophic. Most cases of HHT are caused by mutations in *ENG* or *ACVRL1*, which encode endothelial receptors for TGF- $\beta$ s and BMPs. However, the cellular and molecular mechanisms driving HHT pathobiology are still unclear, though aberrant angiogenesis is thought to play a role. Our lab previously showed that human polymorphic variants of *PTPN14* associate with PAVM formation in both HHT type I and II. Our current study investigates the molecular genetics of PAVM formation in HHT using mouse knockout models for *ENG* and *PTPN14*. Our overarching hypothesis is that *PTPN14* interacts with the TGF $\beta$ /BMP9/endoglin/ *ACVRL1*/Smad4 signaling pathway to stabilize blood vessels such that reduced expression or loss of *PTPN14* will potentiate the formation of AVMs under conditions of genetic loss of an *ENG* allele. To test this hypothesis, we will investigate formation of the retinal vascular network during the first seven days of post-natal life in mice that are germline *Eng*<sup>+/-</sup>. *Ptpn14*<sup>+/-</sup> versus *Eng*<sup>+/-</sup> versus wild type mice. Additionally, we are constructing a mouse model with doxycycline-inducible endothelial cell-specific KO of the *Ptpn14* gene, utilizing a mouse with a *Ptpn14*<sup>flox</sup> allele that will be bred to an EC-specific CRE-ER recombinase. We will utilize immunohistochemistry to assay for arterial, venous, and lymphatic malformations in skin, lung, liver, and gastrointestinal tract as well as molecular techniques, including RNA-Seq, western analysis and IHC of lung tissues, to assess qualitative and quantitative changes in Smad and Taz signaling and global gene expression that may be correlated to clinical findings in HHT. Ultimately, this study has the potential for significant short- and long-term impacts on our understanding of the pathogenesis of vascular malformations and hemorrhage control both in HHT and more broadly, and may provide novel insights into drug treatment for HHT.

## PgmNr 2644: Validation of a genome-wide polygenic risk score on South Asian individuals for prediction of Coronary Artery disease risk.

### Authors:

R. Gupta <sup>1,10</sup>; M. Wallace <sup>6</sup>; S. Mishra <sup>1</sup>; R. Menon <sup>1,10</sup>; T. Deepak <sup>1,10</sup>; M. Deshmukh <sup>1,10</sup>; S. Apte <sup>1</sup>; O. Mathew <sup>1</sup>; B. Anikesh Vijay <sup>2</sup>; R. Gupta <sup>3</sup>; A.S. Mulasari <sup>4</sup>; A. Sigamani <sup>5</sup>; M. Kanchi <sup>5</sup>; S. Santhosh <sup>1,10</sup>; S. Kathiresan <sup>6,7,8,9</sup>; A. Khera <sup>6,7,8,9</sup>; V.L. Ramprasad <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MedGenome Labs Ltd., Bangalore, India; 2) Amrita Institute Medical Sciences, Kochi, India; 3) Eternal Heart Care Centre, Jaipur, India; 4) Madras Medical Mission, Chennai, India; 5) Narayana Hrudayalaya, Bangalore, India; 6) Cardiovascular Disease Initiative of the Broad Institute of Harvard and MIT, Cambridge, MA, USA; 7) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA; 8) Cardiology Division of the Department of Medicine, Massachusetts General Hospital, Boston, MA, USA; 9) Harvard Medical School, Boston, MA, USA; 10) Genomic Medicines Division, MedGenome, Foster City, CA

Recent studies have shown that the predictive power of polygenic risk scores (PRS) is substantial when the effects of common genetic variants are summed across the entire genome to create genome-wide PRS (GPS). The predictive power of GPS can be orthogonal to commonly used clinical tests such as LDL cholesterol, implying that GPS will be important tools in clinical practice and public health planning. GPS are dependent upon allele frequencies and effect sizes that vary between population groups, so calibration and validation will be needed before they can become broadly useful. South Asia (SA) has a very large burden of cardiovascular disease and an appropriate GPS will have enormous impact on human health in this part of the world. We carried out preliminary calibration and evaluation of an existing GPS algorithm (LDPred method and CARDIoGRAMplusC4D GWAS summaries) and demonstrated its utility in predicting individuals with high risk for coronary artery disease (CAD) in South Asians.

In this study, we genotyped 1,173 patients with coronary artery disease and 1,445 age and gender matched healthy controls using Illumina's Global Screening array (GSA) (1,173 patients and 942 controls) or whole genome sequencing (WGS) at 30X (503 controls). Development of a GPS requires a high-quality reference panel for imputation that is optimized for SA populations. We evaluated imputation of the quality passed common genotypes between GSA and WGS using a GenomeAsia Project (GAsP) reference panel containing 30X WGS from 1,739 individuals including 724 individuals from 72 distinct SA population groups. For comparison we used the HRC reference panel or the 1000 Genomes Phase3 (1000G) callset. Imputation with 1000G recovered only 59% of the European CARDIoGRAMplusC4D model variants and was not used to generate a GPS. Imputation with the GAsP reference panel recovered the high proportion (90%) of the variants in the CARDIoGRAMplusC4D GWAS summary. GPS were generated by LDPred and optimized with SA cases and controls from UKBiobank data (630 cases, 6,614 controls). Using the GAsP reference panel, we obtained an odds-ratio of 1.72 for those in the top decile compared with the median GPS. We further observed that the early onset individuals (age  $\leq 45$  years) had 3-fold higher median GPS ( $p < 0.05$ ) as compared to

those with later events. Our study shows that a properly calibrated and validated GPS can be an extremely useful clinical tool for predicting CAD risk in SA groups.



## PgmNr 2645: Transcriptomic changes of individual peripheral blood mononuclear cells in mild stroke patients.

### Authors:

H. Kim <sup>1</sup>; E. Cho <sup>1</sup>; S. Yun <sup>2</sup>; A. Cashion <sup>1</sup>; H. Lee <sup>3</sup>; L. Latour <sup>4</sup>; J. Gill <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) SML, NINR/NIH, Bethesda, Maryland.; 2) Yotta BioMed, Bethesda, Maryland; 3) Yonsei University, College of Nursing, Seoul, Korea; 4) NINDS/NIH, Bethesda, Maryland

---

**Purpose:** Minor and transient ischemic strokes are easily missed with current diagnostic tools. Its diagnosis may be improved by peripheral blood biomarkers. In this study, we examined peripheral blood mononuclear cells (PBMCs) transcriptome profile at the single cell level as the potential biomarker for the minor strokes.

**Methods:** Peripheral blood samples were collected from acute minor ischemic stroke patients (National Institutes of Health Stroke Scale <5) within 24hr after stroke onset and age/gender matched controls. Transcriptome libraries of individual PBMCs were prepared and sequenced.

**Results:** Based on single cell RNA-seq data from 12,418 cells in 11 minor stroke patients, 12 cell type clusters were identified, while 13 cell type clusters were identified from 12,285 cells in 8 controls. Distribution of cell clusters was significantly different between patient and control groups. Cell type clusters commonly identified from both groups were compared to obtain differentially regulated genes. From 9 common clusters, 137 differentially regulated genes were found, and S100A9 from B cells cluster was the most upregulated gene in stroke patients.

**Conclusion:** As the first pilot study of transcriptome changes after minor stroke in a single cell level, it might provide a guidance for the future biomarker study.

## PgmNr 2646: Insights into the genetic architecture of spontaneous coronary artery dissection through sequencing 392 cases and >13,000 UK Biobank controls.

### Authors:

K. Carss <sup>1</sup>; J. Armisen <sup>1</sup>; A. Baranowska <sup>2</sup>; T. Webb <sup>2</sup>; S. Hamby <sup>2</sup>; S. Petrovski <sup>1</sup>; N. Samani <sup>2</sup>; C. Haefliger <sup>1</sup>; D. Adlam <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) AstraZeneca Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK; 2) Leicester NIHR Biomedical Research Centre and Department of Cardiovascular Sciences, University of Leicester, UK

---

Spontaneous coronary artery dissection (SCAD) occurs when one of the coronary arteries is occluded by a tear in the lining of the arterial wall and/or a haematoma in the vessel wall that is not caused by trauma, atherosclerosis or medical treatment. It can cause acute coronary syndrome, myocardial infarction, and sudden death. The prevalence of SCAD has been reported as 0.07% to 1.1%, and it is likely to be significantly underdiagnosed. Pathogenic variants in genes associated with connective tissue disorders have previously been implicated in the aetiology of SCAD, however, very little is known about the genetic architecture of this condition.

We have whole-genome sequenced a cohort of 392 individuals with SCAD with the aim of discovering novel variants and genes associated with SCAD and improving understanding of phenotypic heterogeneity. A multi-pronged approach is being taken to analysis. First, we have identified pathogenic variants in genes previously implicated in SCAD including protein-truncating variants in *COL3A1* (associated with vascular Ehlers-Danlos syndrome), *SMAD3* (associated with Loeys-Dietz syndrome type III), and *PKD1* (associated with autosomal dominant polycystic kidney disease, a known risk factor for SCAD). Second, gene-level collapsing analysis has been performed to identify genes enriched for rare variants in cases compared to >13,000 controls from the UK Biobank. These controls have had WES and have no reported cardiovascular disease or connective tissue disease. We found no gene to be study-wide significantly associated with SCAD. Several genes previously implicated in SCAD or connective tissue disorders rank highly in this experiment including *COL3A1* ( $p = 0.0013$ ), *PKD1* ( $p = 0.0016$ ) and *TGFB2* ( $p = 0.0024$ ). Further analyses of these data, including follow up of potential novel candidate genes, are ongoing and will be presented.

Additional ongoing analyses include variant and pathway level association studies, analysis of structural variants, genotype-phenotype correlation analyses, and candidate gene prioritisation using a machine learning approach. This is the largest SCAD cohort to have undergone whole-genome sequencing, and alongside sequenced and deeply phenotyped UK Biobank controls, provides rich opportunity for novel discoveries.

## PgmNr 2647: Whole exome sequencing analysis elucidates genetic architectures of idiopathic ventricular fibrillation.

### Authors:

P.C. Cha <sup>1</sup>; A. Takahashi <sup>1</sup>; S. Ohno <sup>2</sup>; T. Kamakura <sup>3</sup>; T. Ishikawa <sup>4</sup>; H. Ueda-Ishibashi <sup>5</sup>; M. Shiba-Harada <sup>6</sup>; N. Minamino <sup>4</sup>; K. Hayashi <sup>7</sup>; N. Yagihara <sup>8</sup>; Y. Nakano <sup>9</sup>; H. Itoh <sup>10</sup>; Y. Miyamoto <sup>11</sup>; T. Tanaka <sup>12</sup>; N. Makita <sup>4</sup>; M. Horie <sup>10</sup>; W. Shimizu <sup>13</sup>; K. Kusano <sup>3</sup>; S. Yasuda <sup>3</sup>; T. Aiba <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genomic Medicine, Research Institute, National Cerebral and Cardiovascular Center (NCVC); 2) Department of Biosciences and Genetics, Research Institute, NCVC; 3) Department of Cardiovascular Medicine, NCVC; 4) Omics Research Center, NCVC; 5) Biobank, NCVC; 6) Department of Molecular Innovation in Lipidology, Research Institute, NCVC; 7) Department of Cardiology, Kanazawa University Graduate School of Medicine; 8) Department of Cardiovascular Biology and Medicine, Niigata University Graduate School of Medical and Dental Sciences; 9) Department of Cardiology, Hiroshima University; 10) Department of Cardiovascular and Respiratory Medicine, Shiga University of Medical Science; 11) Department of Preventive Cardiology, NCVC; 12) Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University; 13) Department of Cardiovascular Medicine, Nippon Medical School

**Introduction:** Idiopathic ventricular fibrillation (IVF) is one of the major causes of sudden cardiac death in younger subjects who have no structural heart disease or identifiable inherited primary arrhythmic syndromes. The genetic background of IVF is still unclear.

**Methods:** Whole exome sequencing on 184 subjects (39±19 years, 71% male) who had been resuscitated from VF without any structural heart diseases was conducted. We aligned short read sequences to human reference genome (build38) by using two methods, namely BWA (0.7.15) and Bowtie2 (2.3.2), and performed variant calling by GATK (3.8.1) haplotype caller. To increase accuracy, we annotated (by ANNOVAR) and analyzed only single nucleotide variants (SNVs) commonly detected by both mapping methods and passed standard quality control.

**Results:** A total of 55,107 SNVs, which either affect splicing events or reside in exons of genes/non-coding RNAs, have reported minor allele frequencies (MAF) of  $\leq 0.01$  in publicly available East Asian and Japanese population databases (including the 1000 Genomes, ExAC, Gnomad, HGVD, and ToMMo). 1,581 among these rare SNVs were residing in the 295 genes potentially associated with heritable heart and vascular conditions (Invitae Cardiology Gene Panel). Based on ACMG-AMP Guidelines for clinical interpretation of variants (by InterVar software), we identified 27 pathogenic or likely pathogenic (PLP) SNVs in cardiovascular diseases-related genes in 15% of the IVF subjects (28/184). The proportion of PLP SNVs in cardiovascular diseases-related genes was significantly higher in the 184 IVF patients (27/1581) compared to the 1210 Japanese subjects in the HGVD database (34/3624) ( $P=0.0256$ ). Furthermore, many of the PLP SNVs in IVF patients could be observed in the arrhythmia-related genes (10/480 in IVF vs 7/1366 in HGVD;  $P=0.0048$ ) and cardiomyopathy and muscular diseases-related genes (23/1048 in IVF vs 24/2456 in HGVD;  $P=0.0068$ ).

**Conclusion:** In IVF patients, 15% had some PLP SNVs in genes associated with cardiovascular diseases, but many of these were not directly related to lethal ventricular arrhythmias but rather potential genes associated with cardiomyopathies and/or neuromuscular diseases.

## PgmNr 2648: Association of a genome-wide polygenic risk score for low-density lipoprotein cholesterol and coronary artery disease risk.

### Authors:

K. Chaudhary; G.N. Nadkarni; R. Do

[View Session](#) [Add to Schedule](#)

**Affiliation:** The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York.

---

Coronary artery disease (CAD) is the leading cause of death and disability in the world. Low-density lipoprotein cholesterol (LDL-C) is a causal risk factor for CAD. Recently, a genetic study showed that polygenic risk scores (PRS) associated with CAD can be predictive of CAD. In the present study, we assessed whether PRS associated with LDL-C can be predictive of LDL-C and CAD in individuals from the UK Biobank. We calculated the PRS for LDL-C using summary statistics from Global Lipids Genetics Consortium and calculated the PRS from different sets of SNVs and parameters from the LDpred method in a training dataset of 50,000 self-identified white British. In this analysis, we observed that the optimal parameters, as measured by the variance explained by the LDL-C PRS on LDL-C, was LDpred with  $\rho=0.3$ . We calculated the LDL-C PRS using  $\rho=0.3$  on a test dataset of 324,551 separate individuals also from the UK biobank. We observed consistent strong associations of the LDL-C PRS on LDL-C in White British ( $\beta=0.255$  [0.25 - 0.26],  $P<2.11 \times 10^{-16}$ ) and Non-White British Europeans ( $\beta=0.248$  [0.24 - 0.26],  $P<2.11 \times 10^{-16}$ ) with weaker significant associations in Africans ( $\beta=0.167$  [0.14 - 0.19],  $P=5.53 \times 10^{-34}$ ), East Asians ( $\beta=0.185$  [0.13 - 0.24],  $P=2.11 \times 10^{-11}$ ), and South Asians ( $\beta=0.177$  [0.15 - 0.2],  $P=1.22 \times 10^{-50}$ ). Overall, in a meta-analysis of all ancestries, we observed association of the LDL-C PRS on LDL-C ( $\beta=0.211$  [0.18 - 0.24],  $P=1.09 \times 10^{-48}$ ). Similar trends were observed with the LDL-C PRS on CAD across the different ancestries. Furthermore, individuals in the top 1% of the LDL-C PRS had 1.12 SD units increased LDL-C and 1.6-fold increased risk for CAD compared to the bottom 10% of individuals. Plasma LDL-C was observed to be elevated in the top 1% of individuals with the LDL-C PRS (median= $150 \pm 38$  mg/dL) as compared to the bottom 10% (median= $120 \pm 30$  mg/dL). We expect to replicate these results in the multiethnic clinical care cohort, BioMe, and assess whether individuals with high LDL-C PRS are being adequately recognized for their inherent genetic risk and whether risk modification strategies are being employed.

## PgmNr 2649: Whole-exome sequencing in 249 transposition of the great arteries patients unveils central role of cilia related etiology.

### Authors:

W. Chen; X. Liu; W. Li; B. Han; X. Liu; K. Zhao; Y. Fu; Z. Zhou

[View Session](#) [Add to Schedule](#)

**Affiliation:** Fuwai Hospital Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, Beijing, China

### BACKGROUND

Congenital heart diseases (CHDs) are the most common types of birth defects, affecting approximately 1% of live births and remaining the leading cause of mortality. Transposition of the great arteries (TGA) is a kind of most severe life-threatening congenital heart disease. There was an increased recurrence risk among sporadic TGA cases suggest that genetic factors lead an important role underlying this disease. However, the exact etiology of TGA remains unclear.

### METHODS

We performed whole exome sequencing in 249 non-syndromic TGA samples to quantify the genetic burden and causal factors. 189 obesity cases with normal cardiac phenotype were recruited as a control cohort and another group of 130 TGA patients as a replication cohort. The pathogenicity of de novo variants, loss of function genes, compound heterozygous variants and X-linked variants in CHD related gene and pathway were considered using a semi-automatic automatic bioinformatics filtering process. Loss of function genes were then performed functional enriched analysis.

### RESULTS

Eleven de novo variants, 62 genes with rare loss of function (LOF) variants, 4 genes with compound heterozygous variant and 8 X-linked variants were identified using our analytical strategy. Cilia related genes were occupied a major position in our results: it showed up as strong pathogenic genes in de novo, X-linked and compound heterozygous condition. Notably, ciliopathies related gene *DYNC2LI1* encoding a component of the dynein-2 complex shown a significantly high risk of TGA disease (OR: 2.51; 95% [CI]:1 to 5.22; p value: 0.02). Overall, except one mutation: p.G241D, the rest 4 mutation: p.D80E, p.D80A, p.L130P, p.R232H were located in a critical functional domain. Using sanger sequencing technique, we found 2 more mutation a replication cohort of TGA (n=130) on *DYNC2LI1* gene at the same chromosome loci. Even more interesting, cilium assembly or organization pathway was significantly enriched in functional annotation and enrichment analysis (P-value<0.05).

### CONCLUSIONS

Among the patients in our study, TGA related causal genes arose from in de novo, LOF, compound heterozygous and X-linked condition. A mount numbers of variants affecting genes involved in cilium assembly or organization that exacerbate the widespread genetic susceptibility associated with TGA. The results reveal extensive genetic heterogeneity underlying TGA, and posit potential novel strategy of genetic investigation for rare disease.

## PgmNr 2650: Untargeted lipidomic profiles as endophenotypes for CVD risk.

### Authors:

J.E. Curran <sup>1</sup>; N.B. Blackburn <sup>1</sup>; C. Giles <sup>2</sup>; K. Huynh <sup>2,3</sup>; G.E. Reid <sup>4</sup>; A.C. Leandro <sup>1</sup>; J.M. Peralta <sup>1,5</sup>; J. Blangero <sup>1</sup>; P.J. Meikle <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics and South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley, Brownsville, TX, USA.; 2) Baker Heart and Diabetes Institute, Melbourne, VIC, Australia.; 3) Monash University, Melbourne, VIC, Australia.; 4) University of Melbourne, Melbourne, VIC, Australia.; 5) Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia.

---

Cardiovascular disease (CVD) is the leading cause of mortality worldwide. Risk is attributable to several factors including blood pressure, smoking, increasing age, sex, and genetics. Although CVD risk is heritable and there have been a number of successes localizing QTLs influencing disease risk, the genetic basis is still relatively unknown. Quantitative endophenotypes that are related to disease liability can offer more power for gene localization than dichotomous disease status and thus serve as valuable endophenotypes for disease gene identification. Lipids and their constituent components are important factors in CVD risk. The human lipidome is the total complement of these lipid species and represents a wealth of endophenotypes that may influence CVD risk. To identify CVD endophenotypes, we performed untargeted plasma lipidomic profiling in 500 individuals from 6 extended pedigrees of the San Antonio Family Study (SAFS). Untargeted lipidomic profiling was performed by liquid chromatography-ultrahigh resolution mass spectrometry and conditions optimized for separation of both polar and non-polar lipids. Quantitative measurements were obtained in positive and negative ion modes using MS analysis, and characterization was achieved by leveraging off our existing targeted lipidomic platform and a combination of DDA/DIA MS<sup>2</sup> and MS<sup>3</sup> acquisitions. Prior to processing, the untargeted method identified 135,414 features with less than 10% missing values across all individuals. To identify genetic factors influencing levels of each of these lipid features, we performed variance-component based heritability analyses, using age, sex and their interactions as covariates. Using a very conservative Bonferroni threshold, 72,184 (~50%) lipid features were significantly heritable, with heritabilities of 20-80%. We examined whether variation in these heritable features was associated with future CVD. The 10 most significantly associated features ( $p=7.95 \times 10^{-6}$  to  $p=4.04 \times 10^{-5}$ ) include lipid species in 5 classes with changes of ~0.7 SDUs observed in individuals who develop CVD. The heritability of each of these features was greater than 40% indicating a significant genetic component to variation. These preliminary results provide strong evidence that using an untargeted approach will identify many more lipid species associated with CVD than targeted only profiling methods thus providing a powerful approach for identification of endophenotypes influencing CVD risk.

## PgmNr 2651: Polygenic risk scores substantially improve risk stratification based on traditional non-genetic risk factors.

### Authors:

P. Donnelly; F. Aguilera; M. Sivley; E. Kraphol; V. Plagnol

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genomics plc

---

There has been considerable recent interest in the use of polygenic risk scores (PRS) to identify individuals at increased risk of common diseases. Early risk stratification offers the potential for targeted screening or interventions. Much attention to date has focussed on the use of PRS as standalone tools. But for many diseases clinicians already assess risk based on non-genetic risk factors. A key question is the extent to which PRS add to current risk assessments.

We address this question using coronary artery disease (CAD) because of its major impact on mortality and because there are widely-used risk assessment tools based on non-genetic risk factors, including age, gender, lipid levels, and blood pressure.

For definiteness we consider the American Heart Association (AHA) risk score. We use a CAD PRS developed internally whose performance is comparable to, and in some ways improves on, other published CAD PRS. A subset ( $n = 100K$ ) of UK Biobank is used as training to optimise the AHA/PRS combination. Remaining samples are used to compare the performance of the AHA score with the combined AHA/PRS score.

The CAD PRS is uncorrelated with the AHA non-genetic score ( $r = 0.01$ ), hence clearly provides additional information. The CAD PRS is also almost uncorrelated with family history ( $r = 0.09$ ). The combination of AHA and PRS provides a substantial improvement in classification for CAD (net reclassification improvement of 0.36). Next, we considered the binary classification problem which approximates current guidelines used to determine whether statin prescription is recommended (broadly equivalent to AHA non-genetic score  $> 20\%$ ). Firstly, because the combination of non-genetic and PRS scores improves risk stratification, substantially more individuals ( $\sim 2$ -fold) are above the relevant absolute risk level using the combined score. Secondly, the set of individuals above this level (who would thus be recommended statins) is different: 5.4% of the UK Biobank population not classified at high risk is reclassified as high risk using the combined score. Critically, these individuals meet guideline levels of risk for statin prescription but are not currently identified as such. In addition, 11% of the individuals classified as high risk using AHA score only have medium levels of risk when PRS is incorporated. These data show that CAD risk prediction provides a potential first application for the integration of PRS with traditional risk factors in clinical care.



## PgmNr 2652: Testing of a genome-wide polygenic score for coronary artery disease in individuals with ancestries outside of Europe.

### Authors:

A.C. Fahed<sup>1,2</sup>; G. Hindy<sup>1,2</sup>; A.V. Khera<sup>1,2</sup>; K.G. Aragam<sup>1,2</sup>; Y.I. Chen<sup>6</sup>; K. Chaudhary<sup>3</sup>; H. Krumholz<sup>4</sup>; W.H. Sheu<sup>8</sup>; S.S. Rich<sup>5</sup>; J.I. Rotter<sup>6</sup>; J. Danesh<sup>7</sup>; R. Do.<sup>3</sup>; S. Kathiresan<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA.; 2) Division of Cardiology and Center of Genomic Medicine, Massachusetts General Hospital, Boston, MA; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Center for Outcomes Research and Evaluation, Yale-New Haven Hospital, CT; 5) Center of Public Health Genomics, University of Virginia, Charlottesville, VA; 6) Institute for Translational Genomics and Population Sciences, Harbor-UCLA Medical Center, Los Angeles, CA; 7) Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 8) Division of Endocrinology and Metabolism, Taichung Veterans General Hospital, Taichung 407, Taiwan

---

We recently developed and validated a genome-wide polygenic score (GPS) for coronary artery disease (GPS<sub>CAD</sub>) in individuals of European ancestry. Whether this GPS can generalize to other ancestries is unknown.

We used imputed genotyping data or whole-genome sequencing data to calculate GPS<sub>CAD</sub> in 469,964 individuals (26,476 cases) of European ancestry and 37,977 individuals (3,916 cases) of African (1,371 cases), South Asian (873 cases), East Asian (435 cases), and Hispanic (1,237 cases) ancestries across 6 multi-ethnic cohorts from 4 countries. Those included the UK Biobank, hospital-based case-control studies (VIRGO, TaiChi Consortium, and BRAVE), and hospital-based biobanks (BioMe and Partners Biobank). Self-report ancestry was validated by principle component analysis to 1000 Genomes Project data. We measured the strength of GPS<sub>CAD</sub> in predicting CAD in each ancestry of each cohort, using a logistic regression model adjusted for the first four principle components of ancestry. We then performed a meta-analysis for similar ancestries across cohorts to compare the predictive power of GPS<sub>CAD</sub> across different ancestries.

The strength of association of GPS<sub>CAD</sub> in European ancestry was as reported previously, with an odds ratio per standard deviation (OR<sub>SD</sub>) of 1.60 (1.45-1.75) for CAD. The European GPS<sub>CAD</sub> predicted CAD reasonably well in East Asian (OR<sub>SD</sub> = 1.66 [1.45-1.88]), South Asian (OR<sub>SD</sub> = 1.47 [1.36-1.58]), and Hispanic (OR<sub>SD</sub> = 1.53 [1.44-1.65]) ancestries. The score remained predictive, but had a reduced strength, in individuals of African ancestry (OR<sub>SD</sub> = 1.25 [1.14-1.37]). Within each ancestry, variability in the Area Under the Curve (AUC) and OR<sub>SD</sub> was confounded with the study design and case definition.

A polygenic score of CAD developed in individuals of European ancestry is generalizable to non-European ancestry in diverse settings but shows weakest strength of association in African ancestry.



## PgmNr 2653: Coronary artery disease polygenic risk score is associated with subclinical carotid atherosclerosis measures: The Long-Life Family Study (LLFS).

### Authors:

M.F. Feitosa<sup>1</sup>; A.L. Kuipers<sup>2</sup>; M.K. Wojczynski<sup>1</sup>; L. Wang<sup>1</sup>; E. Barinas-Mitchell<sup>2</sup>; A.M. Kulminski<sup>3</sup>; B. Thyagarajan<sup>4</sup>; L.J. Lee<sup>5</sup>; T. Perls<sup>6</sup>; K. Christensen<sup>7</sup>; A.B. Newman<sup>2</sup>; J.M. Zmuda<sup>2</sup>; M.A. Province<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St Louis, Missouri, USA; 2) Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA, USA; 3) Social Science Research Institute, Duke University, Durham, NC, USA; 4) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA; 5) Sergievsky Center, Taub Institute, Departments of Epidemiology and Neurology, Columbia University, New York, NY, USA; 6) Department of Medicine, Boston University School of Medicine, Boston, MA, USA; 7) Department of Epidemiology, Biostatistics and Biodemography, Danish Aging Research Center, University of Southern Denmark, Odense, Denmark

Polygenic risk scores (PRS) for coronary artery disease (CAD) may quantify an individual's risk for developing atherosclerosis. Carotid intima-media thickness (IMT) and carotid plaque measures are associated with coronary artery atherosclerotic lesions. We evaluated whether the PRS for CAD is associated with carotid IMT, plaque severity and plaque presence, assessed using high resolution B-mode ultrasonography in up-to 2405 European ancestry individuals from 499 families recruited for exceptional longevity (age:  $72.1 \pm 11.7$ , range: 42-110 years old). To generate the PRS, 177 independent SNPs ( $r^2 < 0.2$ ) identified from genome-wide association studies (GWAS,  $p < 5.0 \times 10^{-8}$ ) for CAD were aligned to their risk alleles. The PRS for each participant was computed using the allele count of 177-SNPs multiplied by CAD effect size ( $\beta$ -coefficients or odds ratio using natural-logarithm (log) transformation). The products of allele counts and  $\beta$ -coefficients were summed over all 177 CAD SNPs for the total PRS (mean  $\pm$  S.E. =  $11.08 \pm 0.48$ ). Carotid IMT and plaque severity (ranging from 0 (no) to 30 (highest) plaque index +1), were log transformed and adjusted for age, sex, field centers, cardiovascular risk factors (smoking, diabetes, systolic blood pressure, hypertension, low-density lipoprotein, high-density lipoprotein, and waist circumference) and accounting for population stratification (principal components). Then, linear mixed effects models accounting for kinship (lme4 function in R-package) were employed in residuals of carotid IMT and plaque severity with PRS as the independent variable. A generalized estimating equations approach (SAS® 9.4) was used for plaque presence (Yes/No) and included the PRS, cardiovascular risk factors and principal components in the model. The CAD PRS was significantly associated with carotid IMT ( $\beta \pm$  S.E. =  $0.140 \pm 0.046$ ,  $p = 2.23 \times 10^{-3}$ ), plaque severity ( $\beta \pm$  S.E. =  $0.191 \pm 0.044$ ,  $p = 1.48 \times 10^{-5}$ ) and plaque presence ( $\beta \pm$  S.E. =  $0.387 \pm 0.105$ ,  $p = 2.17 \times 10^{-4}$ ) considering Bonferroni correction (0.05/3 phenotypes,  $p = 0.0167$ ). Our findings show that CAD PRS is associated with early markers of atherosclerosis even among adults selected for exceptional survival independent of established cardiovascular disease risk factors. Patients identified by high polygenic risk scores may ultimately benefit from prevention strategies, including early lifestyle change and therapeutic interventions, to help modify the trajectory of CAD.

## PgmNr 2654: Phenome-wide association of functional coding variants and cardiometabolic phenotypes in American Indians.

### Authors:

N. Franceschini <sup>1</sup>; Y. Shan <sup>7</sup>; K. Haack <sup>2</sup>; P.E. Melton <sup>3,4</sup>; S. Sanna <sup>5</sup>; S.A. Cole <sup>2</sup>; Y. Li <sup>6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, Univ North Carolina at Chapel Hill, Chapel Hill, NC; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) The Curtin UWA Centre for Genetic Origins of Health and Disease, Faculty of Health Sciences, Curtin University and Faculty of Health and Medical Sciences, The University of Western Australia, Crawley, Western Australia; 4) School of Pharmacy and Biomedical Sciences, Faculty of Health Sciences, Curtin University, Bentley, Western Australia; 5) Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands; 6) Departments of Genetics and Computer Science, University of North Carolina, Chapel Hill, NC; 7) Department of Biostatistics, University of North Carolina, Chapel Hill, NC

---

There has been a great interest in phenotyping individuals with predicted protein-altering functional variants (rare loss of function and missense variants) to understand the health effects in populations. In our ongoing genetic investigation in American Indians, we recently genotyped 1,127 participants from the Strong Heart Family Study, in a case-control study of chronic kidney disease, for single nucleotide variants (SNV) and small indels identified by whole exome sequencing in a subset of individuals and that had predicted protein-altering function. Among 1,206 SNVs and indels that passed quality control (average minor allele count [MAC] = 20, range of 1 to 1,064), ~43% were not present in publicly available repositories, and may be specific to American Indians. Most of the novel coding variants were missense SNVs/indels (n=228), frameshift substitutions (n=73), stop-gain or stop-loss variants (n=8), and splice acceptor/donor (n=11). We performed association analyses of SNVs/indels with a MAC>10 for 32 cardiometabolic biomarkers, using variance component models to account for relatedness, while adjusting for age, sex, center and case-control status. Using a multiple-testing adjusted significance threshold of  $p < 5.5 \times 10^{-6}$  to account for 9,122 effective independent tests including 32 traits tested, we identified 11 trait-variant associations. For example, we detected a missense SNV at *ABCA10* (p.G1369W, MAC=17,  $p=8 \times 10^{-9}$ ) associated with increased fasting triglycerides and explaining 2.5% of trait variability. *ABCA10* protein is involved in macrophage lipid homeostasis suggesting that it is a cholesterol-responsive gene. A missense SNV located at *TRPM3* (p.V1249M, MAC=185,  $p=5 \times 10^{-8}$ ) was associated with lower fasting insulin concentration and accounted for 1.7% of serum insulin variance. *TRPM3* protein is a non-selective cation channel expressed in pancreatic  $\beta$ -cells, which has been shown to regulate insulin secretion. Additional findings include a novel missense SNV at *EXTL2* (MAC=23,  $p=9 \times 10^{-9}$ ) associated with serum creatinine, and a missense SNV at *PNPLA5* (MAC=13,  $p=3 \times 10^{-7}$ ) associated with increased HbA1c. In conclusion, our study focused on predicted damaging coding variants in American Indians identified new gene associations with cardiometabolic phenotypes, demonstrating the advantages of strategies that leverage whole exome sequencing findings to select predicted functional variants for association screenings in less genetic characterized populations.

## PgmNr 2655: Oligogenic inheritance of a human heart disease involving a genetic modifier.

### Authors:

C. Gifford<sup>1,2</sup>; S. Ranade<sup>1,2</sup>; R. Samarakoon<sup>1,2</sup>; K. Ivey<sup>1,2,3</sup>; D. Srivastava<sup>1,2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Gladstone Institute of Cardiovascular Disease, Gladstone Institutes, UCSF, San Francisco, California; 2) Roddenberry Stem Cell Center at Gladstone, San Francisco, California; 3) Department of Pediatrics, UCSF, San Francisco, California; 4) Department of Biochemistry and Biophysics, UCSF, San Francisco, California

---

Complex genetic mechanisms are thought to underlie many human diseases, yet experimental proof of this model has been elusive. Here, we show that a human cardiac anomaly can be caused by a combination of rare, inherited heterozygous mutations. Whole-exome sequencing of a nuclear family revealed that three offspring with childhood-onset cardiomyopathy had inherited three missense single-nucleotide variants in the *MKL2*, *MYH7*, and *NKX2-5* genes. The *MYH7* and *MKL2* variants were inherited from the affected, asymptomatic father and the rare *NKX2-5* variant (minor allele frequency, 0.0012) from the unaffected mother. We used CRISPR-Cas9 to generate mice encoding the orthologous variants and found that compound heterozygosity for all three variants recapitulated the human disease phenotype. Analysis of murine hearts and human induced pluripotent stem cell-derived cardiomyocytes provided histologic and molecular evidence for the *NKX2-5* variant's contribution as a genetic modifier.

## PgmNr 2656: RYR2 and KCNQ1 ion channel single nucleotide polymorphisms are contributing to myocardial infarction susceptibility in middle-aged Mexican men.

### Authors:

R. Gomez <sup>1</sup>; L. Lopez <sup>1</sup>; M.deL. Lopez G <sup>1</sup>; C. Santana <sup>2</sup>; G. Noris <sup>2</sup>; A. de la Pena <sup>3,4</sup>; M. Flores-Garcia <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) CINVESTAV-IPN, Departamento de Toxicología, Mexico City, Select a Country; 2) Biología Molecular Diagnóstica S.A. de C.V., Queretaro, Qro., Mexico; 3) Facultad de Medicina, Departamento de Farmacología, Universidad Nacional Autónoma de México; 4) Departamento de Biología Molecular, Instituto Nacional de Cardiología, Mexico City, Mexico

---

Cardiovascular disorders (CVD) are the leading cause of death in low and middle-income countries where more than three-quarter of CVD caused deceased occur. Of these, ischaemic heart disease contributes to 85% of deaths, exhibiting a higher trend in middle age men (18 – 50 years). Genome-wide association studies have identified common risk-variants related to CVD, which frequencies have exhibited multi-ethnic variations among populations. Next-generation sequencing is a powerful tool to study and discover rare gene variants contributing to complex disease development. Herein, sixteen young men (range: 26 - 46 years;  $41.05 \pm 4.71$ ) and their controls were studied by sequencing 174 genes, (TruSight Cardio Sequencing kit, Illumina®) with Illumina Miseq System, to identify their genetic contribution to acute myocardial infarction risk. Quality control analysis guaranteed reliable results, which were evaluated to identify those variants with potential to cause disease using ten *in silico* algorithms. Our results suggested the contribution of two SNP in ion channel genes: ryanodine receptor 2 (RYR2; rs561321743; H877P), and potassium voltage-gated channel subfamily Q member 1 (KCNQ1; rs199472800; R555H). R555H has been related to abnormal potassium channel functions. This variant was also found in affected individuals with long QT syndrome, reinforcing our findings. H877P has been considered as a gene variant likely benign, but it is feasible that this it ion channel single nucleotide polymorphism could be in linkage disequilibrium with causal variants given that RYR2 polymorphisms have been associated with ventricular arrhythmias risk. Hitherto, the allele frequencies of these two ion-channel single nucleotide polymorphisms are unknown in the Mexican population, making further studies necessary to reveal their clinic implication.

## PgmNr 2657: Systematic whole-exome sequencing in unexplained cardiac arrest: Results from the Canadian CASPER registry and biobank.

### Authors:

S. Grondin <sup>1</sup>; B. Davies <sup>2</sup>; C.C. Cheung <sup>2</sup>; C. Steinberg <sup>3</sup>; J.S. Healey <sup>4</sup>; M.S. Green <sup>5</sup>; S. Sanatani <sup>6</sup>; P. Angaran <sup>7</sup>; H. Duff <sup>8</sup>; L. Arbour <sup>9</sup>; R. Leather <sup>9</sup>; C.S. Simpson <sup>10</sup>; M. Gardner <sup>11</sup>; C. Seifer <sup>12</sup>; G. Lettre <sup>1</sup>; M. Talajic <sup>1</sup>; Z.W. Laksman <sup>2</sup>; J.D. Roberts <sup>13</sup>; A.D. Krahn <sup>2</sup>; R. Tadros <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, University of Montreal, Research Center, Montreal Heart Institute, Montreal, Quebec, Canada; 2) University of British Columbia, Vancouver, British Columbia, Canada; 3) Institut Universitaire de Cardiologie et Pneumologie de Québec, Quebec City, Quebec, Canada; 4) Population Health Research Institute, McMaster University, Hamilton, Ontario, Canada; 5) University of Ottawa Heart Institute, Ottawa, Ontario, Canada; 6) British Columbia Children's Hospital, Vancouver, British Columbia, Canada; 7) St. Michael's Hospital, Toronto, Ontario, Canada; 8) Libin Cardiovascular Institute of Alberta, University of Calgary, Calgary, Alberta, Canada; 9) Royal Jubilee Hospital, Victoria, British Columbia, Canada; 10) Queen's University, Kingston, Ontario, Canada; 11) Queen Elizabeth II Health Sciences Center, Halifax, Nova Scotia, Canada; 12) St. Boniface Hospital, Winnipeg, Manitoba, Canada; 13) Western University, London, Ontario, Canada

---

**Introduction:** Ventricular fibrillation (VF) in the absence of coronary disease or left ventricular (LV) dysfunction often remains unexplained. Given its heritability, identifying the genetic culprit is important for family screening. **Hypothesis:** Systematic whole-exome sequencing (WES) identifies a pathogenic or likely pathogenic (P/LP) variant in a significant proportion of unexplained cardiac arrest (UCA) survivors. **Methods:** We performed WES in consecutive unrelated survivors of UCA, defined as VF in the absence of overt structural or electrical disease. Joint variant calling was performed as per the GATK best practice guidelines. For the present study, we limited our analysis to rare (allele frequency, AF <1/10,000 in the Genome Aggregation Database, gnomAD) protein-coding variants in 60 genes unequivocally associated with cardiac arrhythmia, cardiomyopathy or systemic disease with a cardiac manifestation. Variant classification was performed according to the American College of Medical Genetics and Genomics (ACMG) guidelines. **Results:** WES was performed in 234 UCA probands (66% males; age at arrest 38±13). WES identified a total of 17.6M high quality genetic variants, of which 172 were rare variants in 60 cardiac disease genes (mean 0.7 variants/case). We observed a significant enrichment of rare missense RYR2 variants in UCA compared to gnomAD (combined AF 4.1% vs. 2.0%; P=0.004) as well as an enrichment of rare truncating variants in cardiomyopathy genes (2.8% vs. 1.1%; P=0.003). Of the 234 probands, 30 (12.8%) had ≥1 P/LP variants as per ACMG criteria. Phenotype-driven clinical genetic testing prior to research WES only identified the P/LP variant in 12 (40%) of the 30 cases, suggesting an added value of systematic genetic testing compared to the current standard of phenotype-driven testing. **Conclusions:** Systematic comprehensive genetic testing of cardiac genes identifies a disease-causing variant in 13% of unexplained cardiac arrest survivors. Truncating variants in cardiomyopathy genes are enriched in UCA supporting an association of structural genes with ventricular arrhythmias in the absence of overt cardiomyopathy. Further studies are needed to understand the implications of these

findings in cascade family screening.

## PgmNr 2658: Multi-ancestry whole genome sequencing study of carotid intima media thickness and carotid plaque.

### Authors:

N.R. Hasbani <sup>1</sup>; J.G. Broome <sup>2</sup>; J.G. Terry <sup>3</sup>; N.D. Palmer <sup>4</sup>; J.C. Bis <sup>5</sup>; H. Xu <sup>6</sup>; J.E. Curran <sup>7</sup>; L.A. Cupples <sup>8</sup>; A. Correra <sup>9</sup>; D. Herrington <sup>10</sup>; R.S. Vasan <sup>8</sup>; W.S. Post <sup>11</sup>; on behalf of the TOPMed Atherosclerosis Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Human Genetics, and Environmental Sciences, University of Texas School of Public Health, Houston, TX.; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Department of Radiology, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Medicine, University of Washington, Seattle, WA; 6) Department of Medicine, University of Maryland, School of Medicine, Baltimore, MD; 7) Department of Human Genetics, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX; 8) Department of Epidemiology, Boston University School of Public Health, Boston, MA; 9) Departments of Medicine and Pediatrics, University of Mississippi Medical Center, Jackson, MS; 10) Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem NC; 11) Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD

---

Carotid intima-media thickness (CIMT) and presence of carotid plaque, evaluated non-invasively using ultrasound, are markers of atherosclerosis and useful in detecting subclinical cardiovascular disease. The largest published genome-wide association study (GWAS) found 11 CIMT and 5 plaque loci. Utilizing whole genome sequencing (WGS) generated by the National Heart, Lung, and Blood Institute's Trans-Omics for Precision Medicine (TOPMed) program, we sought to investigate potentially novel genetic associations with CIMT and carotid plaque that may have been missed by GWAS based on genotyping arrays. WGS data was available for 22,443 participants from 9 studies who had CIMT measured and 20,254 participants from 7 studies who had carotid plaque evaluated. Participants were of European, African-American, Hispanic, or East Asian ancestry. CIMT was transformed through inverse rank-based normalization. Linear and logistic mixed models accounting for relatedness, implemented in SAIGE, were used to perform single variant tests for CIMT and plaque, respectively. Analyses were adjusted for age, sex, study, ancestry group, and principal components and limited to variants with a minor allele frequency (MAF) >0.1%. Using a significance threshold of  $P < 5 \times 10^{-8}$ , we identified two loci (*APOE* and *CCBE1*) associated with CIMT and one locus (*HNF4G*) with plaque. While the *APOE* locus has been previously linked to CIMT, *CCBE1* and *HNF4G* represent potential novel associations. *CCBE1* plays a role in extracellular matrix remodeling and is important to the formation of lymphatic vessels. *HNF4G* is a part of a family of transcription factors known to play important roles in development, homeostasis, and metabolism. The index variant for *CCBE1* (rs747874008) was rare in all ancestry groups (MAF=0.12%). In contrast, the index variant for *HNF4G* (rs1817002) was intergenic and common across all ancestry groups (MAF=26.4%). Population specific analyses suggest associations with both variants are primarily driven by European participants. Through our analysis, we confirmed an association between *APOE* and CIMT, and identified two potentially novel loci for CIMT and plaque. The *CCBE1* variant may have been missed by previous GWAS due to its low

MAF. However, the common *HNF4G* variant was evaluated in prior array-based GWAS and was not associated despite being well imputed. Replication is required in order to confirm or refute the association of these novel loci with carotid atherosclerosis.



## **PgmNr 2659: Multi-ethnic whole genome sequence analysis of fibrinogen, fibrin D-dimer, tissue plasminogen activator & plasminogen activator inhibitor 1 within the TOPMed program.**

### **Authors:**

J.E. Huffman<sup>1,2</sup>; B. Rodriguez<sup>2,3</sup>; L.M. Raffield<sup>4</sup>; P.S. de Vries<sup>5</sup>; J.A. Brody<sup>6</sup>; H. Chen<sup>5</sup>; M.R. Brown<sup>5</sup>; J. Haessler<sup>7</sup>; J.P. Lewis<sup>8</sup>; N. Pankratz<sup>9</sup>; L. Yanek<sup>10</sup>; X. Guo<sup>11</sup>; R. Bowler<sup>12</sup>; L. Almasy<sup>13,14</sup>; L.F. Bielak<sup>15</sup>; A.P. Reiner<sup>7,16</sup>; A.D. Johnson<sup>2,3</sup>; A.C. Morrison<sup>5</sup>; N.L. Smith<sup>16,17,18</sup>; TOPMed Hematology and Hemostasis Working Group

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Center for Population Genomics, VA Boston Healthcare System, Jamaica Plain, MA.; 2) The NHLBI's Framingham Heart Study, Framingham, MA; 3) NHLBI's Population Science Branch, Framingham, MA; 4) Department of Genetics, University of North Carolina, Chapel Hill, NC; 5) Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas Health Science Center at Houston, Houston, TX; 6) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA; 7) Division of Public Health, Fred Hutchinson Cancer Research Center, Seattle, WA; 8) Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 9) University of Minnesota School of Medicine, Minneapolis, MN; 10) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 11) Institute for Translational Genomics and Population Sciences, Department of Pediatrics, LABioMed at Harbor-UCLA Medical Center, Torrance, CA; 12) Division of Pulmonary Medicine, Department of Medicine, National Jewish Health, Denver, CO; 13) Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA; 14) Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 15) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI; 16) Department of Epidemiology, University of Washington, Seattle, WA; 17) Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, WA; 18) Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle, WA

---

Compared with array-based, imputation-based, and exome-focused analyses, whole genome sequencing (WGS) data provides better coverage of the genome and better representation of non-European variants.

To better understand the genetics underlying several hemostasis traits, we leverage Freeze 6 deep whole genome sequences from NHLBI's Trans-Omics for Precision Medicine (TOPMed) program to investigate plasma levels of 4 hemostasis measures: fibrinogen (n= 32,572), fibrin D-dimer (n=19,049), tissue plasminogen activator (tPA; n=4,393), and plasminogen activator inhibitor 1 (PAI-1; n=7,857). Phenotypes were centrally harmonized across up to 12 studies that included participants of European, African, Asian, and Hispanic ancestry. Association analyses were conducted using inverse normalized and rescaled residuals adjusting for age, sex, study, TOPMed phase, study-specific parameters, self-reported ancestry, 11 ancestry informative principal components, and a

kinship matrix. All analyses were conducted on the Analysis Commons cloud computing platform using the SMMAT function implemented in GENESIS. Single-variant analyses included all variants with a minor allele count  $\geq 40$ . Gene-based aggregate analyses used 3 strategies for variant selection: 1) loss of function (LOF), 2) LOF and deleterious missense (LDM), and 3) coding, enhancer and promoter variants. The latter aggregation tests were restricted to variants with a minor allele frequency (MAF)  $< 0.05$ , whereas no restrictions were implemented for LOF aggregation tests.

Significantly associated regions were found in single variant tests for fibrinogen ( $n=7$ ) and D-dimer ( $n=3$ ). All were in loci previously associated with these phenotypes, and the majority were common variants in high linkage disequilibrium with previously reported variants. The most significant association for fibrinogen was a previously reported rare missense mutation (rs148685782,  $p=6.8 \times 10^{-48}$ , MAF=0.003, *FGG*) located within the region containing the fibrinogen structural genes: *FGA*, *FGB*, *FGG*. LOF and LDM aggregation tests demonstrated associations with these genes only. No significant associations for genes with  $>5$  alternate alleles were identified for D-dimer. No significant associations were detected for tPA or PAI-1.

Fine-mapping analyses are planned within several regions for fibrinogen to leverage the resolution provided by WGS. Meta-analysis using external cohorts imputed to the TOPMed reference panel is planned to improve power for discovery.

## PgmNr 2660: Whole genome sequencing analysis reveals the contribution of rare deleterious variants in the development of coronary artery disease.

### Authors:

H. Ieki <sup>1,2</sup>; K. Ito <sup>2</sup>; S. Koyama <sup>2</sup>; M. Akiyama <sup>3</sup>; C. Terao <sup>3</sup>; H. Matsunaga <sup>1,2</sup>; K. Ozaki <sup>2,4</sup>; Y. Onouchi <sup>2,5</sup>; S. Nomura <sup>1,6</sup>; H. Akazawa <sup>1</sup>; M. Horikoshi <sup>7</sup>; Y. Momozawa <sup>8</sup>; K. Matsuda <sup>9</sup>; Y. Murakami <sup>10</sup>; M. Kubo <sup>3</sup>; H. Aburatani <sup>6</sup>; Y. Kamatani <sup>3</sup>; I. Komuro <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Cardiovascular Medicine, The University of Tokyo, Tokyo, Japan; 2) Laboratory for Cardiovascular Disease, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 3) Laboratory for Statistical Analysis and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 4) Division for Genomic Medicine, Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Japan; 5) Department of Public Health, Chiba University Graduate School of Medicine, Chiba, Japan; 6) Genome Science Division, Research Center for Advanced Science and Technologies, The University of Tokyo, Tokyo; 7) Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 8) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 9) Department of Computational Biology and Medical Science, Graduate school of Frontier Sciences, The University of Tokyo, Tokyo, Japan; 10) Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

---

Coronary artery disease (CAD) is a leading cause of death worldwide. According to a European twin study, the ratio of genetic contribution to the development of CAD is estimated to be more than 50%. Although genome-wide association studies (GWAS) revealed numerous disease susceptibility loci, the explained heritability by GWAS did not reach the expected ratio partly because their target was limited to low frequency to common variants. Therefore, investigation of rare variants is mandatory to unveil the entire genetic architecture of CAD. But the contribution of rare variants has not been fully characterized thus far. Here, we conducted whole genome sequencing (WGS) on 4417 Japanese samples (1781 myocardial infarction cases, 2636 healthy controls) and analyzed rare variants. Functional annotations for the variants called in the WGS were performed using ANNOVAR and RefSeq. For the analysis, we grouped together frameshift insertion/deletion, stop gain/loss, splicing variants as loss of function (LOF) variants. We also grouped missense variants with combined annotation dependent deletion (CADD) score greater than 20 as LOF variants. Next, we excluded the variants reported in 1000 Genomes Phase 3 in East Asian cohorts to have just rare variants. We then compared the number of these rare LOF variants between case and control by Fisher's exact test. We observed a significant excess of LOF variants in *LDLR* (14 in case and 0 in control,  $P = 3e-6$ ). We also found significant differences in *PCNX4* (11 in case and 1 in control,  $P = 3.4e-4$ ), *BCLAF1* (0 in case and 16 in control,  $P = 3.9e-4$ ), *TMEM255B* (8 in case and 0 in control,  $P = 6.9e-4$ ), *TMC7* (10 in case and 1 in control,  $P = 7.8e-4$ ). To validate the result, we explored LOF variants in another Japanese cohort (200 myocardial infarction cases and 832 controls) and confirmed a significant burden of LOF variants in *LDLR* (8 in case and 0 in control). These results suggest that rare LOF variants can be a risk for CAD and contribute to the heritability. Because of relatively small sample size and classical statistical

method for evaluation, our results should be re-confirmed by large sample size cohort studies with a sophisticated statistical method tailored to rare-variant analysis.

## PgmNr 2661: High-resolution genome-wide association study and clinical characterization of “*monogenic*” variants in coronary artery disease.

### Authors:

S. Koyama <sup>1</sup>; K. Ito <sup>1</sup>; C. Terao <sup>2</sup>; M. Akiyama <sup>2</sup>; H. Matsunaga <sup>1,3</sup>; H. Ieki <sup>1,3</sup>; K. Ozaki <sup>1</sup>; Y. Onouchi <sup>1</sup>; S. Nomura <sup>3</sup>; H. Akazawa <sup>3</sup>; M. Horikoshi <sup>4</sup>; Y. Momozawa <sup>5</sup>; K. Matsuda <sup>6</sup>; Y. Murakami <sup>6</sup>; M. Kubo <sup>7</sup>; H. Aburatani <sup>8</sup>; F. Matsuda <sup>9</sup>; Y. Kamatani <sup>3,9</sup>; I. Komuro <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 2) Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) Department of Cardiovascular Medicine, The University of Tokyo, Tokyo, Japan; 4) Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Centre for Integrative Medical Sciences, Yokohama, Japan.; 5) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.; 6) Institute of Medical Science, The University of Tokyo, Tokyo, Japan.; 7) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.; 8) Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan; 9) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.

Coronary artery disease (CAD) is a leading cause of death worldwide. CAD is known as a common, heritable disease. Numerous genome-wide association studies (GWAS) were conducted to elucidate the genetic architecture of CAD and have identified over 160 disease susceptibility loci. To contribute broader knowledge on the genetics of CAD, we conducted a large-scale genetic analysis of CAD in combination with whole-genome sequencing.

We sequenced genomes of 1782 CAD cases and 3148 controls recruited from the Japanese population. By the whole genome sequencing, we found several “pathogenic” variants for CAD. To leverage enriched disease-specific haplotype information, we constructed a novel reference panel for haplotype imputation. Newly developed reference panel significantly improved imputation quality in the Japanese population, especially for rare to very rare [minor allele frequency (MAF) < 0.001] variants. Using these densely imputed genotypes, we performed a large-scale GWAS for CAD (25,859 cases and 69,359 controls) in the Japanese population. We identified 9 previously unreported loci at the genome-wide significance level and additional associated independent signals. These independent signals included rare and high impact coding variants in CAD-relevant genes. Especially we found rare, coding variants in established causal genes [*LDLR*, MAF 0.05%, odds ratio (OR) 7.2, *PCSK9* MAF 1%, OR 1.5, *APOB* MAF 3%, OR 1.5] as independent signals. These variants are specific to the East Asian population and have a large effect size in CAD development. To characterize the carriers of these rare variants, we performed phenome-wide analyses using a biobank-based dataset. These variants are enriched in acute coronary syndromes, and the carriers showed worse clinical phenotypes exemplified by worsened cholesterol profile and early-onset myocardial infarction. As expected from these surrogate markers, individuals with these variants showed increased long-term cardiovascular mortality (Hazard ratio 1.21 [1.09-1.39],  $p=6 \times 10^{-3}$ ). In summary, population and disease-specific reference panel provided improved imputation performance and accurate estimation of the prevalence and penetrance of rare-pathogenic variants in the context of GWAS.

## PgmNr 2662: Pathways affecting polygenic risk for supralvalvar aortic stenosis in Williams-Beuren syndrome.

### Authors:

B.A. Kozel <sup>1</sup>; P.C.R. Parrish <sup>1,2</sup>; R.H. Knutsen <sup>1</sup>; C.J. Billington, Jr <sup>1</sup>; A.E. Pickett <sup>1</sup>; R.P. Mecham <sup>3</sup>; Y.-P. Fu <sup>1</sup>; D. Liu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland.; 2) Department of Genetics, University of Washington, Seattle, WA; 3) Department of Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO

Williams Beuren syndrome (WBS) is caused by the deletion of 26-28 genes on chromosome 7 and produces a characteristic set of multi-organ phenotypes. Haploinsufficiency for elastin (*ELN*) causes the cardiovascular features of the condition, including supralvalvar aortic stenosis (SVAS), but severity varies widely. To identify factors that impact SVAS risk, we performed exome sequencing in 104 individuals with WBS and looked for differences in variant burden among the phenotypic groups. Single variable analysis revealed no statistically significant effect of sex, race, or WBS deletion size on SVAS outcomes. Likewise, neither second hits in *ELN* itself nor other genes in the WBS critical region had significant effect on the SVAS phenotype. Exome wide gene-based analyses (SKAT-O) performed on nonsynonymous variants in an extreme phenotype subset (n=62) revealed no single gene with exome wide significance. However, gene set enrichment showed a preponderance of variants affecting the adaptive immune system (False Discovery Rate (FDR) =  $1 \times 10^{-7}$ ). Studies in *Eln*<sup>+/-</sup>; *Rag1*<sup>+/-</sup> mice showed larger aortic caliber in the double mutant animal (p<0.01), confirming the role of the adaptive immune system in modifying aortic disease in elastin insufficiency.

Subsequent gene set enrichment analysis performed on a subset of variants with CADD\_phred >10 and more than 10% difference in allele frequency between those with no and severe SVAS identified modifiers in extracellular matrix (ECM), GPCR signaling, lipid metabolism and immune pathways. Focused burden evaluation of those pathways by SKAT-O yielded positive results for each with FDR = 0.001-0.03, while Qtest Wald testing resulted in even more significant findings with ECM variant burden FDR =  $2 \times 10^{-26}$ , while the other three pathways had FDR on the order of  $10^{-7}$ . Investigation of *Eln*<sup>+/-</sup> mice on diverse genetic backgrounds confirmed a role for variation in the ECM in modifying vascular features of elastin insufficiency.

Taken together, our studies have allowed us to detect important genomic variations in adaptive immunity and ECM pathways that contribute to SVAS risk in individuals with WBS. Replication is needed but our findings show that focus on pathway level, rather than gene level analysis may reveal more meaningful results in small cohorts. Orthogonal methods in animal models can help to confirm suggestive results from exploratory studies.

\*PCR Parrish is the study's first author and BA Kozel is the senior author.

## PgmNr 2663: Genetic influences on coronary heart disease act on atherosclerosis in type 2 diabetes.

### Authors:

T. Lu <sup>1,2</sup>; V. Forgetta <sup>1</sup>; O.H.Y. Yu <sup>1,3</sup>; G. Thanassoulis <sup>4,5</sup>; C.M.T. Greenwood <sup>1,6,7,8</sup>; J.B. Richards <sup>1,6,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, Montreal, Quebec, Canada; 2) Quantitative Life Sciences, McGill University, Montreal, Quebec, Canada; 3) Division of Endocrinology, Jewish General Hospital, Montreal, Quebec, Canada; 4) Department of Medicine, McGill University, Montreal, Quebec, Canada; 5) Preventive and Genomic Cardiology, McGill University Health Centre and Research Institute, Montreal, Quebec, Canada; 6) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 7) Gerald Bronfman Department of Oncology, McGill University, Montreal, Quebec, Canada; 8) Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Quebec, Canada; 9) Department of Twin Research and Genetic Epidemiology, King's College London, Strand, London, United Kingdom

**Background:** Type 2 diabetes (T2D) increases the risk of coronary heart disease (CHD), yet the specific mechanisms involved remain poorly described. Traditional clinical risk factors, including blood pressure, serum lipids, glycemic control, obesity and smoking fail to identify all individuals who will develop CHD, particularly among the diabetics, suggesting that some risk of CHD in diabetics is not captured by these clinical risk factors. Polygenic risk scores (PRS) provide an opportunity to explore the mechanisms involved in disease, because this data driven approach surveys the genome and makes no prior mechanistic assumptions. We tested whether a PRS for CHD influenced risk of CHD in patients with T2D and which known CHD mechanisms were associated.

**Methods:** We first tested the association of a previously published CHD PRS with CHD amongst patients with T2D in UK Biobank [N = 21,102 patients with T2D, of whom, 1,898 (9.0%) developed CHD]. To better understand which mechanisms were influenced by this PRS, we tested its association with CHD traditional risk factors amongst these patients. However, atherosclerotic burden is not assessed in UK Biobank. We therefore undertook coronary angiograms, clinical risk factor measurement and genome-wide genotyping in a cohort of 352 patients with T2D and tested whether the PRS for CHD was associated with multivessel stenotic disease, as detected by angiogram, and/or traditional CHD risk factors.

**Results:** In the UK Biobank we found that the CHD PRS was strongly associated with CHD amongst patients with T2D (OR per standard deviation = 1.50; 95% CI 1.43-1.57;  $p < 2 \times 10^{-16}$ ). Amongst patients with T2D, the CHD PRS was not associated with hypertension, hypertriglyceridemia, glycemic control, obesity or smoking, and was moderately associated with hyperlipidemia (OR = 1.22 per standard deviation increase; 95% CI 1.12-1.33;  $p = 3.5 \times 10^{-6}$ ). On the other hand, the CHD PRS was strongly associated with multivessel stenosis (OR = 1.65; 95% CI 1.25-2.20;  $p = 4.9 \times 10^{-4}$ ). Further, the CHD PRS was associated with increased number of major stenotic lesions (OR = 1.35; 95% CI 1.08-1.69;  $p = 9.4 \times 10^{-3}$ ).

**Conclusions:** Amongst individuals with T2D, the CHD PRS was strongly associated with atherosclerotic burden and this effect is largely independent of traditional clinical risk factors. The mechanisms leading to increased atherosclerotic burden should be investigated to understand CHD risk in T2D.



## PgmNr 2664: Returning polygenic risk information for coronary artery disease and other traits to biobank participants.

### Authors:

M. Marttila <sup>1</sup>; T. Paajanen <sup>1</sup>; H. Marjonen <sup>1</sup>; A. Haukkala <sup>2</sup>; H. Kääriäinen <sup>1</sup>; K. Kristiansson <sup>1</sup>; M. Perola <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Institute of Health and Welfare, Helsinki, Finland; 2) 2Faculty of Social Sciences, University of Helsinki, Finland; 3) 3Research Program for Clinical and Molecular Metabolism, Faculty of Medicine, University of Helsinki, Finland

---

We have recruited 3449 participants from a biobanked population-based study cohort FinHealth to a separate study called P5 FinHealth. Here, we will return polygenic risk scores (PRS) for three common diseases via internet portal to these volunteering participants. To validate the PRS, have tested how for coronary artery disease (CAD), type 2 diabetes and venous thromboembolism affect the risk of disease within the following ten years in participants from a population-based study FINRISK .

Based on follow-up data from national health care registries we validated PRS in whole genome genotyped population based FINRISK cohorts with multiple registry follow-up for incident cases (N=21726) and used validated PRS to model P5.fi FinHealth participants from FinTerveys 2017 (N= 3449) using Cox regression models. We estimated the impact of genetic and traditional risk factors on a risk of disease within the next 10 years.

We compared the CAD classification of Cox regression model with traditional risk factors and polygenic additive model with 6,6M variants. From basic model risk class 10-20% PRS reclassified 205 (17%) participants to highest >20% risk class. In FINRISK participants risk for CAD was higher for men and smoking adds to risk of disease in all PRS classes. Risk for disease increases with BMI being highest in BMI class 30-35. Our initial analysis in Finnish population cohorts shows that individuals with very high genetic risk, the PRS class over 2, may get the disease significantly younger than 60-70 years which is conventionally considered the peak.

The validation showed that adding PRS to the traditional risk factors significantly changed the risk enabling reclassification of as many as 17% of the participants. Polygenic risk information is returned to volunteering participants through an internet portal. Changes in life style are followed up using questionnaires through the portal and the morbidity will be collected through health registers. P5 study is a pilot for P6 which will recruit 100 000 participants and include at least 6-10 diseases for similar feedback study, including randomized controlled trial for PRS stratified intervention.



## PgmNr 2665: Genome-wide association study for atrial fibrillation identifies 25 novel loci.

### Authors:

H. Matsunaga <sup>1,2</sup>; K. Ito <sup>2</sup>; S. Koyama <sup>2</sup>; M. Akiyama <sup>3</sup>; C. Terao <sup>3</sup>; H. Ieki <sup>1,2</sup>; K. Ozaki <sup>2,4</sup>; Y. Onouchi <sup>2,5</sup>; S. Nomura <sup>1,6</sup>; H. Akazawa <sup>1</sup>; M. Horikoshi <sup>7</sup>; Y. Momozawa <sup>8</sup>; K. Matsuda <sup>9</sup>; Y. Murakami <sup>10</sup>; M. Kubo <sup>3</sup>; H. Aburatani <sup>6</sup>; Y. Kamatani <sup>3</sup>; I. Komuro <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Cardiovascular Medicine, The University of Tokyo, Tokyo, Japan; 2) Laboratory for Cardiovascular Disease, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 3) Laboratory for Statistical Analysis and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 4) Division for Genomic Medicine, Medical Genome Center, National Center for Geriatrics and Gerontology, Obu, Japan; 5) Department of Public Health, Chiba University Graduate School of Medicine, Chiba, Japan; 6) Genome Science Division, Research Center for Advanced Science and Technologies, The University of Tokyo, Tokyo; 7) Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 8) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 9) Department of Computational Biology and Medical Science, Graduate school of Frontier Sciences, The University of Tokyo, Tokyo, Japan; 10) Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Atrial fibrillation (AF) is one of the common arrhythmias in the world and can lead to stroke, heart failure and sudden death. Genome-wide association studies (GWAS) have revealed more than 150 AF-related loci and enhanced our understanding of the underlying mechanism of AF. While most of previously reported loci were discovered in Europeans, some loci were only identified in East Asian and other populations, which emphasizes the significance of GWAS in non-European populations. Moreover, AF is often related to structural heart diseases and other comorbid chronic conditions, but the shared genetic basis between AF and such disorders remains unclear. To tackle those problems, we carried out a Japanese GWAS, followed by a trans-ethnic meta-analysis for AF and exploring the association of AF with other diseases.

First, we conducted a GWAS in the Japanese population that encompassed 10,250 cases and 148,998 controls. We identified 32 loci associated with AF ( $P < 5.0 \times 10^{-8}$ ), of which 4 loci were novel. Next, we performed a trans-ethnic meta-analysis by MANTRA using the result from a European GWAS and discovered 21 additional new loci ( $\log_{10}$  Bayes' factor  $> 6$ ). In total, our study identified 25 novel loci associated with AF. These new loci were involved in cardiac development, contractility and electrophysiological function. Gene set enrichment analysis revealed the association of AF with FHL2 subnetwork ( $P = 1.12 \times 10^{-11}$ ) and SRF subnetwork ( $P = 1.26 \times 10^{-11}$ ), which were prevalent in cardiomyopathy. In tissue and cell type enrichment analysis, we found significant enrichment in cardiovascular system (Heart Ventricles,  $P = 3.79 \times 10^{-7}$ , Heart Atria,  $P = 2.47 \times 10^{-6}$ , Atrial Appendage,  $P = 2.95 \times 10^{-6}$ ) and musculoskeletal system (Muscle Striated,  $P = 5.17 \times 10^{-4}$ , Muscle Skeletal,  $P = 5.17 \times 10^{-4}$ ). We evaluated the genetic correlations between AF and other diseases using Japanese GWAS by means of bivariate LD score regression. AF showed significant positive genetic correlation with blood pressure and coronary artery disease. Finally, we explored associations between genome-wide significant loci for AF and clinical measurements. We then observed that several loci were

associated with cardiovascular traits, such as elevated blood pressure, decreased ejection fraction and left ventricular enlargement.

In conclusion, we identified 25 novel loci for AF and showed the shared genetic basis between AF and other associated diseases. Our results may help us understand the underlying mechanisms of AF.

## PgmNr 2666: Exome sequencing in multiplex families with left sided cardiac defects has high yield for disease gene discovery.

### Authors:

K.L. McBride <sup>1 2 3 4</sup>; D. Cunningham <sup>1</sup>; S.M. Fitzgerald-Butt <sup>5</sup>; R. Neinast <sup>1</sup>; G. Zender <sup>1</sup>; D. Corsmeier <sup>4</sup>; D. Gordon <sup>4</sup>; V. Garg <sup>1 2 3</sup>; P. White <sup>2 4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center Cardiovascular Research, Nationwide Children's Hospital, Columbus, OH.; 2) Dept Pediatrics, College of Medicine, Ohio State University; 3) The Heart Center, Nationwide Children's Hospital, Columbus, OH.; 4) Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, OH.; 5) Dept of Medical and Molecular Genetics, School of Medicine, Indiana University, Indianapolis, IN

The genetic architecture of congenital heart disease (CHD) is complex. In addition, cardiac phenotyping can be challenging and parsing individuals with additional congenital anomalies or neurodevelopmental disorders that may indicate a multisystem syndrome is complicated. Recent large-scale exome sequencing projects have identified single gene causes in as many as 10% of trios. Most have been in individuals with multisystem disease, with the yield in isolated CHD of 1-3%. We sought to overcome these problems by a family based approach to CHD gene discovery, using multiplex families to better phenotype cases and use segregation as an additional criterion for variant pathogenicity. We performed exome sequencing on 19 families with 2 or more individuals with isolated left sided cardiac defects, specifically aortic valve stenosis, coarctation of the aorta and hypoplastic left heart syndrome. After quality control and filtering for minor allele frequency (<0.001), we ranked loss of function and missense variants with a minimum CADD score of 30. Five families had segregating predicted pathogenic variants in previously described genes known to cause CHD (*CTBP2*, *MATR3*, *MCTP2*, *MYH6*, *NOTCH1*). Segregating variants in three novel CHD gene candidates were identified; two showed functional changes (*BMP10*, *CASZ1*), and functional studies are ongoing for one (*ROCK1*). A variant in *BMP10* was shown to have normal mRNA and protein expression in transfected HEK293 cells but the variant protein was not secreted by cell based assay. A variant in the nuclear localizing signal domain of *CASZ1* was shown have normal mRNA and protein expression. Transfection of the mutant into HEK293 cells demonstrated the protein did not localize to the nucleus, and compared to wild type was unable to transactivate a tyrosine hydroxylase luciferase reporter construct. We identified two additional segregating predicted pathogenic variant candidates, but luciferase transactivation assays studies for a *HEY1* variant and potassium current studies in *Xenopus* oocytes for a *KCNJ2* variant failed to show any functional differences. Lowering the CADD cut point to 20 has identified candidates in six other families, leaving three families without a reasonable disease gene candidate. We demonstrate that family based approaches have a much higher yield for identifying left sided CHD causing genes compared to parent-case trio approaches, at over 40% (8/19) for known or functionally validated novel CHD genes.

## PgmNr 2667: Assessing the role of rare genetic variations in heart failure via whole exome sequencing.

### Authors:

G. Povysil <sup>1</sup>; O. Chazara <sup>2</sup>; J. Armisen <sup>2</sup>; C. Haefliger <sup>2</sup>; D.B. Goldstein <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Genomic Medicine, Columbia University Irving Medical Center, New York, NY; 2) Centre for Genomics Research, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK

---

Several sequencing studies have found causal genetic variants for distinct subtypes of heart failure (HF) such as hypertrophic or dilated cardiomyopathy. However, the role of genetic variants in unselected HF patients has not been assessed so far.

We performed whole exome sequencing of almost 6,000 cases with unselected HF from the CHARM (Candesartan in Heart Failure-Assessment of Reduction in Mortality and Morbidity) and CORONA (Controlled Rosuvastatin Multinational Trial in Heart Failure) studies and compared them to a set of 13,000 unrelated controls. Using a gene-based collapsing approach, we looked for genes with an excess of rare and presumably deleterious variants among cases compared to controls.

We detected a significant enrichment of rare protein-truncating variants (PTVs) in the *TTN* gene. Truncating variants in *TTN* are a known cause of dilated cardiomyopathy, but only if they are located in exons that are spliced into most of the *TTN* transcripts present in the heart. We used the presence of PTVs in such exons as additional criterion for defining qualifying variants, which increased the statistical significance of enrichment further. In order to investigate the burden of pathogenic variants in other known cardiomyopathy genes, we performed a diagnostic analysis of all cases and classified variants based on the American College of Medical Genetics and Genomics (ACMG) guidelines for clinical sequence interpretation criteria. We defined diagnostic variants as those that were classified as “pathogenic” or “likely pathogenic” according to ACMG and could be linked to heart failure. In addition to PTVs in *TTN*, we found diagnostic variants in other cardiomyopathy genes such as *MYBPC3*, *MYH7*, *SCN5A*, and *TNNT2*.

Using whole exome sequencing, we have demonstrated that even in a group of unselected HF patients a significant number of individuals harbor diagnostic variants in genes that have previously been associated with cardiomyopathy.

## PgmNr 2668: Increased burden of nsSNVs in cardiomyopathy genes predicts dilated but not hypertrophic cardiomyopathy.

### Authors:

M. Puckelwartz <sup>1,3</sup>; L. Pesce <sup>2</sup>; L. Dellefave-Castillo <sup>3</sup>; M. Wheeler <sup>4</sup>; T. Pottinger <sup>3</sup>; A. Robinson <sup>3</sup>; A. Gacita <sup>3</sup>; Z. Schoppen <sup>3</sup>; W. Pan <sup>3</sup>; G. Kim <sup>5</sup>; J. Wilcox <sup>6</sup>; A. Anderson <sup>6</sup>; E. Ashley <sup>4</sup>; S. Day <sup>7</sup>; T. Cappola <sup>8</sup>; G. Dorn <sup>9</sup>; E. McNally <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pharmacology, Northwestern University, Chicago, IL.; 2) Computation Institute, University of Chicago, Chicago, IL.; 3) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago IL.; 4) Department of Medicine/Cardiovascular Medicine, Stanford University, Stanford, CA; 5) Department of Medicine, University of Chicago, Chicago, IL 60637; 6) Department of Medicine, Northwestern University, Chicago, IL 60611; 7) Department of Internal Medicine, The University of Michigan, Ann Arbor, Michigan; 8) Perelman School of Medicine, Division of Cardiovascular Medicine and Penn Cardiovascular Institute and Department of Medicine, University of Pennsylvania, Philadelphia, PA; 9) Washington University School of Medicine, St. Louis, MO

**Rationale:** Dilated and hypertrophic cardiomyopathy are important causes of heart failure and both cardiomyopathy subtypes result from autosomal dominant inheritance with variable expressivity and penetrance.

**Hypothesis:** We hypothesized that additional genetic variation contributed to the manifestation of the pathogenesis of dilated cardiomyopathy (DCM), in which the left ventricle becomes enlarged and poorly contractile, versus hypertrophic cardiomyopathy (HCM), a condition accompanied by a thickened ventricle. We conducted whole genome sequencing to isolate genes responsible for cardiomyopathy expressivity.

**Methods and Results:** Genome sequencing was performed on subjects with dilated (n=71) or hypertrophic (n=56) cardiomyopathy. Variation was assessed in cardiomyopathy genes (n=102) derived from gene panels routinely tested during cardiomyopathy genetic testing. Genetic variation was correlated with measures of heart function and size. Principal component analysis of left ventricle measures using echocardiographic data separated the two cardiomyopathy subtypes (DCM and HCM). Regression of the first principal component using all nonsynonymous single nucleotide variants (nsSNVs) in cardiomyopathy panel genes showed that the number of nsSNVs predicted DCM but not HCM. The likelihood of DCM in the cohort significantly increased as the number of cardiomyopathy gene nsSNVs increased ( $p < 0.02$ ). The increase in cardiomyopathy gene nsSNVs also significantly associated with reduced left ventricular ejection fraction and increased left ventricular diameter, both measures indicative of DCM. Genomewide gene expression in the heart was characterized with data from the Genotype-Tissue Expression (GTEx) database intersecting left ventricle and left atrial RNAseq data. Resampling methods identified genes with deviant cumulative allele frequencies in DCM vs HCM, and potential cardiomyopathy modifying genes were identified in the cardiomyopathy gene panel and all cardiac genes as defined by GTEx.

**Conclusions:** Dilated cardiomyopathy subjects carried a greater burden of nsSNVs in cardiomyopathy genes. In contrast, nsSNV burden in cardiomyopathy genes did not correlate with the

probability or manifestation of left ventricular measures of hypertrophic cardiomyopathy. These data support a complex genetic architecture where increased cardiomyopathy gene variation establishes a genetic background that predisposes to dilated cardiomyopathy and decreased left ventricular function.

## PgmNr 2669: *Cis-ANXA11* and *DYDC2* and *trans-KCNMA1* variants are associated with heterogeneous phenotypes in arrhythmogenic dilated cardiomyopathy.

### Authors:

E. Purevjav<sup>1,2</sup>; W. Zhang<sup>3</sup>; U. Mendsaikhan<sup>3</sup>; U. Munkhsaikhan<sup>1,2</sup>; F. Xu<sup>4</sup>; L. Lu<sup>4</sup>; J.A. Towbin<sup>1,2,3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Cardiology, University of Tennessee Health Science Center, Memphis, Tennessee.; 2) Children's Foundation Research Institute, Le Bonheur Children's Hospital, Memphis, TN; 3) The Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN; 5) Pediatric Cardiology, St. Jude Children's Research Hospital, Memphis, TN

**Objectives:** Cardiomyopathies are genetic diseases of heart muscle caused by mutations in cytoskeletal genes. Disease severity and penetrance of mutations vary greatly among patients carrying the same mutations and single-gene variants often do not reliably predict cardiomyopathy phenotypes.

**Background:** The chromosome 10q21-q23 locus was previously associated to familial arrhythmogenic dilated cardiomyopathy (aDCM), heart failure, Wolff-Parkinson-White (WPW) syndrome, mitral valve prolapse (MVP) and/or mitral regurgitation (MR). However, the exact variants responsible for heterogeneous phenotypes remained unknown.

**Methods:** A large family of 62 members was studied using whole exome and direct sequencing. Phenotype-genotype correlation and systems genetics analysis were performed. Quantitative co-segregation studies combined with systems genetics analysis were performed using online functional enrichment analysis tools.

**Results:** We identified missense *KCNMA1*-M802T, *ANXA11*-I457V, and *DYDC2*-P123R variants at 10q21-q23. The proband and ten family members carried *ANXA11*-I457V and *DYDC2*-P123R identified *in cis* (digenic heterozygosity). Seven digenic variant-carriers were affected with DCM (n=3), heart failure (n=1), left ventricular dysfunction (n=3), arrhythmia (n=2), MVP (n=2) and MR (n=3). A single *KCNMA1*-M802T was identified in 11 individuals, including 4 affected with MVP (n=3), MR (n=4), ventricular arrhythmia (n=2), and WPW (n=1). All three variants as a multigenic heterozygosity were identified in three family members and all were affected. Computed genetic and phenotype QTL correlations predicted that rare variants in *CORIN*, *DSP*, *TTN*, *KCNMB3*, *TRM7* and *OTX1* may effect on function of *KCNMA1*, *ANXA11*, and *DYDC2* at the major 10q21-q23 locus.

**Conclusions:** Different variant combinations identified at 10q21-q23 underlie the clinical heterogeneity of aDCM phenotypes. The *in cis* digenic heterozygous *ANXA11*-I457V and *DYDC2*-P123R variants co-segregate with the DCM and HF phenotypes in this family, while *KCNMA1*-M802T is associated with MVP/MR and arrhythmias. Rare variants in cardiomyopathy-related genes located beyond the 10q21-q23 may function as the genetic modifiers in inter-, sub-, and intra-familial phenotypic variability.

## PgmNr 2670: Impact of healthy lifestyle and genetic variants on lipid metabolites and incident risk of coronary heart disease.

### Authors:

J. Si<sup>1,2</sup>; J. Li<sup>1</sup>; C. Yu<sup>1</sup>; Y. Guo<sup>3</sup>; Z. Bian<sup>3</sup>; L. Yang<sup>4,5</sup>; I. Millwood<sup>4,5</sup>; R. Walters<sup>4,5</sup>; Y. Chen<sup>4,5</sup>; H. Du<sup>4,5</sup>; J. Chen<sup>6</sup>; Z. Chen<sup>5</sup>; L. Li<sup>1</sup>; L. Liang<sup>2</sup>; J. Lv<sup>1,7,8</sup>; the China Kadoorie Biobank Collaborative Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology and Biostatistics, School of Public Health, Peking University Health Science Center, Beijing, China; 2) Departments of Epidemiology and Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Chinese Academy of Medical Sciences, Beijing, China; 4) Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, United Kingdom; 5) Clinical Trial Service Unit & Epidemiological Studies Unit (CTSU), Nuffield Department of Population Health, University of Oxford, United Kingdom; 6) China National Center for Food Safety Risk Assessment, Beijing, China; 7) Key Laboratory of Molecular Cardiovascular Sciences (Peking University), Ministry of Education, Beijing, China; 8) Peking University Institute of Environmental Medicine, Beijing, China

Adherence to healthy lifestyles is associated with a considerable reduction of coronary heart disease (CHD) risk. We quantified how much of the effects of healthy lifestyles on CHD risk are mediated through lipid metabolites, and compared the effect of healthy lifestyle and genetic scores for two LDL-C lowering targets on the lipidomic profile. A prospective case-control sample of 4,681 participants from the China Kadoorie Biobank aged 30 to 79 years without prior cardiovascular disease or statin use was selected for a metabolomics study. Smoking, alcohol consumption, dietary habit, physical activity, and BMI and waist circumference, assessed by questionnaire and physical measurements at baseline, were used to define five healthy lifestyles. A comprehensive lipid profile in baseline plasma samples was measured by nuclear magnetic resonance. Genetic instruments for lipid-lowering effects of statins and ATP citrate lyase (ACLY) inhibition were used to examine their impact on lipid metabolites. Adherence to healthy lifestyles was associated with 50 components of the lipid profile (FDR < 0.05), including substantially lower very low-density lipoprotein (VLDL) particles and their cholesterol concentration, smaller VLDL particle size, increased high-density lipoprotein (HDL) particles and their cholesterol concentration, and larger HDL particle size. The triglycerides within all lipoproteins, including most HDL particles, were lower among participants who adhered to a healthy lifestyle. A total of 35 lipid metabolites showed significant mediation effect from healthy lifestyle to CHD. The strongest mediator, ratio of apolipoprotein B to A1, accounted for 11% of the reduced CHD risk associated with healthy lifestyle. All lipid metabolites jointly explained 14% of the protective effect of healthy lifestyle on CHD. Genetic scores for the targets of statins and ACLY showed similar effect on reducing concentrations of intermediate- and low-density lipoprotein particles, and cholesterol within these particles, while healthy lifestyle more strongly improved VLDL- and HDL-related measures, and triglycerides in almost all lipoprotein subfractions. The effects of adherence to healthy lifestyles on lower CHD risk were partly mediated by an improved lipid profile. Lifestyle interventions and lipid-lowering drug may target different components of the lipid profile, suggesting that they are not redundant strategies but could be combined for improved benefits.



## PgmNr 2671: A genome-wide polygenic score for coronary artery disease to stratify risk among South Asians.

### Authors:

M. Wang <sup>1</sup>; R. Menon <sup>2</sup>; S. Mishra <sup>2</sup>; D. T. <sup>2</sup>; M. Deshmukh <sup>2</sup>; S. Apte <sup>2</sup>; O. Mathew <sup>2</sup>; R. Gupta <sup>2</sup>; R. V. L. <sup>2</sup>; S. Santhosh <sup>2</sup>; A. Vijay B <sup>3</sup>; R. Gupta <sup>4</sup>; A. Mullasari <sup>5</sup>; A. Sigamani <sup>6</sup>; M. Kanchi <sup>6</sup>; S. Kathiresan <sup>1,7,8,9</sup>; A. Khera <sup>1,7,8,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cardiovascular Disease Initiative of the Broad Institute of Harvard and MIT, Cambridge, MA, USA.; 2) MedGenome Labs Ltd., Bangalore, India; 3) Amrita Institute Medical Sciences, Kochi, India; 4) Eternal Heart Care Centre, Jaipur, India; 5) Madras Medical Mission, Chennai, India; 6) Narayana Hrudayalaya, Bangalore, India; 7) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.; 8) Cardiology Division of the Department of Medicine, Massachusetts General Hospital, Boston, MA, USA.; 9) Harvard Medical School, Boston, MA, USA.

---

Genome-wide polygenic scores (GPS) integrate information from many sites of common DNA variation into a single measure of inherited risk. We recently developed a GPS for coronary artery disease that enabled stratification of a primarily European ancestry population into varying trajectories of coronary risk, with the upper tail of the distribution having risk equivalent to that of rare monogenic mutations such as those related to familial hypercholesterolemia. The use of a GPS to identify high-risk individuals may be particularly useful amongst South Asians (who develop coronary artery disease at higher rates and significantly earlier in life) but the predictive value in this population had remained uncertain. Here, we derive a GPS using the LDpred computational algorithm based on the largest GWAS summary statistics to date and validate it in 8,025 South Asian participants of the UK Biobank. We built a South Asian specific reference distribution for the GPS using whole genome sequencing data of 1,741 participants and systematically compared GPS in an independent cohort of 1,059 cases and 718 controls. We observed an odds ratio per standard deviation increase in the GPS of 1.46 (95%CI 1.35 -1.57), AUC = 0.60, results overall similar to those previously observed in studies of European ancestry. As compared to average risk (defined as the middle quintile of the score distribution) the GPS identified 14%, 2.5%, and 0.6% of the population with 2-fold, 3-fold, and 4-fold increased risk for CAD. These results confirm the predictive utility of a GPS among a South Asian population and provide a framework for ancestry-specific score validation and testing. Ongoing efforts are designed to integrate these scores with rare monogenic mutations and optimize a strategy to disclose GPS to interested individuals, with the ultimate goal of empowering high-risk individuals to take advantage of targeted screening or prevention measures prior to disease onset.

## PgmNr 2672: Whole exome sequencing reveals novel pathogenic variants in left ventricular noncompaction cardiomyopathy.

### Authors:

F. Xu <sup>1</sup>; J. Collyer <sup>2,3</sup>; U. Munkhsaikhan <sup>2,3</sup>; W. Zhang <sup>4</sup>; L. Lu <sup>1</sup>; E. Purevjav <sup>2,3</sup>; J.A. Towbin <sup>2,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN; 2) Department of Pediatrics, University of Tennessee Health Science Center, Memphis, TN; 3) Children's Foundation Research Institute, Le Bonheur Children's Hospital, Memphis, TN; 4) The Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 5) Pediatric Cardiology, St. Jude Children's Research Hospital, Memphis, TN

**Background:** Left ventricular noncompaction (LVNC), a subtype of cardiomyopathy, has been recognized as a genetically and phenotypically heterogeneous cardiac muscle disorder. Although much progress has been made, the genetic basis still remains incomplete. The aim of this study is to identify the genetic causes of LVNC using whole exome sequencing.

**Methods:** Study included 42 patients with LVNC and 17 unaffected family members from 33 families. Whole exome sequencing was followed by read mapping, variant calling, annotation, and bioinformatics filtering to prioritize pathogenic or likely pathogenic variants. The disease-causing variants were selected according to MAF < 0.01 in 1000G, ExAC, ESP6500, or gnomAD, variant class (stop gain/loss, frameshift, nonsynonymous, and splice site variant), functional impact (deleterious predicted by SIFT, Polyphen2, MutationTaster, PROVEAN, and CADD). The candidate variants were further verified by Sanger sequencing.

**Results:** 59 pathogenic or likely pathogenic variants were identified in 35 genes, in which 25 were novel variants. Pathogenic variants in *TTN* (10 variants in 7 families), *MYH7* (6 variants in 6 families), *MYBPC3* (4 variants in 3 families), and *DSP* (3 variants in 3 families) were the most common. Two pathogenic variants were found in *ACADVL*, *CPT2*, *DMD*, *DSC2*, *TNNT2*, and *PLEC*. One pathogenic variant was identified in 25 genes including four sarcomere associated genes, *ANKRD1*, *HAND1*, *LDB3*, and *OBSCN*. Digenic and polygenic variants were identified in 9 individuals/pedigrees. For example, a 22-year-old female patient carried pathogenic variants in genes such as *TTN*, *ACADVL*, *CPT2*, *DSP*, and *AARS2*. Trigenic variants in *MYH7*, *ANKRD1*, and *NRG1* were co-segregated in a family with two family members affected with dilated cardiomyopathy (DCM) and LVNC.

**Conclusions:** Our results unravel several known and novel LVNC associated mutations and expands the genetic spectrum of this disorder. The diverse di-, tri- and polygenic profile of inheritance may be associated with heterogeneous LVNC phenotypes.

## PgmNr 2673: Pathway-specific polygenic risk scores identify subgroups with distinct pathway contributions in people with high overall polygenic risk scores for coronary artery disease.

### Authors:

Y. Ye <sup>1</sup>; S. Zekavat <sup>1,2</sup>; X. Chen <sup>3</sup>; P. Natarajan <sup>4,5</sup>; H. Zhao <sup>1,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT.; 2) School of Medicine, Yale University, New Haven, CT.; 3) Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT.; 4) Center for Genomic Medicine and Cardiovascular Research Center, Massachusetts General Hospital, Boston.; 5) Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA; 6) Department of Biostatistics, Yale School of Public Health, New Haven, CT.; 7) Department of Genetics, Yale School of Medicine, New Haven, CT.

---

Polygenic risk scores (PRSs) have been shown to have robust risk prediction performance for coronary artery disease (CAD), with high CAD PRSs reflecting an increased risk from multiple underlying pathways. However, it has not been studied whether there is heterogeneity in genetic risks among individuals with similar high PRS, and what proportion of the PRS have pleiotropy with the known pathways for CAD. Here we propose to develop pathway-specific PRSs (PS-PRSs) to represent contributions of pathways and investigate whether clinically meaningful subgroups exist among people with high CAD PRS based on PS-PRSs. We considered the UK Biobank samples consisting of 407,060 European participants, divided into training and testing sets. State-of-the-art methods were used to generate the CAD PRSs, with the optimal PRS being the AnnoPred PRS using 233,538 variants validated in the testing set (15,668 cases and 309,980 controls) with area under the receiver-operator curve (AUC) = 0.639 (95% CI, 0.635-0.644) which was significantly higher than the AUC = 0.628 (95% CI, 0.623-0.633) given by LDpred. We partitioned the optimal CAD PRS into multiple PS-PRSs corresponding to the following CAD risk factors: low-density lipoprotein (LDL-C), triglycerides (TG), body mass index (BMI), systolic blood pressure (SBP), insulin resistance (IR) and smoking. Each PS-PRS had around 25,000 variants that were selected based on the GWAS summary statistics of the pathway-corresponding CAD risk factors. Variants belonging to none of the PS-PRSs were integrated as the unknown-pathway PRS. Of individuals having the top 5% overall CAD PRS, the contributions to the overall PRS from specific pathway were 0.151 (95% CI, 0.137-0.164) for LDL-C, 0.083 (95% CI, 0.069-0.097) for TG, 0.029 (95% CI, 0.016-0.043) for BMI, 0.069 (95% CI, 0.056-0.083) for SBP, 0.038 (95% CI, 0.024-0.052) for IR, 0.048 (95% CI, 0.035-0.062) for smoking and 0.488 (95% CI, 0.476-0.500) for unknown components according to the linear regression model where the outcome was the overall CAD PRS. The subgroups with high PS-PRSs had higher levels or prevalence of the corresponding risk factors. For example, the mean LDL-C level in the top and bottom decile of LDL-C PS-PRS was  $3.92 \pm 0.94$  mmol/L versus  $3.40 \pm 0.81$  mmol/L ( $P < 2.2e-16$ ). These results suggest that individuals with overall high CAD PRS are heterogeneous with respect to different pathways, and PS-PRSs might help identify subgroups developing CAD through specific pathways.

## PgmNr 2674: From GWAS to Mendelian randomization: *UGT1A1* variant identified as causally influencing bilirubin level and reduced risk of hypertension in Africans.

### Authors:

G. Chen; A. Adeyemo; J. Zhou; A. Doumatey; A. Beneley; D. Shriner; C. Rotimi

[View Session](#) [Add to Schedule](#)

**Affiliation:** CRGGH/NHGRI/National Institutes of Health, Bethesda, Maryland.

---

**Objective:** Serum total bilirubin is associated with several clinical outcomes, including hypertension, cardiovascular disease, diabetes and drug metabolism. Here we describe our findings of the first large-scale GWAS of serum bilirubin in continental Africans and investigate the causal effect of bilirubin on hypertension in African ancestry individuals.

**Methods:** Participants included 1,094 West Africans enrolled as part of the African American Diabetes Mellites (AADM) study and 1,739 African Americans enrolled from Washington DC in the Howard University Family Study (HUFs). After genotyping on a high density GWAS array and imputation, a generalized linear mixed model was used to test association with serum bilirubin levels while adjusting for covariates (age, sex, BMI, type 2 diabetes and significant principal components) and relatedness as a random effect. Genome-wide conditional analysis and CAVIARBF were used to fine-mapping significant loci. Mendelian randomization was used to assess the causal effect of bilirubin on hypertension.

**Results:** *UGT1A1* rs887829 was significantly associated with serum bilirubin levels with p value of  $9.13 \times 10^{-54}$  [effect allele (T) frequency 0.40,  $\beta$  (SE) = 0.59 (0.04)] in West Africans, a finding that was replicated in HUFs study with p value of  $9.51 \times 10^{-69}$  [effect allele (T) frequency 0.45,  $\beta$  (SE) = 0.59 (0.03)]. The joint association testing of both study samples yielded a p value of  $3.93 \times 10^{-128}$  and a model  $r^2$  of 0.16. Genome-wide conditional analysis and regional fine mapping resolved rs887829 as a causal variant on bilirubin level with the highest posterior inclusion probability (PIP) of 0.99. Additive Bayesian Networks (ABN) modeling was used to show serum bilirubin level as mediator in the causal relationship between *UGT1A1* rs887829 variant and hypertension. Two-stages least squares Mendelian randomization showed a causal effect of bilirubin associated with lower risk of hypertension with OR of 0.64 [0.51, 0.81]. Cox proportional hazard model showed that bilirubin significantly decreased the risk of hypertension with observational multivariate hazard ratio of 0.55 [0.51,0.60].

**Conclusion:** This first GWAS of serum bilirubin in continental Africans confirms rs887829 (*UGT1A1*) as a major causal locus influencing bilirubin levels. We also show the causal effect of bilirubin associated with reduced the risk of hypertension.

## PgmNr 2675: Complete sequence of the 22q11.2 allele in 1,053 subjects with 22q11.2 deletion syndrome reveals modifiers of conotruncal heart defects near the *CRKL* gene.

### Authors:

Y. Zhao <sup>1</sup>; A. Diacou <sup>1</sup>; H. Johnston <sup>2</sup>; D. McDonald McGinn <sup>3</sup>; G. Repetto <sup>4</sup>; A. Swillen <sup>5</sup>; J. Vermeesch <sup>5</sup>; M. Digilio <sup>6</sup>; B. Marino <sup>7</sup>; C. Bearden <sup>8</sup>; S. Eliez <sup>9</sup>; T. van Amelsvoort <sup>10</sup>; M. van den Bree <sup>11</sup>; D. Gothelf <sup>12</sup>; D. Cutler <sup>2</sup>; E. Zackai <sup>3</sup>; A. Bassett <sup>13</sup>; B. Emanuel <sup>3</sup>; E. Goldmuntz <sup>14</sup>; L. Mitchell <sup>15</sup>; T. Wang <sup>16</sup>; B. Morrow <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Albert Einstein College of Medicine, Bronx, New York.; 2) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 3) Division of Human Genetics, Children's Hospital of Philadelphia and Department of Pediatrics, Perleman School of Medicine, University of Pennsylvania, Philadelphia, USA; 4) Center for Human Genetics, Facultad de Medicina Clinica Alemana-Universidad del Desarrollo, Santiago, Chile; 5) Center for Human Genetics, University of Leuven (KU Leuven), Leuven, Belgium; 6) Department of Medical Genetics, Bambino Gesù Hospital, Rome, Italy; 7) Department of Pediatrics, La Sapienza University of Rome, Rome, Italy; 8) Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California at Los Angeles, Los Angeles, CA, USA; 9) Department of Pediatrics, Gynecology, Obstetrics, La Sapienza University of Rome, Italy; 10) Department of Psychiatry and Psychology, Maastricht University, Maastricht, the Netherlands; 11) Psychological Medicine and Clinical Neuroscience, Cardiff University, Cardiff, Wales; 12) The Child Psychiatry Unit, Edmond and Lily Saffra Children's Hospital, Sackler Faculty of Medicine, Tel Aviv University and Sheba Medical Center, Tel Aviv, Israel; 13) Clinical Genetics Research Program, Centre for Addiction and Mental Health and Department of Psychiatry, University of Toronto, University Health Network, Toronto General Hospital, Toronto, Ontario, Canada; 14) Division of Cardiology, Children's Hospital of Philadelphia and Department of Pediatrics, Perleman School of Medicine, University of Pennsylvania, Philadelphia, USA; 15) Environmental Sciences, University of Texas School of Public Health, Houston, Texas, USA; 16) Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, USA

---

The 22q11.2 deletion syndrome (22q11.2DS) results from non-allelic homologous recombination between low copy repeats termed LCR22. About 65% patients with the typical deletion of 3 million base pairs (Mb) from LCR22A-D have congenital heart disease, mostly of the conotruncal type (CTDs), while others have a healthy heart. In this study, we tried to determine if variants in the hemizygous LCR22A-D region can partially explain such variability in CTDs based upon whole genome sequence data of 1,053 22q11.2DS subjects, all with the same 3Mb LCR22A-D deletion. We found a significant association between CTDs and a cluster of 73 common single nucleotide variants in a single LD block of the 0.7 Mb LCR22C-D interval harboring the *CRKL* gene ( $P$  ranging from  $10^{-3}$  to  $2.59 \times 10^{-5}$ , 63 passed FDR correction, Alternate Allele Frequency  $>0.05$  from logistic regression analyses). Most of the variants confer increased risk to CTDs with a median OR of 2.96, ranging 1.64 from 4.75, while alternate alleles of five are protective with a median OR of 0.51, ranging 0.33 from 0.52. The association signal was not driven individually or collectively by rare variants in the same region. Associations of four variants, representing two independent association signals were consistently

found in three independent cohorts of the general population based on meta-analyses ( $P=6.0\times 10^{-3}$ , OR=0.86 [0.76-0.96] for rs178252;  $P=0.04$ , OR=1.10 [1.00-1.16] for rs178252, 165912 and rs6004160). The 73 variants map to a single topologically associated domain. The variant, rs178252, locates in an open chromatin region and resides in the double-elite enhancer, GH22J010947, predicted to regulate *CRKL* (CRK like proto-oncogene, cytoplasmic adaptor) expression. Approximately 23% of patients with nested LCR22C-D deletions have CTDs and inactivation of *Crkl* in mice causes CTDs, implicating this gene as a modifier. Both rs178252 and rs6004160 are eQTLs of *CRKL*. Further, set based tests identified an enhancer predicted to target the *CRKL* gene that is significantly associated with CTD risk (GH22J020946, SKAT test  $P=7.21\times 10^{-5}$ ) in the 22q11.2DS cohort. These findings suggest CTD penetrance variances among the 22q11.2DS population may be explained in part by disturbance of enhancer-target gene interactions, leading to dysregulation of genes within the LCR22C-D region, in particular for regulation of *CRKL* expression.

## PgmNr 2676: Using eQTL effect size correlation across tissues to cluster coronary artery disease variants.

### Authors:

A. Conklin <sup>1</sup>; L. Stolze <sup>1,2</sup>; M. Whalen <sup>1</sup>; C. Romanoski <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cellular and Molecular Medicine, University of Arizona, Tucson, Arizona.; 2) Genetics Graduate Interdisciplinary Program, University of Arizona, Tucson, Arizona.

A primary methodology in systems genetics is integrating heterogeneous datasets. Interpreting emergent or complex phenotypes relies on synthesizing multiple forms of phenotype data with genotype in order to infer the impact of genetic variants. In order to make inferences about human tissue expression qualitative trait locus (eQTL) data, we implement an effect-size based clustering strategy to associate eQTL data with disease risk variants identified in genome-wide association studies. Our approach relies on clustering eQTLs identified by the Genotype Tissue Expression Consortium using effect size observations across tissues as a proxy for cell-type localization of eQTL effects. Specifically, we use hierarchical density-based spatial clustering of applications with noise to detect clusters of eQTL/eGene associations using correlation as a distance metric. This approach allows for the unsupervised detection of eQTL/eGene association clusters in which cluster detection is dependent on eQTL covariance in effect size across human tissues.

After correcting for linkage disequilibrium, we performed clustering on a random subset (n=250,000) of Genotype Tissue Expression Consortium eQTLs. We found that coronary artery disease associated loci (nominal p-value < 0.05) are enriched in 30 out of 87 eQTL clusters; these 30 clusters contained unique 1728 eQTLs with associations to 114 unique transcripts as 2579 unique eQTL/eGene pairs. We find that one cluster enriched for coronary artery disease risk loci has eQTLs with high effect size in the left ventricle and atrial appendage of the heart and in skeletal muscle. This cluster contains 64 unique eQTLs with associations to 13 unique transcripts as 64 eQTL/eGene pairs.

For future work, we will attempt to determine whether cell-type localization can explain the effect size patterns of clustered eQTLs, and we will investigate whether any epistatic interactions may exist between clustered eQTLs.



## PgmNr 2677: The human leukocyte antigen locus and susceptibility to rheumatic heart disease in Oceanian, South Asian, and European populations.

### Authors:

K. Auckland<sup>1</sup>; B. Mittal<sup>2</sup>; B.J. Cairns<sup>3</sup>; N. Garg<sup>4</sup>; S. Kumar<sup>5</sup>; A.J. Mentzer<sup>1</sup>; J. Kado<sup>6</sup>; M.L. Perman<sup>6</sup>; M. Mirabel<sup>7</sup>; J.K. Kauwe<sup>8</sup>; K.J. Robson<sup>9</sup>; A.C. Steer<sup>10</sup>; A.V.S. Hill<sup>1</sup>; T. Parks<sup>1,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Babasaheb Bhimrao Ambedkar University, Lucknow, India; 3) MRC Population Health Research Unit, University of Oxford, Oxford, United Kingdom; 4) Sanjay Gandhi Postgraduate Institute of Medical Sciences, Lucknow, India; 5) All India Institute of Medical Sciences, New Delhi, India; 6) Fiji National University, Suva, Fiji; 7) Institut National de la Santé et de la Recherche Médicale U970, Paris, France; 8) Brigham Young University, Provo, Utah, United States of America; 9) MRC Weatherall Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford, United Kingdom; 10) Murdoch Children's Research Institute, Melbourne, Australia; 11) London School of Hygiene & Tropical Medicine, London, United Kingdom

---

**Background** Rheumatic heart disease (RHD) is an autoimmune disease of the heart valves that causes an estimated 319,400 deaths and 10.5 million disability-adjusted life-years each year globally. Several candidate gene studies have linked the disease to the human leukocyte antigen (HLA) locus but with little consistency. More recently, two genome-wide association studies (GWAS), one in the Aboriginal Australian population and our own unpublished study in South Asian and European populations, have associated HLA variants and alleles with susceptibility. However, the specific variants implicated in the two studies differ, finding signals in the class II and class III regions, respectively. Moreover, a negligible HLA signal was observed in our earlier GWAS set in the Pacific region. Therefore, in an effort to clarify the role of the HLA locus in susceptibility to RHD, we sought to identify common HLA variants with consistent effects across distinct ancestral groups. **Methods** We undertook a GWAS of RHD susceptibility in 1707 cases and 4060 controls of Oceanian, South Asian and European ancestry. In addition to our previous case-control study from the Pacific region, we included cases of RHD and healthy controls from Lucknow, Northern India, as well as cases of mitral stenosis (a surrogate for RHD) and matched controls from the UK Biobank study. Having performed linear mixed model analyses in each of the seven ancestral strata, we performed fixed and random effects meta-analyses at each of the 4.2 million variants present in all seven groups with frequency of 5% or more. **Results** The lead HLA signal mapped to a missense variant (rs2894232) in *TNXB* located within the class III region (odds ratio 0.80, 95% CI 0.72-0.89,  $P=5.74 \times 10^{-5}$ ). There was little heterogeneity in this signal across populations ( $I^2=0\%$ ,  $P=0.97$ ). This region is upstream of several immune genes, including components of the complement pathway, and has previously been implicated in age related macular degeneration. **Discussion** This analysis is the largest study of RHD susceptibility reported to date, and the first to identify an HLA variant impacting RHD susceptibility with consistent effects across ancestral groups. While offering important new insight into pathogenesis, more work is required to understand the causal relevance of this signal, which is likely to be complex, to RHD susceptibility.



## PgmNr 2678: African-specific variants reveal novel susceptibility genes for carotid intima-media thickness in Sub-Saharan African populations with sex-differences: An AWI-Gen study.

### Authors:

P.R. Boua<sup>1,2,3</sup>; A. Choudhury<sup>2</sup>; H. Sorgho<sup>1</sup>; J.T. Brandenburg<sup>2</sup>; S. Hazelhurst<sup>2,4</sup>; D. Sengupta<sup>2</sup>; H. Tinto<sup>1</sup>; C.G. Mathew<sup>2,5</sup>; M. Ramsay<sup>2,3</sup>; as members of AWI-Gen and the H3Africa Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Research Unit of Nanoro (CRUN)/IRSS, Burkina Faso; 2) Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; 3) Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, South Africa; 4) School of Electrical and Information Engineering, Faculty of Engineering and the Built Environment, University of the Witwatersrand, South Africa; 5) Department of Medical and Molecular Genetics, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom

Atherosclerosis precedes the onset of many clinical manifestations of cardiovascular diseases such as stroke and myocardial infarction which are the leading causes of death worldwide. Consequently, genetic susceptibility to atherosclerosis has been widely investigated, but not yet in sub-Saharan Africans (SSA). Considering the evolutionary roots and the genetic diversity on the continent, GWASs in SSA represent an opportunity to shed new light on understanding the genetic architecture of complex traits such as atherosclerosis. We used carotid intima-media thickness (cIMT), to investigate genetic susceptibility to atherosclerosis in SSA.

Our sample included 7894 middle-aged adults (3963 women, 3931 men) aged 40 to 60 years, from the AWI-Gen study (with participants from East, South and West Africa). cIMT was measured by ultrasound and genotyping was performed on the H3Africa SNV Array (~2.3M SNVs). After imputation using the Sanger Imputation Server African reference panel, we tested for association using BOLT-LMM with adjustment for age, sex and 8 principal components (PCs). In the combined sample we identified two new African-specific genome-wide significant loci (*SIRPA* ( $p=4.7E-08$ ) and *FBXL17* ( $p=2.5E-08$ )). Sex-stratification analysis revealed two male-specific loci (*SNX29* ( $p=6.3E-9$ ) and *MAP3K7* ( $p=5.3E-8$ )), and two female-specific loci (*ARNT2* ( $p=2.4E-09$ ) and *PROK* ( $p=1.0E-08$ )). We replicated regional associations with known loci for atherosclerosis and CVDs, but with different lead SNVs. In a gene-based analysis there were significant associations with *CALD1* ( $p=5.9E-07$ ) for the entire dataset and *FLT4* ( $p=4.3E-07$ ) for female-specific analysis. We found significant enrichment for oestrogen response genes for female-specific signals. The genes identified in our study showed previous biological evidence of involvement with atherosclerosis/CVDs (*SIRPA*, *FBXL17*, *MAP3K*, *ARNT2*, *PROK1*) and sex-difference (*SNX29*, *ARNT2*).

Our study successfully replicated loci and identified novel African-specific variant and gene associations for cIMT. Sex-stratified analyses highlighted sex differences in atherosclerosis risk. Transferability of signals from non-African studies provide opportunities for fine-mapping using African data. Greater inclusion of African populations in medical genomics is important for accelerating

discoveries and identifying new genetic associations with traits for variants absent from other populations.

## PgmNr 2679: Whole genome sequencing and associations with coagulation factors VII and VIII and von Willebrand factor: The TOPMed program.

### Authors:

P.S. de Vries <sup>1</sup>; M.R. Brown <sup>1</sup>; J.E. Huffman <sup>2,3</sup>; L.M. Raffield <sup>4</sup>; B. Rodriguez <sup>3,5</sup>; J.A. Brody <sup>6</sup>; J. Haessler <sup>7</sup>; L.R. Yanek <sup>8</sup>; J.P. Lewis <sup>9</sup>; L. Almas <sup>10,11</sup>; N. Pankratz <sup>12</sup>; X. Guo <sup>13</sup>; A.P. Reiner <sup>7,14</sup>; A.D. Johnson <sup>3,5</sup>; N.L. Smith <sup>14,15,16</sup>; A.C. Morrison <sup>1</sup>; TOPMed Hematology and Hemostasis Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 2) VA Boston Healthcare System, Jamaica Plain, MA; 3) The NHLBI's Framingham Heart Study, Framingham, MA; 4) Department of Genetics, University of North Carolina, Chapel Hill, NC; 5) NHLBI's Population Science Branch, Framingham, MA; 6) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, WA; 7) Division of Public Health, Fred Hutchinson Cancer Research Center, Seattle, WA; 8) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 9) Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 10) Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, PA; 11) Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 12) Department of Laboratory Medicine and Pathology, School of Medicine, University of Minnesota, Minneapolis, MN; 13) Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, CA; 14) Department of Epidemiology, University of Washington, Seattle, WA; 15) Kaiser Permanente Washington Health Research Institute, Kaiser Permanente Washington, Seattle, WA; 16) Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle, WA

---

**Background:** Coagulation factor VII (FVII) and factor VIII (FVIII), and its carrier protein von Willebrand factor (vWF) are implicated in modulating the risk of arterial and venous thrombosis. The NHLBI's Trans-Omics for Precision Medicine (TOPMed) program brings together extensive whole genome sequence (WGS) resources and hemostasis phenotypes, and capitalizes on advances in computational analysis to better understand the genetic architecture of hemostatic factors. **Methods:** We leveraged Freeze 6 WGS from TOPMed. Plasma levels of FVII (n= 16,335), FVIII (n= 19,766), and vWF (n= 14,020) were harmonized across 9 studies that included participants of European, African, Asian, and Hispanic ancestry. Association analyses were conducted across all individuals using inverse normalized and rescaled residuals adjusting for age, sex, ancestry, principal components, and a kinship matrix. Analyses were conducted on the Analysis Commons cloud computing platform using the SMMAT function implemented in GENESIS. Single-variant analyses assessed all variants with a minor allele count  $\geq 40$ . Aggregate analyses grouped low-frequency and rare variants (minor allele frequency [MAF]  $< 0.05$ ) by gene, using 3 strategies for selection of variants within gene-based aggregation units: 1) loss of function (LOF) variants; 2) LOF and deleterious missense variants; and 3) coding, enhancer and promoter variants. **Results:** Single-variant analyses identified significant associations ( $P < 5E-8$ ) at 4 known loci for FVII, 8 for FVIII, and 8 for vWF. New FVIII associations

included rs538727675 located between *FNDC3B* and *GHSR* (MAF=0.0015; P=2.2E-8) and rs114894279 downstream of *TBL1XR1* (MAF=0.012; P=3.4E-8). A new vWF association was identified with rs147142418 in *DPF3* (MAF=0.017; P=1.1E-8). Conditional analyses revealed multiple independent signals at *F7*, *VWF*, *STAB2*, and *ABO*. Gene-based aggregate analyses identified associations at 1 known locus for FVII (*F7*), and 3 known loci each for vWF and FVIII (*VWF*, *STAB2*, and *ABO*). Additionally, LOF variants in *CD36* were associated with FVIII, representing a novel locus for FVIII. The driving variant, rs3211938, causes CD36 deficiency and is associated with a range of hematological phenotypes. **Conclusions:** WGS analysis of hemostatic factors yielded novel genetic associations with FVIII and vWF. Replication of these findings will be completed in up to 30,000 individuals from studies with imputed genotypes based on a TOPMed reference panel.

## PgmNr 2680: Trans-ethnic meta-analysis of genetic determinants of blood pressure trajectory.

### Authors:

M.H. Gouveia <sup>1</sup>; A.R. Bentley <sup>1</sup>; M.A. Nalls <sup>2</sup>; L. Hampton <sup>2</sup>; C. Guanjie <sup>1</sup>; E. Tarazona-Santos <sup>3</sup>; M.F. Lima-Costa <sup>4</sup>; A. Adeyemo <sup>1</sup>; D. Shriner <sup>1</sup>; C.N. Rotimi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, United States of America; 2) Data Tecnica International, Glen Echo, Maryland, USA; 3) Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; 4) Instituto de Pesquisa René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, Minas Gerais, Brazil

---

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with cross-sectional blood pressure (BP) traits, however, there are limited GWAS of longitudinal BP data. We performed GWAS followed by meta-analysis of longitudinal BP traits using 21,900 multi-ethnic individuals from four studies: i) the Atherosclerosis Risk in Communities Study (ARIC); ii) the Bambui-Epigen Cohort Study of Aging (Bambui, Brazil); iii) the Health, Aging and Body Composition Study (Health ABC) and iv) the Multi-Ethnic Study of Atherosclerosis (MESA). We found 10 new loci genome-wide significantly associated with longitudinal BP outcomes, and most of these were ancestry specific with six found in African Americans, three in European Americans and one in Chinese and Hispanic American individuals. The strongest signal was the intronic rs140659580-C allele associated with higher systolic BP average (mean of measurements over time) ( $p = 1.45 \times 10^{-10}$ ) in Hispanic and Chinese American individuals from the MESA study. rs140659580 is in *GLG1* which is highly expressed in the aorta and coronary arteries and was previously associated with varicose veins. Furthermore, the T allele of rs73601659 was associated with increasing systolic BP trajectory (rate of change over time) ( $p = 4.10 \times 10^{-8}$ ) in African Americans. This variant is in *CHD9*, which has been previously associated with baseline plasma renin activity and plasma fibrinogen concentration, which are established predictors of hypertension. By comparing the meta-analysis results using longitudinal (BP trajectory and average) and baseline traits, we observed a highest correlation between the results using baseline and average outcomes ( $\rho=0.48$ ) and as expected, trajectory was more correlated with average ( $\rho=0.35$ ) than with baseline results ( $\rho=0.18$ ). In summary, in this study longitudinal data of cohorts of diverse ancestry facilitated discovery of novel BP signals, most of which were ancestry-specific. Notably, and despite some degree of overlap between the associations using longitudinal and cross-sectional outcomes, there is evidence for different genetic associations underlying these traits.

## **PgmNr 2681: Identification of disease susceptibility through 3,283 protein-polygenic-risk-score phenome-wide association studies in Penn Medicine BioBank electronic health records.**

### **Authors:**

Y. Ko <sup>1</sup>; G. Sirugo <sup>1</sup>; M. Guerraty <sup>1</sup>; R. Kember <sup>2</sup>; R.L. Judy <sup>1</sup>; B. Sun <sup>3</sup>; M. Vujkovic <sup>1</sup>; S. Sheth <sup>1</sup>; D. Marchadier <sup>1</sup>; A. Munshi <sup>1</sup>; N. Hand <sup>1</sup>; A. Butterworth <sup>3</sup>; S. Damrauer <sup>2</sup>; D. Rader <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Institute for Translational Medicine and Therapeutics, University of Pennsylvania School of Medicine, Philadelphia, PA; 2) Department of Surgery, University of Pennsylvania School of Medicine, Philadelphia, PA; 3) NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK

Polygenic risk scores (PRS) have recently shown promise for predicting disease susceptibility based on cumulative effects of genetic variants. Here we generated polygenic scores using the protein genome-wide association studies (GWAS) in the UK INTERVAL study. Although many of these proteins have not previously been associated with disease, we tested a protein polygenic score (PPS) against all phenotypes in the Penn Medicine BioBank (PMBB), by integrating the PPS with corresponding electronic health records (EHRs).

The protein GWAS in INTERVAL, also described as protein quantitative trait loci (pQTL), were analyzed on the SOMAscan platform to quantify plasma proteins. SOMAscan is an aptamer-based assay that profiles more than 3,000 proteins for each sample, enabling unbiased discovery of potential biomarkers for disease.

We generated PPS using LDpred method for the 3,283 proteins GWASs from INTERVAL, a cohort of 3,301 European subjects, using summary statistics. We first validated the PPS against a subset of PMBB samples for which we had measured plasma protein levels (protein n=565). We then tested the association between PPS and PMBB EHR.

Working with the Regeneron Genetics Center, we obtained genome-wide SNP genotype data of 17,560 individuals in PMBB (n=5,988 African Americans, n=11,580 European ancestry after quality control). EHRs were available for each from the University of Pennsylvania Health System. Adopting the phenome-wide association study (PheWAS) approach, we asked which phenotype in the EHR is associated with each PPS, across nearly 2,000 phenotypes aggregated from the International Classification of Disease 9th edition (ICD9) codes to representative phenotypes.

We identified a set of significant associations with many congenital disease phenotypes, including Forkhead box protein C2 (FOXC2), a known transcription factor mutated in congenital heart disease. In the PheWAS analysis, FOXC2 was shown to be associated with valvular heart disease and diseases that affect the heart chambers.

In summary, the novel method of PPS-PheWAS analysis connects cumulative genetic effects to disease through integration of large-scale EHR data with plasma proteome. This approach allows us to directly identify disease susceptibility of the genetically determined protein values.

## PgmNr 2682: Interrogating the proteome to identify putatively causal protein biomarkers of platelet count and mean platelet volume: The Framingham Heart Study.

### Authors:

D. Lee <sup>1</sup>; C. Yao <sup>1</sup>; A. Bhan <sup>2</sup>; T. Schlaeger <sup>2</sup>; J. Keefe <sup>1</sup>; B. Rodriguez <sup>1</sup>; S. Hwang <sup>1</sup>; M. Chen <sup>1</sup>; D. Levy <sup>1</sup>; A. Johnson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Framingham Heart Study, National Heart, Lung, and Blood Institute, Framingham, MA; 2) Department of Pediatrics, Boston Children's Hospital, Boston, MA

---

Mean platelet volume (MPV) and platelet count (PLT) are commonly used platelet metrics associated in some studies with cardiovascular disease (CVD) and mortality. Identifying causal protein biomarkers for MPV and PLT may yield insights into CVD mechanisms and improve platelet regeneration therapies or platelet transfusion outcomes. To this end, we conducted integrative analyses of GWAS hits for PLT and MPV with protein quantitative trait loci (pQTLs) of plasma proteins, Mendelian randomization (MR) for protein to PLT/MPV inference, and protein-PLT/MPV trait association in the Framingham Heart Study (FHS). Seventy-one proteins were selected based on prior associations with CVD. A total of 4,348 participants (52% female, mean age 58 years) were included in the protein-trait association analyses for MPV and 4,272 participants (53% female, mean age 58 years) were included for PLT. Among 71 proteins, 11 were associated with MPV and 31 were associated with PLT after Bonferroni correction ( $P < 0.05/71$ ). We identified 15 pQTLs (4 *cis* and 11 *trans*) that overlapped with MPV GWAS SNPs and 17 pQTLs (5 *cis* and 12 *trans*) that overlapped with PLT GWAS SNPs at  $P < 1.25e-07$  for *cis* and  $P < 7.04e-10$  for *trans*. MR analyses identified 4 putatively causal proteins for MPV and 4 proteins for PLT at  $P < 0.05/37$ . Integrating all the results identified 3 putatively causal proteins for PLT. Glycoprotein V (GP5) and granulin (GRN) were positively correlated with PLT and melanoma cell adhesion molecule (MCAM) was negatively correlated with PLT in both protein-trait association and MR analyses. GP5 is a well-known receptor on platelets, but MCAM and GRN are not well-characterized as platelet proteins. Utilizing iPSC-derived megakaryocyte clones that produced variable numbers of functional platelets, we conducted megakaryocyte RNA-sequencing and analyzed transcriptome-wide differences between low- and high-platelet producing clones. Results were concordant with our population-based causal associations; GRN and GP5 were over-expressed and MCAM was down-regulated in high-platelet producing clones. In total, the results suggest that GP5, GRN and MCAM are causally linked to platelet generation and/or turnover and may play important roles in CVD via a platelet-based mechanism. These may also represent potential therapeutic biomarkers or targets for CVD prevention, and genes that could be manipulated to increase platelet production in bioreactors for transfusion medicine purposes.

## PgmNr 2683: A genome wide association study on postural orthostatic tachycardia syndrome.

### Authors:

H. Qu <sup>1</sup>; J. Glessner <sup>1</sup>; N. Williams <sup>1</sup>; J. Qu <sup>1</sup>; X. Chang <sup>1</sup>; A. Wenocur <sup>1</sup>; F. Mentch <sup>1</sup>; M. Lemma <sup>1</sup>; M. Behr <sup>1</sup>; L. Tian <sup>1</sup>; P. Sleiman <sup>1,2,3</sup>; J. Boris <sup>4</sup>; H. Hakonarson <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.; 2) The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; 3) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; 4) Pediatric Cardiologist, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania

---

Postural orthostatic tachycardia syndrome (POTS) causes significant functional impairment and psychological distress to the patients, and is affecting about 0.2% to 1.0% US population. The pathophysiology of POTS has been described as being a failure of ability to maintain vascular tone with a compensatory tachycardia. There is associated chronic orthostatic intolerance and severe debilitating symptoms. However, the actual genetic mechanism of POTS is not understood. We performed a genome wide association study (GWAS) to determine the genetic underpinnings of POTS. We recruited 465 patients diagnosed with POTS and available family members. We obtained a biological sample (blood or saliva) for DNA isolation and GWAS analysis using the Illumina Infinium Global Screening Array. We performed transmission disequilibrium testing (TDT) in available family trios. With Bonferroni correction, significance was defined as  $7.14E-08$ . TDT analysis was performed in 99 family trios, and case control analysis in 207 unrelated patients versus 4,328 ethnicity-matched controls. An intronic variant in the gene encoding phosphorylase 1 regulatory subunit 12B (*PPP1R12B*) for myosin, located at chromosome 1q32.1, demonstrated significance at  $P=5.44E-16$ . This was replicated in both independent TDT and case-control analysis. Also, a non-coding variant of human zinc ribbon domain antisense RNA 1 (*ZNRD1-AS1*), in the major histocompatibility class (*MHC*) I region located on chromosome 6, demonstrated significance at  $P=1.82E-09$ . *PPP1R12B*, which dephosphorylates the regulatory light chain of myosin II, could be a clinically plausible gene for POTS, as myosin II functions in multiple locations in the body, including myocardium and vascular smooth muscle. *ZNRD1-AS1* is a long non-protein coding RNA. Many patients have onset of symptoms after an infection or concussion, suggesting an autoimmune etiology involving the *MHC* region. While *ZNRD1-AS1* is considered to be non-coding in human, its variants could interfere with RNA expression of other genes. Further exome and RNA sequencing is underway to delineate these loci to help determine further clinical validity.



## PgmNr 2684: Meta-analysis of 26,643 individuals identifies two genetic loci associated with left ventricular ejection fraction.

### Authors:

C. Schaefer<sup>1</sup>; K.K. Thai<sup>1</sup>; C. Jiang<sup>1</sup>; D.K. Ranatunga<sup>1</sup>; T.J. Hoffmann<sup>2,3</sup>; A.S. Go<sup>1</sup>; M.G. Ehm<sup>4</sup>; D.M. Waterworth<sup>4</sup>; A.C. Lindsay<sup>4</sup>; N. Risch<sup>1,2,3</sup>; H. Choquet<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Kaiser Permanente Northern California (KPNC), Division of Research, Oakland, CA 94612, USA; 2) Institute for Human Genetics, UCSF, San Francisco, CA 94143, USA; 3) Department of Epidemiology and Biostatistics, UCSF, San Francisco, CA 94158, USA; 4) GlaxoSmithKline, King of Prussia, PA 19406, USA

**Purpose:** Left ventricular ejection fraction (EF) is an indicator of cardiac function, assessed in individuals with heart failure and other cardiac conditions. Although family studies indicate that EF has an important genetic component, with heritability estimates up to 0.61, to date only 3 EF-associated loci have been reported in a Japanese cohort. Here, we conducted a genome-wide association study (GWAS) of EF in the multiethnic Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort.

**Methods:** Our discovery GWAS analysis included 22,160 GERA participants from 4 ethnic groups (African American, East Asian, Hispanic/Latino, and non-Hispanic white) having EF measured. For each participant, the 1<sup>st</sup> EF measurement was extracted from electronic health records. Genetic association analyses with EF were performed using linear regression adjusted for age, sex, and ancestry principal components. Confirmation of GERA findings was conducted in the UK Biobank (UKB) European sample (N=4,483 individuals) and a meta-analysis combining GERA and UKB results was performed. We replicated meta-analysis findings using summary statistics from a published GWAS of EF conducted in the Biobank Japan Project (BBJ) (N=19,516). Sensitivity analyses of the effects of EF-associated loci on EF adjusting for antecedent conditions (heart failure/cardiomyopathy, hypertension, myocardial infarction, or atrial fibrillation) were performed in GERA.

**Results:** Our discovery GWAS analysis across the 4 GERA ethnic groups identified a unique locus, *BAG3*, significantly associated with EF (lead SNP rs17617337;  $P=4.72 \times 10^{-11}$ ). The association between *BAG3* rs17617337 and EF replicated in UKB at Bonferroni significance ( $P=3.42 \times 10^{-7}$ ). A meta-analysis combining GERA and UKB identified an additional locus: *TMEM40* (rs11719526;  $P=3.10 \times 10^{-8}$ ). The association between *TMEM40* rs11719526 and EF replicated in BBJ ( $P=0.016$ ). In GERA, *BAG3* rs17617337 was associated with heart failure/cardiomyopathy and *TMEM40* rs11719526 was associated with atrial fibrillation. When we included these heart conditions as covariates in the models, the associations between *BAG3* rs17617337 or *TMEM40* rs11719526 and EF remained significant in GERA.

**Conclusions:** This largest study conducted to date on EF reports 2 novel loci, including *BAG3*, consistent with prior work involving this gene in cardiac conditions. However, sensitivity analyses demonstrated significant effects of the lead SNPs on EF independently of these conditions.

## PgmNr 2685: Genome-wide genotype imputation using SNP dense arrays and meta-analysis implicates *RTN4* variants association with blood pressure traits in Arab population.

### Authors:

A.T. Thangavel <sup>1</sup>; P. Hebbar <sup>1</sup>; S.E. John <sup>1</sup>; A. Channanath <sup>1</sup>; J.A. Abubaker <sup>1</sup>; M. Abu-Farha <sup>1</sup>; J. Pitkaniemi <sup>2</sup>; J. Tuomilehto <sup>2</sup>; F. Al-Mulla <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dasman Diabetes Institute, Kuwait, Kuwait, Kuwait; 2) University of Helsinki, Helsinki, Finland

**Aim:** Conglomeration of low levels of physical activity and high calorie food intake in Middle East resulted in a high prevalence of hypertension along with obesity. Global genetic studies have implicated many single nucleotide variations for hypertension however, Arab population lacks in convincingly determined genetic variants for hypertension. In this study, we imputed genome-wide genotypes using in-house genotype data from two genotyping arrays and then performed statistical tests for associations and followed by performing meta-analysis. The study implicates rs2920844\_T, rs2968781\_G of *RTN4* gene with elevated levels DBP and SBP.

**Method:** we used genome-wide genotype data of 1434 and 1298 unrelated individuals of Arab ethnicity genotyped on Illumina Cardio-Metabo BeadChip and Illumina HumanOmniExpress BeadChip respectively to perform genome-wide genotype imputation using 1KGP as reference panel. We harnessed quality genotypes (RSQ>0.5 with MAF≥5%) from imputed data and performed statistical association analysis (using RvTests software) on DBP and SBP traits. Further, we combined association results of both data sets with meta-analysis using Metal program. We considered variant-trait association signals with genome wide p-value threshold <5E-08. In addition, we performed fine mapping analysis using the FINEMAP tool and formed 95% credible causal SNPs around the top signals.

**Result:** Imputation of genotyped markers generated 6.08 million (with a mean RSQ value of 0.822) and 1.52 million variants (with a mean RSQ value of 0.759 and mean imputation quality score >0.50 and MAF≥ 5%) in Illumina OmniExpress and Cardio-Metabo data respectively. Statistical association analysis followed by meta-analysis of two data sets showed association of rs2920844\_T (DBP:beta=5.617, p-value=1.94E-08; SBP: beta=5.483, p-value=4.18E-08), rs2968781\_G (DBP:beta=5.618,p-value= 1.93E-08; SBP: beta= 5.688,p-value= 1.29E-08) of *RTN4* gene with elevated levels of blood pressure traits. Both the effect alleles were alternate alleles and showed a frequency of 15% in imputed data. Further, fine mapping analysis included them in 95% credible set of causal variants with posterior inclusion probability 0.45.

**Conclusion:** Literature evidences suggest that reticulon 4 (*RTN4*) is implicated in coronary heart disease. Isoforms of *RTN4* have been demonstrated to regulate blood pressure. Hence, the rs2920844\_T and rs2968781\_G of *RTN4* could be novel regulators of blood pressure in Arab population.

## PgmNr 2686: Non-genetic and genetic risk factors for aortic stenosis in patients with a bicuspid aortic valve.

### Authors:

T. Nguyen<sup>1</sup>; X. Xu<sup>1</sup>; C.S. Ong<sup>2</sup>; J. Shahram<sup>3</sup>; J. Gosnell<sup>3</sup>; S. Shrestha<sup>3</sup>; J. Muehlschlegel<sup>3</sup>; T. Sundt<sup>2</sup>; S. Body<sup>1</sup>; BAV Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Beth Israel Deaconess Medical Center, Boston, MA; 2) Massachusetts General Hospital, Boston, MA; 3) Brigham and Women's Hospital, Boston, MA

---

### Introduction

Bicuspid aortic valve (BAV) is the most common congenital valvular heart disease and causes ~40% of aortic stenosis (AS) requiring aortic valve replacement (AVR) surgery. AS in the BAV population presents at a younger age, with AVR occurring ~15 years earlier than those with a tricuspid aortic valve (TAV) (1,2). In TAV disease, genetic risk factors for AS include variation in the number of Kringle IV repeats in the *LPA* gene that encodes the Apo(a) protein, determining both protein function and also levels of the lipoprotein(a) particle, a low-density lipoprotein. Its role in AS in BAV individuals has not yet been reported. We investigated non-genetic and genetic risk factors of AS in the BAV population.

### Methods

We examined 3,249 individuals with echocardiography-confirmed BAV to identify risk factors for AS. Patients were identified from a health system-wide registry and included individuals age >20 years with detailed aortic valve phenotyping of BAV and AS. Separately, we genotyped individuals with the Illumina GSA array to identify genetic risk factors for AS, including the *LPA* locus.

### Results

Of the 3,249 patients with BAV, 1,300 (40%) had no to mild AS, 398 (12%) had moderate AS, and 1,551 (48%) had severe AS. Multinomial logistic regression showed that AS severity increased with advancing age and COPD due to smoking (OR=1.38, CI 1.06-1.80), but not with gender (OR=0.88, CI 0.74-1.05). A unicuspid valve phenotype was associated with increased AS severity (OR=6.52, CI 3.43-12.4), while the aortic valve leaflet configuration (R-L vs R-Non) was not. Genotyping results will be provided at the meeting.

### Conclusions

In patients with BAV, more severe aortic stenosis is associated with older age, a unicuspid aortic valve phenotype, and history of COPD. Valve configuration was not associated with more severe aortic stenosis. In contrast to published associations of AS in tricuspid aortic valves, we did not observe associations with gender, diabetes and other cardiovascular risk factors.

## PgmNr 2687: Importance of the additional risk factors for the heart failure manifestation and progression in patients carrying mutations in the *TNNT2* “hot” codon.

### Authors:

E. Zaklyazminskaya; A. Bukaeva; G. Radzhabova; M. Polyak; A. Dombrovskaya; Y.U. Frolova; S. Dzemeshkevich

[View Session](#) [Add to Schedule](#)

**Affiliation:** Petrovsky Russian Research Centre of Surgery, Moscow, Russian Federation

**Introduction.** Non-ischemic dilated cardiomyopathy (DCM) is a group of progressive myocardial disorder with a poor prognosis. The probability of death or heart transplantation is 70% in 5 years after initial diagnosis. But clinical manifestation and progression rate may vary significantly within members of the family. Up to 80 different genetic forms of DCM are known with a predominantly autosomal dominant mode of inheritance. Even these genetic forms can be significantly influenced by an individual's risk factors and present mixed pathogenesis.

The aim of this study was to demonstrate the importance of gene-environmental interaction in patients with DCM carrying pathogenic variants affecting codon 173 in the *TNNT2* gene. Position Arg173 places in T1-fragment of the cTnT, highly conservative, and when mutated alter sarcomere organization, Ca<sup>2+</sup>-regulation and contractility.

**Methods.** The cohort included 92 DCM probands. Clinical evaluation included personal and family history taking, general examination, ECG, Echo-CG, cardiac MRI. Genetic testing was performed by NGS of the target panel flanking 81 DCM-associated genes with following Sanger sequencing of all clinically significant findings, and cascade familial screening.

**Results.** We revealed three heterozygous pathogenic variants in the *TNNT2* gene affecting codon 173: (p.Arg173Gln, p.Arg173Leu, and p.Arg173Trp) in 3 index cases diagnosed with familial DCM. Cascade familial screening revealed 3 more mutation carriers. Two patients have had well-known risk factors associated with DCM. Two male probands (p.Arg173Leu and p.Arg173Trp) were initially diagnosed with toxic cardiomyopathy soon after chronic alcohol consumption at 40 and 38 y.o., respectively) and one female (p.Arg173Leu, younger daughter) was diagnosed with peripartum cardiomyopathy at 20 y.o. Heart transplantation was recommended to all of them due to rapid heart failure progression. Carrier of p.Arg173Trp variant (39 y.o.), his daughter (11 y.o.), and female carrier of the p.Arg173Leu variant (29 y.o.) had no identifiable risk factors. Contractility and LV ejection fraction was slightly decreased but preserved for years.

**Conclusion.** The mutation rate in the *TNNT2* gene is 3.2% in the Russian DCM cohort. We suggest that “second hit” (genetic or environmental) might be crucial for rate of cardiac remodeling in patients with mutations affecting codon 173 of the *TNNT2* gene.

This study was supported by Russian Science Foundation research grant ?16-15-10421

## PgmNr 2688: A systems genetics and integrative biology approach: Discovery of novel heart disease biology and putative heart disease drug target candidates.

### Authors:

A. Walentinsson; B. Tyrberg; C. Kumar

[View Session](#) [Add to Schedule](#)

**Affiliation:** Translational Science & Experimental Medicine, Early Cardiovascular, Renal and Metabolism, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden

Diseases of the heart constitute a major health care burden encompassing a range of comorbid conditions including Myocardial Infarction (MI), Coronary Artery Disease (CAD), Heart Failure (HF), and Atrial Fibrillation (AF). Being *multifactorial* in nature, these conditions result from multiple additive gene contributions (*polygenic*) in combination with lifestyle and environmental factors. In the past decade, there has been major progress in identifying genetic loci contributing to heart disease risk. Genome-wide association studies (GWAS) have identified >160 loci for CAD and >30 for AF, whereas other heart conditions, like HF, is less studied. Still, the majority of GWAS SNPs reside in non-coding regions and the causal genes remain largely unknown.

In this study we set out to identify and characterize causal biology in human heart disease(s) leveraging public domain genomics data. To pinpoint casual genes at heart disease GWAS loci likely to act through the heart ("cardiac effector genes"), we cross-interrogated GWAS SNPs (GWAS Catalogue,  $p \leq 5.0 \times 10^{-8}$ ), along with their proxies, against expression quantitative trait loci (eQTL) from heart ventricle, atrium and coronary arteries (GTEx Consortium).

We identified a total of 87 cardiac effector genes distributed across 49 genomic loci, from 387 GWAS SNPs (9446 with proxies) originating from 54 heart disease studies. The number of identified genes per disease category was as follows; 58 (67%) CAD, MI and/or Coronary Artery Calcification, 10 (11%) AF, 8 (9%) Sudden Cardiac Arrest, 5 (6%) HF, 5 (6%) Cardiac Hypertrophy and/or Left Ventricular Function and Mass and 1 (1%) Dilated Cardiomyopathy. Twenty-three (26%) genes are non-protein coding (long non-coding-, antisense RNAs or transcribed pseudogenes), 63 (72%) are novel in the context of Cardio-Vascular Disease biology, and 4 (5%) are targets for approved drugs.

To identify causal biological mechanisms, cardiac effector genes were further analyzed in a pathway and network biology context. Biological themes emerging from these analyses could provide novel mechanistic entry points into treatment of heart disease.

In summary, retrospective review of drugs in development demonstrates that those targeting proteins with a genetic underpinning are more likely to progress to marketed drugs. Hence, the present findings not only provide novel insights into hitherto unexplored heart disease biology but also expands the repertoire of novel drug target candidate genes.

## PgmNr 2689: The causal link between rheumatoid arthritis, usual walking pace, and coronary heart disease: A network Mendelian randomization study.

### Authors:

T.L. Yang; S.S. Dong; J.C. Feng

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China

---

It is believed that patients with rheumatoid arthritis (RA) have a 1.5-2.0 fold increased risk of developing heart diseases. Due to joint pain/stiffness or other joint problems, physical fitness (e.g. walking speed) is often reduced in RA patients. However, the causal link between RA, usual walking pace, and CAD remains unclear. Here we aim to examine whether usual walking pace mediate the causal pathway from RA to CAD using a network Mendelian randomization design. Summary statistics from genome-wide association studies (GWASs) were used for 2-sample Mendelian randomization analyses. The European GWAS of 14,361 RA cases and 43,923 controls were used to identify instruments and genetic association estimates for RA. GWAS summary data for usual walking pace (N = 452,264) of the UK Biobank cohort were obtained from the GenAtlas database. Summary data for CAD from the meta-analysis of UK Biobank cohort (34,541 CAD cases and 261,984 controls) and CARDIoGRAMplusC4D (88,192 cases and 162,544 controls) were downloaded from CardiOmics.net. Inverse variance weighted (IVW) analysis was used as the main method for MR estimates. As complementary methods, MR-Egger, weighted median, and weighted mode analyses were also performed. Potential horizontal pleiotropy was assessed by MR-Egger and MR-PRESSO. MR analyses showed that RA was positively associated with the risk of CAD (OR = 1.0227, CI: 1.0122 to 1.0333,  $P = 2.06 \times 10^{-5}$ ) and negatively associated with usual walking pace (OR = 0.9959, CI: 0.9939 to 0.9979,  $P = 8.49 \times 10^{-5}$ ). Significant negative causal effect of usual walking pace on CAD was also observed (OR = 0.3859, CI: 0.3072 to 0.4849,  $P = 2.89 \times 10^{-16}$ ). Mediation analysis showed that usual walking pace might account for about 15% of the association between RA and incident CAD. In conclusion, our results indicate that RA may be causally associated with slower walking pace, which mediates a significant proportion of the relationship between RA and CAD.



# PgmNr 2690: Detection of mutations associated with thrombophilia in the *MTHFR*, *F2* and *F5* genes in Colombian patients: Incidence in diagnosis, treatment and cost for the Colombian health system.

## Authors:

A.J. Muñoz Bolaños <sup>1</sup>; Y. Gómez P <sup>1</sup>; S.J. Maradei Anaya <sup>2</sup>; T. Bernal <sup>3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Biotecnología y Genética S.A., Bogotá, Colombia, Bact., MSc.; 2) Biotecnología y Genética S.A., Bogotá, Colombia, MD, MSc.; 3) Biotecnología y Genética S.A., Bogotá, Colombia, MD

---

**Introduction.** Factor V Leiden (FVL) mutation is the most common cause of hereditary thrombophilia (TP) among the white population. The FVL carrier frequency ranges from 2% in Southern Europe to 15% in southern Sweden. Heterozygous carriers frequency is estimated to be 2.2% in Hispanics. The ethnic distribution of the prothrombin G20210A is similar to that of FVL. It is estimated that >25% of Hispanics and between 10 and 15% of North American Caucasians are homozygous for the “thermolabile” variant of *MTHFR* (C677T).

**Methods.** Results of *MTHFR*, prothrombin and FVL mutational tests from 150 patients with TP diagnosis were included. Tests were performed in a private genetics laboratory of Bogota, Colombia. This is a descriptive study in which mutations C677T and A1298C were investigated in *MTHFR*, G20210A mutation of *F2* and R506Q of *F5*, using real-time PCR. Demographic and clinical data from patients were correlated with tests results, according to ACMG guidelines recommendations.

**Results:** 40 blood samples were tested for *MTHFR* C667T mutation, 12 samples for *MTHFR* A1298C mutation, 64 samples for *F2* G202110A mutation and 34 samples for *F5* R506Q mutation. 75% of the *MTHFR* tests were positive for the C667T mutation and 42% for A1298C mutation. On the contrary, only 6,25% and 3% were positive for *F2* G202110A mutation and *F5* R506Q mutation respectively.

**Discussion.** The global incidence of TP is high so laboratory testing for hereditary TP and selected acquired TP is common. A wide variety of both non-genetic and genetic factors have been recognised as predisposing risk factors for thrombosis. Given the costs of testing and multiple pre-analytic and analytic variables affecting it, careful patient selection and timing of testing and diligent application to patient management are critical to providing high-value clinical care. Collaboration between the ordering providers and performing laboratories has the potential to achieve these goals. Data from several studies strongly suggest that the pathogenesis of thrombosis is multifactorial and requires interactions between both inherited and acquired risk factors. We found that, despite the recommendations of the ACMG, physicians continue to order mutational studies for the *MTHFR* gene in thrombosis patients. Likewise, they order mutational studies for the *F2* and *F5* genes without adhering to the proper selection criteria of patients, causing additional and unnecessary costs to the Colombian health system.





## PgmNr 2691: Integration of polygenic risk score in CAD risk models for clinical use.

### Authors:

A. Bolli <sup>1</sup>; A. Di Rocco <sup>2</sup>; P. Di Domenico <sup>1</sup>; G. Busby <sup>3</sup>; G. Bottà <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Allelica, Rome, Italy; 2) University of Rome 'La Sapienza', Rome, Italy; 3) Big Data Institute, University of Oxford, Oxford, UK

Different groups have developed Polygenic Risk Scores (PRS) for Coronary Artery Disease (CAD) based on inference from millions of Single Nucleotide Polymorphisms (SNP). In one study, a PRS-based model was able to identify the 8% of the individuals of the European population with a CAD risk, comparable to that conferred by monogenic mutations. In a second study, men in the top 20% of PRS distribution reached a threshold of 10% cumulative CAD risk by 61 years of age, ten years earlier than men in the bottom 20% distribution.

However, although PRS-based risk models demonstrate higher predictive performances than any currently used risk factor they are not yet used in clinical practice. We therefore addressed the following questions to build the use case for the use of PRS into current clinical risk models. Following analyses are based on the UK Biobank population of European descent.

First, we assessed the impact of adding published PRSs to clinical risk models (e.g., Framingham score). At present, a 10% 10-year risk threshold is used in risk models to decide whether life-style or medical interventions should be implemented. We found that the addition of PRS to the Framingham score determined consistent reclassification of individuals based on the 10% risk threshold.

Second, we tested if PRSs were able to stratify individuals already considered at risk on the basis of conventional risk factors. For men with family history of CAD for example, we detected a strong PRS-based stratification with CAD prevalence ranging from 2.5% to 23% for the first and last percentile of PRS distribution. This indicates that even in individuals with family history, risk trajectories can be substantially modulated by the addition of PRSs. Additionally, we find no correlation between PRS and LDL-cholesterol blood levels indicating that these are orthogonal risk factors.

Finally since the majority (99.9%) of SNPs composing such PRSs have a low median effect size ( $10^{-6}$ ), researchers have argued about the relevance of such SNPs in contributing to the overall PRS predictive performances. By generating PRS with different subsets of SNPs, we found that full set PRS (6.6 millions SNPs) displays higher Positive Predictive Values compared with either effective-SNPs PRS (top 0.1% SNPs) or genome-wide significant PRS (74 SNPs). This shows that even low weight SNPs play a role in classifying a large proportions of individuals into the higher-risk CAD category.

## **PgmNr 2692: A polygenic risk score for atrial fibrillation is associated with the development of post-operative atrial fibrillation following cardiac surgery, independent of a patient's history of atrial fibrillation.**

### **Authors:**

M. Shuey; E. Farber-Eger; J.D. Mosley; F.T. Billings; N.J. Cox

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt University Medical Center, Nashville, TN.

---

**Background:** Post-operative atrial fibrillation (PoAF) is the most common complication of cardiac surgery, occurring in 20 to 40% of patients. The clinical risk factors of PoAF are well described, however, recent work has attempted to incorporate genetic variants into prediction models. Our objective was two-fold: 1) develop an algorithm to identify adult cardiac surgery patients and determine PoAF status using the Vanderbilt University Medical Center biobank (BioVU) and 2) develop a polygenic risk score (PRS) of AF and test its ability to predict PoAF.

**Methods:** Using current procedural codes and natural language processing we developed an algorithm capable of identifying adult patients that developed PoAF following cardiac surgery and prior to hospital discharge. Algorithm performance was validated by blinded chart review. Next using linkage disequilibrium-driven clumping and thresholding over a range of p-value cut-offs, we developed multiple PRSs for AF using GWAS summary statistics from the AFGEN Consortium. The optimal PRS was determined by the maximal AUC in a logistic regression model with AF as the outcome in all BioVU subjects. We then calculated the PRSs for all non-Hispanic White subjects in the BioVU cardiac surgery cohort and tested the association between the GPS and PoAF using a logistic regression model. To determine the potential for PRS to enhance perioperative risk prediction models of PoAF we incorporated GPS to models including genetic ancestry and predefined clinical risk factors.

**Results:** The frequency of PoAF was 40.5% (n=2,177) in the BioVU adult cardiac surgery cohort (n=5,364). Development of PoAF was significantly associated with known clinical risk factors, including history of AF (p<0.001, for all). We estimated the GPSs for 721 subjects that developed PoAF and 908 subjects that did not. The GPS was significantly associated with development of PoAF (p<0.001) and remained an independent predictor after adjustment for known risk factors and genetic ancestry (p<0.001).

**Conclusions:** We successfully developed an electronic algorithm for the identification of PoAF following cardiac surgery in a DNA databank for genetic studies. Further, AF PRS is an independent predictor of PoAF and the addition of GPSs to prediction models may improve clinical prediction of PoAF particularly in patients who are at an increased genetic risk of developing AF but who have not yet developed AF by surgery date.

## PgmNr 2693: Improved prediction of common complex diseases using family history informed genetic risk scores.

### Authors:

B.N. Wolford <sup>1</sup>; I. Surakka <sup>2</sup>; M.E. Gabrielsen <sup>3</sup>; A.H. Skogholdt <sup>3</sup>; B.M. Brumpton <sup>3,4,5</sup>; J.B. Nielsen <sup>2,3,6</sup>; S.E. Graham <sup>2</sup>; L.G. Fritsche <sup>7</sup>; H.M. Kang <sup>7</sup>; S. Lee <sup>7</sup>; K. Hveem <sup>3,8</sup>; C.J. Willer <sup>1,2,3,9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA; 2) Department of Internal Medicine: Cardiology, University of Michigan, Ann Arbor, Michigan, USA; 3) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim, Norway; 4) Medical Research Council Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom; 5) Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.; 6) Department of Epidemiology Research, Statens Serum Institute, Copenhagen, Denmark; 7) Department of Biostatistics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA.; 8) HUNT Research Centre, Department of Public Health and Nursing, Norwegian University of Science and Technology, Levanger, Norway; 9) Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA

### Introduction

Major improvements in human health and longevity could be seen if individuals at high-risk of preventable diseases were identified and treated pre-emptively. One approach is via accurate estimation of a genetic risk score (GRS) which captures polygenic disease risk with individuals in the top percentile of GRS having an odds of disease equal to or exceeding monogenic disease risk for some complex diseases (Khera et al). We hypothesize for most common diseases and in most populations, family history will improve identification of at-risk individuals in population screening compared to using GRS alone.

### Methods and Results

By combining genotype and imputation, dense phenotyping and comprehensive electronic health records, and self-reported family history in the Nord-Trøndelag Health Study (N=69,635), we demonstrate the additive relationship between family history and GRS and estimate a new metric, the Family History informed Genetic Risk Score (FHiGRS). FHiGRS provides a better predictor of disease prevalence than GRS or family history alone, for both Myocardial Infarction (MI) using a Coronary Artery Disease (CAD) GRS, and for Type 2 Diabetes (T2D) using a T2D GRS. For example, in classification of MI cases the area under the curve of the receiver operator characteristic (AUROC) for FHiGRS was 0.65 versus 0.58 for GRS alone, and for classifying T2D cases the AUROC for FHiGRS was 0.71 versus 0.61 for GRS alone.

For CAD, the odds ratio (OR) for MI in individuals in the top 5% of CAD GRS with positive family history compared to the remaining population is 2.85 (95%CI=2.48-3.27) while the OR for all individuals in the top 5% CAD GRS compared to the overall prevalence in the population is 1.98 (95%CI=1.80-2.18). Similarly, the OR for T2D in the top 5% of T2D GRS with family history (OR=4.51, 95%CI=3.70-5.50) is almost double that compared to the GRS that does not take family history into account (OR=2.42, 95%CI=2.13-2.75).

### Conclusion

Self-reported family history is easily ascertainable early in life (e.g. does not require lab measurements) and is an indicator of shared environmental factors and genetic predisposition to disease. This work suggests family history is an informative addition to GRS, and FHiGRS can be used to accurately identify individuals at greatest risk of CAD or T2D. We are evaluating broad applicability by estimating FHiGRS in UK Biobank and for additional complex disease traits.

## PgmNr 2694: Polygenic risk scores in prediction of complications and mortality in cardiometabolic diseases and common cancers.

### Authors:

N. Mars<sup>1</sup>; E. Widén<sup>1</sup>; A.S. Havulinna<sup>1,2</sup>; T.T.J. Kiiskinen<sup>1</sup>; M. Kurki<sup>1,3,4</sup>; J. Karjalainen<sup>1,5,6</sup>; P. Palta<sup>1,7</sup>; M. Daly<sup>1,5</sup>; A. Palotie<sup>1,4,5</sup>; S. Ripatti<sup>1,5,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland, FIMM, HiLIFE, University of Helsinki, Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Program in Medical and Population Genetics and Genetic Analysis Platform, Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 4) Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Analytic and Translational Genetics Unit, Department of Medicine, and the Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts; 5) Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 7) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia; 8) Department of Public Health, Clinicum, University of Helsinki, Helsinki, Finland

---

Genome-wide polygenic risk scores (PRS) are strongly associated with first cardiometabolic and cancer events, but evidence about their impact on predicting subsequent disease events and disease complications is limited.

Using Finnish nationwide biobank data (FinnGen,  $n = 135,300$ ) with genome-wide genotyping and up to 46 years of prospective follow-up within nationwide registries, we studied whether disease-specific PRSs are associated with recurrent disease events and complications for coronary heart disease (CHD; 19,767 cases), atrial fibrillation (AF; 12,809), type 2 diabetes (T2D; 17,519), breast cancer (3,904), and prostate cancer (3,263).

Compared to individuals without CHD, those with only one CHD event had on average 0.29 standard deviations (SD) higher CHD PRS, and those with two or more CHD events, 0.51SD higher PRS ( $p = 2.25 \times 10^{-40}$  for one vs two or more events). The CHD PRS was associated with the risk of CHD in both T2D patients (HR 1.26, 95% CI 1.23–1.29,  $p = 1.34 \times 10^{-67}$ ) and in T1D patients (1.25, 1.16–1.33,  $p = 2.30 \times 10^{-10}$ ). Compared to individuals without T2D, T2D patients without complications had 0.36SD higher T2D PRS, and those with one or more complications 0.53SD higher T2D PRS ( $p = 4.17 \times 10^{-21}$ ). Compared to individuals without AF, those without heart failure had 0.48SD higher AF PRS, and those with heart failure following or at diagnosis of AF had 0.56SD higher AF PRS ( $p = 0.001$ ).

The CHD PRS was also associated with CHD deaths (HR 1.15 per SD increment, 95% CI 1.12–1.18,  $p = 2.20 \times 10^{-24}$ ), and a PRS combining both the CHD PRS and T2D PRS was strongly associated with CHD death (HR 1.23, 1.19–1.26,  $p = 9.79 \times 10^{-45}$ ). In a smaller subset of 10,258 men with cancer death information available, the prostate cancer PRS was associated with death due to cancer (HR 1.12, 1.03–1.22,  $p = 0.007$ ).

Our results show that the polygenic risk scores derived for first events of diseases are also associated with the risk for recurrent disease events and death in CHD, for diabetes complications in T2D, and for cancer-related deaths. These results suggest that the potential utility of polygenic risk scores extends beyond primary prevention and may be beneficial also for disease prognosis, secondary prevention, and disease management in common complex diseases.

## PgmNr 2695: Exome sequencing of Takotsubo cardiomyopathy identifies novel and known rare variants in candidate genes.

### Authors:

L.C. Kwee <sup>1</sup>; J.W. Abdulrahim <sup>1</sup>; I.C. Siegler <sup>2</sup>; W.E. Kraus <sup>3</sup>; R.B. Williams <sup>2</sup>; S.K. Smith <sup>4</sup>; S.H. Shah <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Duke Molecular Physiology Institute, Duke University, Durham, NC; 2) Behavioral Medicine Research Center, Duke University, Durham, NC; 3) Division of Cardiology, Department of Medicine, Duke University, Durham, NC; 4) Department of Anesthesiology, Duke University, Durham, NC

**Introduction:** Takotsubo cardiomyopathy (TC) is a severe but usually transient cardiovascular syndrome characterized by left ventricular (LV) dysfunction and apical LV ballooning. It predominantly occurs in older women after a physical or emotional stress trigger, but its biological etiology remains unclear. Proposed mechanisms include catecholamine-mediated toxicity, vascular dysfunction due to increased susceptibility to endothelin, and dysregulated myocardial calcium regulation. We hypothesized that identifying rare variants in genes related to these pathways may help clarify the biological underpinnings of susceptibility to TC.

**Methods:** The study population consisted of individuals from the CATHGEN biorepository with a diagnosis of TC based on the following parallel approaches: ICD-9 search; echocardiogram reports of hypocontractile, akinetic, aneurysmal, or dyskinetic LV apex in patients without coronary artery disease; and health record keyword search (Takotsubo, broken heart, stress cardiomyopathy, apical ballooning). All cases were adjudicated by deep chart review using clinical diagnostic criteria for TC. Whole exome sequencing was performed on all cases and 98 candidate genes involved in endothelin signaling, calcium regulation, and familial cardiomyopathy were selected.

**Results:** Of 340 potential TC patients, eight had a diagnosis substantiated after chart review (TC set), leaving 8583 non-Takotsubo CATHGEN subjects for comparison. We identified 36,258 candidate gene variants, 17 of which were present only in the TC set. Two rare variants (gnomAD maf<0.001%) were identified in genes not previously associated with cardiomyopathy: rs771740107 in *EDNRA* (endothelin receptor) and rs746335414 in *CACNG1* (calcium channel subunit). The remaining TC set-specific variants were in known familial cardiomyopathy genes: six novel variants in *PRDM16*, *MYH6*, *TTN*, *EYA4*, and *HFE*, and known rare variants in *PRDM16*, *TTN*, *EYA4*, *CTNNA3*, *LDB3*, *ABCC9*, *MYH6*, and *ACTC1*.

**Conclusion:** Using exome sequencing, we have identified both novel and known variants present only in TC patients in CATHGEN. Identification of variants in known cardiomyopathy genes suggests that TC may share pathophysiological elements with more common cardiomyopathies, while the TC-specific variants in endothelin and calcium regulation genes indicate these pathways may be relevant for future study into TC susceptibility.

## PgmNr 2696: Rare loss-of-function variant analysis in lone atrial fibrillation.

### Authors:

J. Lazarte <sup>1,2</sup>; J. Wang <sup>1</sup>; J.F. Robinson <sup>1</sup>; J.S. Dron <sup>1</sup>; A.D. McIntyre <sup>1</sup>; H. Cao <sup>1</sup>; Z. Laksman <sup>3</sup>; R.A. Hegele <sup>1,2</sup>; J.D. Roberts <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Robarts Research Institute, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 2) Department of Medicine, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada; 3) Department of Medicine, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, Canada

---

Atrial fibrillation (AF) is a complex heritable disorder that substantially increases the risk for stroke, sudden death and all-cause mortality. Lone AF is a subset defined by early diagnosis (less than 60 years) and with no evidence of cardiopulmonary disease. Genome-wide association studies have associated loci near genes for structural integrity and function of muscles with AF susceptibility. In addition, some studies have identified rare loss-of-function (LoF) variants in dilated cardiomyopathy (DCM) genes in lone AF. The increased burden of these variants in lone AF patients may indicate a shared predisposition between AF and DCM.

We performed a retrospective study on 252 adult patients and a control group of 503 healthy European subjects. Exome sequencing and microarray assays were performed at the McGill Genome Center while control's exome data was retrieved from 1KG Consortium. Principal component analysis was performed to define a group with common ancestry using SVS software. Variants within candidate DCM genes were filtered by minor allele frequency of  $\leq 0.1\%$ , a Combined Annotation Dependent Depletion (CADD) Phred score of  $>20$  and a LoF variant type (frameshift, splice acceptor, splice donor and stop gain). All variants were confirmed with Sanger sequencing. Cascade screening was attempted on all cases positive for variants.

The AF cohort, an ethnically homogenous composite of 181 lone AF cases, had 6 heterozygous LoF variants, including 4 novel variants, in 6 individuals, while the controls had 4 in 503 individuals. Half of variants in the cases were in the *TTN* gene (3/6). The prevalence of rare LoF variants in the lone AF cohort was 3.3% while in the control cohort 0.79%, the odds ratio of rare LoF variants found in lone AF patients compared to controls is 4.28 (95%CI: 1.19-15.33, P-value: 2.5E-02). Cascade screening in one family demonstrated co-segregation of one of the rare *TTN* variants and AF.

Patients with lone AF carry a significant burden of heterozygous LoF variants in DCM genes. Our results are in agreement with a previous study identifying *TTN* variants in lone AF patients. Further, we identified novel variants in different DCM genes and confirmed the co-segregation of a *TTN* variant and AF in one family. Our findings may indicate a common genetic predisposition between lone AF and DCM, which may help elucidate novel pathophysiologic mechanisms, and improve diagnosis and treatment.



## PgmNr 2697: Targeted sequencing of linkage region in Dominican families implicates *PRIMA1* and the *SPATA7-PTPN21-ZC3H14-EML5-TTC8* locus in carotid-intima media thickness and atherosclerotic events.

### Authors:

L. Wang<sup>1,2</sup>; N. Dueker<sup>1</sup>; A. Beecham<sup>1</sup>; S.H. Blanton<sup>1,2</sup>; R.L. Sacco<sup>1,2,3,4</sup>; T. Rundek<sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, Florida.; 2) John T. McDonald Department of Human Genetics, Miller School of Medicine, Miami, Florida; 3) Department of Neurology, Public Health Sciences, Miami, Florida; 4) Evelyn F. McKnight Brain Institute, Miller School of Medicine, University of Miami, Miami, Florida

Carotid intima-media thickness (cIMT) is a subclinical marker for atherosclerosis. Previously, we identified a quantitative trait locus (QTL) for total cIMT on chromosome 14q. Using a common variants (CVs)-based approach via high throughput genotyping, we have found evidence for association in *PRiMA1*, *FOXN3* and *CCDC88C*. Herein, we further evaluated the genetic contribution of the QTL to cIMT by surveying both CVs and rare variants (RVs) via resequencing the linkage region. We sequenced all exons within the QTL and genomic regions of *PRiMA1*, *FOXN3* and *CCDC88C* using next generation sequencing coupled with targeted sequence enrichment. Dominican families (N=7) with evidence for linkage to the QTL (family specific LOD>0.1) were sequenced. Unrelated Dominicans from the Northern Manhattan Study (NOMAS, N=561) were used for validation using HumanCoreExome with custom content. Single-variant-based analysis was performed for CVs and gene-based analysis were performed for RVs, adjusting for significant covariates identified using a polygenic screen. Top cIMT-associated CVs were further tested for association with clinical atherosclerotic events, including myocardial infarction, ischemic stroke or vascular death, in NOMAS. Total cIMT was calculated as a composite measure of the near and the far wall IMT of all carotid sites from both sides. The family size ranged from 11 to 25, with 116 individuals in total from the 7 families. The strongest evidence for association with CVs was found in *PRiMA1* ( $p=8.2 \times 10^{-5}$  in families,  $p=0.01$  in NOMAS at rs12587586), and in the five-gene cluster *SPATA7-PTPN21-ZC3H14-EML5-TTC8* locus ( $p=1.3 \times 10^{-4}$  in families,  $p=0.01$  in NOMAS at rs2274736). A top marker in *PRiMA1* (rs7152362) was associated with fewer atherosclerotic events (OR=0.67;  $p=0.02$  in NOMAS) and smaller cIMT ( $\beta=-0.58$ ,  $p=2.8 \times 10^{-4}$  in Family,  $\beta=-0.01$ ,  $p=0.047$  in NOMAS). No evidence for association with RVs was found in *PRiMA1*. Within the five-gene cluster, evidence for association was found for exonic RVs ( $p=0.02$  in families,  $p=0.28$  in NOMAS), which was enriched among RVs with higher functional potentials ( $p=0.05$  in NOMAS for RVs in the top functional tertile). In summary, targeted resequencing provided validation and novel insights into the genetic architecture of cIMT in Dominicans. We found that RVs with higher functional potentials have stronger effects for cIMT. Furthermore, our data support the clinical relevance of subclinical atherosclerosis-associated CVs.

## PgmNr 2698: Excess of population-specific rare functional variants in *ANGPTL3* increase hypertriglyceridemia and coronary artery disease risk in Punjabi Sikhs.

### Authors:

D.K. Sanghera <sup>1</sup>; C. Bejar <sup>1</sup>; G.S. Wander <sup>2</sup>; S. Ralhan <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Pediatrics, Sect Gen, Univ Oklahoma HSC, Oklahoma City, Oklahoma.; 2) Dayanand Medical College and Hero DMC Heart Institute, Ludhiana, Punjab, India

---

Elevated circulating triglyceride levels are well-established risk factor for cardiovascular disease and a principal cause of mortality in individuals with type 2 diabetes. Despite high heritability (50-80%) of triglycerides, previous genome-wide association studies (GWAS) have only been able to account for a fraction of the heritability (<10%) in genes involved in triglyceride (TG) and lipid metabolism. In contrast to the existing common association signals with blood lipid phenotypes, there is extreme paucity of data on the spectrum of rare and private variants in diverse ethnic groups especially in populations of Asian Indians. Most rare variants are of recent origin, they may vary between ethnic groups.

We sequenced 940 individuals [572 cases with high serum TG (>150mg/dl) and 368 controls with low TG (<100mg/dl)] and performed replication in additional 2,239 samples from the same Punjabi population. All 3,179 participants were part of the Asian Indians Diabetic Heart Study. A known common polymorphism in *ANGPTL3* (M259T) present in 10% of African Americans and associated with lower plasma levels of TG in Africans and Europeans, was not observed in this population. We observed three rare missense variants, earlier found in Caucasians (Asn296Ser, Ala364Val, and Phe295Leu), earlier found in Caucasians, were associated with high TG levels (ranging between 290 mg/dl to 313 mg/dl) in our sample. Additionally, our data revealed several new variants at 5'UTR, 3'UTR, missense and loss-of-function variants, which were not found in any public database. Large proportion of these variants were present in people with >75%TG and some rare variants were only present with subjects with lower TG levels (<25%).

In summary, our study found notable differences in the profiles of rare/private variants and their association with lipid traits between Panjabi population from North India and Europeans. Results of our currently ongoing validation study of the most significant rare variants for their association with TG, other lipid traits, and cardiovascular disease risk in additional samples of South Asians, Europeans and Mexicans populations will be presented.

This study was supported by NIH grants -R01DK082766 (NIDDK) and NOT-HG-11-009 (NHGRI), R218 RS&G Service, U.S. Federal Government contract # HHSN268201100037C (NHLBI).

## PgmNr 2699: Comprehensive structural variation map for 19,652 humans and impact on translation for disease phenotypes utilizing long reads.

### Authors:

F.J. Sedlazeck<sup>1</sup>; B. Yu<sup>2</sup>; A.J. Mansfield<sup>1</sup>; H. Chen<sup>2</sup>; O. Krasheninina<sup>1</sup>; A. Tin<sup>3</sup>; Q. Qi<sup>4</sup>; S. Zarate<sup>5</sup>; J.L. Traynelis<sup>1</sup>; V. Menon<sup>1</sup>; J. Hu<sup>1</sup>; H. Doddapaneni<sup>1</sup>; G.A. Metcalf<sup>1</sup>; J. Coresh<sup>3</sup>; R.C. Kaplan<sup>4,6</sup>; D.M. Muzny<sup>1</sup>; G. Jun<sup>2</sup>; R.A. Gibbs<sup>1</sup>; W.J. Salerno<sup>1</sup>; E. Boerwinkle<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) School of Public Health, University of Texas Health Science Center at Houston, Houston, TX; 3) Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 4) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 5) DNAnexus, Mountain View, CA; 6) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

---

Although structural variations (SV) have been shown to impact expression and play a key role in certain diseases we still lack a large-scale study to include SVs and investigate their impact on the protein level. Here we have analyzed 19,652 whole genomes (30x+) human samples across multiple ethnicities to capture the so far hidden variability within and among populations across multiple ethnicities. To achieve this we have implemented Parliament2 our novel population level SV calling pipeline. We have discovered and genotyped 304,533 SVs ranging from small deletions to large rearrangements and amplifications across GRCh38. We further analyzed the ancestral alleles for these SVs with respect to chimps and gorillas.

We optimized the use of long reads for population studies to validate and extend the SV catalog. We developed a novel sampling framework (SVCollector) that operates based on the variations called in the short read data to maximize the validation efficiency. Using this framework we sequence 18 PacBio genomes with matching RNA-Seq data to validate and obtain further insights on these samples. Using the 18 PacBio genomes we were able to validate a large proportion of the SVs from the short read cohort. This analysis resulted in haplotype resolved SNP and SVs per sample across the genome. We are currently in the process to genotype insertions and initially missed SVs into the association cohort study using Paragraph. Paragraph is a novel graph-based genome method to genotype SVs in short reads even though these SVs were initially missed.

Overall our study highlights multiple ethnicity-specific SVs, which alter genes and impact their protein levels and structure. In total, we identified 2,336 protein-coding genes impacted by common SVs (MAF >0.05) and 234 genes that are affected by multiple SVs specific to only one ethnic group (MAF >0.05). Given the new set of SVs, we have analyzed 4,156 proteins in a subset of 4,021 multi-ethnic samples. We also identified 58 significant SV-protein associations that comprise 33 cis- and 25 trans-acting relationships, explaining a median of 7-9% protein variations in African- and European-Americans. Our findings elucidate a more complete mapping from a comprehensive genomic architecture to regulatory and translational pathways to endophenotypes and ultimately to disease endpoints. Furthermore, this study is a role model on how to efficiently integrate long read

technologies for haplotype resolved variant detection.

# PgmNr 2700: Novel loci linked to impaired ejection fraction and risk of cardiomyopathy among childhood cancer survivors of African-descent: A report from the St. Jude Lifetime Cohort.

## Authors:

Y. Sapkota<sup>1</sup>; D.A. Mulrooney<sup>1,2</sup>; G.T. Armstrong<sup>1</sup>; C.L. Wilson<sup>1</sup>; Z. Wang<sup>1,3</sup>; J. Zhang<sup>3</sup>; M.M. Hudson<sup>1</sup>; L.L. Robison<sup>1</sup>; Y. Yasui<sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Epidemiology and Cancer Control, St Jude Children's Research Hospital, Memphis, Tennessee.; 2) Oncology, St Jude Children's Research Hospital, Memphis, Tennessee.; 3) Computational biology, St Jude Children's Research Hospital, Memphis, Tennessee.

**Background:** Survivors of childhood cancer are at increased risk of treatment-related cardiomyopathy. To our knowledge, no study has assessed genetic risk of cardiomyopathy in childhood cancer survivors of African-descent.

**Methods:** We utilized whole-genome sequencing data (36.8-fold) of 301 childhood cancer survivors of African-descent from the St. Jude Lifetime Cohort (SJLIFE) study and examined associations between genetic variants and ejection fraction (EF) as a continuous outcome. Common (MAF>0.05) and rare/low-frequency (MAF≤0.05) variants were assessed separately using single-variant and burden/SKAT tests, respectively, adjusting for demographic characteristics and childhood cancer treatment exposures.

**Results:** We identified two novel loci for reduced EF at 1p13.2 (rs6689879: MAF=0.16; per-allele reduction=4.2%;  $P=1.7 \times 10^{-9}$ ) and 15q25.3 (rs9788776: MAF=0.08; per-allele reduction=5.6%;  $P=3.3 \times 10^{-8}$ ). None of the burden/SKAT results was statistically significant. Among 246 survivors of African-descent exposed to either anthracyclines or heart radiation, both loci retained genome-wide significance: rs6689879 (per-allele reduction=4.3%;  $P=3.0 \times 10^{-8}$ ) and rs9788776 (per-allele reduction=6.0%;  $P=3.4 \times 10^{-8}$ ). The association between the 1p13.2 locus and EF was replicated among 1645 survivors of European-descent from the SJLIFE study who were also exposed to anthracyclines or heart radiation, with an attenuated effect size (rs6689879: MAF=0.30; per-allele reduction=0.5%;  $P=0.045$ ); the 15q25.3 locus is monomorphic in Europeans. Both loci were significantly associated with an increased risk of clinically-diagnosed cardiomyopathy: Common Terminology Criteria for Adverse Events (CTCAE) grade 2+ [rs6689879: OR=1.83;  $P=6.2 \times 10^{-4}$  and rs9788776: OR=4.44;  $P=4.5 \times 10^{-3}$ ] and CTCAE grade 3+ [rs6689879: OR=3.76;  $P=1.8 \times 10^{-14}$  and rs9788776: OR=3.24;  $P=0.039$ ]. rs6689879 maps to *MAGI3* which encodes a scaffolding protein that acts at cell-cell junctions, regulating cellular and signaling processes. rs9788776 is located ~61.7 kb away from *AGBL1* which encodes a glutamate decarboxylase that catalyzes the deglutamylation of polyglutamylated proteins. The *AGBL1* locus has been shown to be associated with reduced activated partial thromboplastin time.

**Conclusions:** This study among childhood cancer survivors of African-descent identified two novel loci for reduced EF and increased risk of cardiomyopathy. Further research is needed to elucidate the pathophysiology underlying these associations.

## PgmNr 2701: Assessing the transferability of genetic associations with aortic stenosis to non-European ethnicities.

### Authors:

M. Hoekstra<sup>1,2</sup>; H.Y. Chen<sup>1,2</sup>; L. Dufresne<sup>2</sup>; G. Thanassoulis<sup>1,2</sup>; J.C. Engert<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McGill University, Montréal, QC; 2) McGill University Health Centre and Research Institute, Montréal, QC

---

Aortic stenosis (AS) is a debilitating heart valve disease with no known medical treatments. While genetic association studies in Europeans have highlighted novel therapeutic targets for AS, it is unknown whether these findings are transferable to non-European ethnicities. We have derived a genetic risk score (GRS) for AS from risk variants previously discovered in European populations. We determined the extent to which the unweighted GRS predicted AS in 55,192 Europeans, 1,917 African-Americans and 3,482 Latin-Americans from the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. In addition, we assessed each variant independently for associations with AS. To investigate the influence of ancestry on these associations, we estimated global ancestry proportions in each admixed individual. In total, 6 variants from 5 different loci were included in the GRS. As expected, the GRS was a strong predictor of AS in Europeans (OR per risk allele [95% CI], 1.10 [1.08, 1.13];  $P=4.4e-17$ ). While the GRS demonstrated a consistent effect in African-Americans (OR = 1.14 [0.97, 2.34]), it was not significantly associated with AS in either admixed population. These results remained virtually unchanged after adjustment for non-European ancestry. Of the variants included in the GRS, only rs10455872 in *LPA* and rs2069832 in *IL6* were associated with AS in African-Americans, with rs10455872 demonstrating the largest effect seen to date (OR per risk allele, 4.52 [1.50, 13.6];  $P=0.0074$ ). Consistent with European populations, each risk allele of rs10455872 was associated with increased levels of lipoprotein(a) in individuals with mixed European-African ancestry from the UK Biobank ( $P=3.1e-20$ ). In both GERA African-Americans and Latin-Americans, carriers of this variant had significantly more European ancestry ( $P=1.6e-07$  and  $P=2.2e-07$ , respectively). Notably, none of the risk variants investigated were significantly associated with AS in Latin-Americans. Our results suggest that genetic associations with AS in European populations may differ in admixed populations, hindering the predictive power of a GRS. Additional work is needed to determine if non-European ethnicities possess unique risk loci for this disease.

## PgmNr 2702: Miransertib (ARQ 092) prevents the formation of *PIK3CA*<sup>H1047R</sup>-driven vascular malformations in mice.

### Authors:

P. Kobińska<sup>1</sup>; S.D. Castillo<sup>1</sup>; H. Sabata<sup>1</sup>; Y. Yu<sup>2</sup>; B. Schwartz<sup>2</sup>; M. Graupera<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vascular Signalling Laboratory, Institut d'Investigacio Biomedica de Bellvitge, Hospitalet de Llobregat, Barcelona, Spain; 2) ArQule, Inc, Burlington, MA 01803, USA

Venous malformations are congenital lesions that consist of abnormal dilated venous channels with scarce mural cell coverage, they can be of different sizes and be present in any tissue, such as subcutaneous tissue or internal organs. Venous malformations are painful and disfiguring, many leading to bleeding, obstruction of organs and intravascular coagulopathy with an overall incidence of 1:5000. Most venous malformations are caused by either somatic mutations in the *TEK* or the *PIK3CA* genes in endothelial cells, both types of mutations leading to an increased PI3K signalling. PI3K activation in endothelial cells triggers the activation of the AKT serine/threonine kinase promoting a plethora of cellular effects, including increased cell proliferation and loss of mural cell coverage, which contribute to the pathogenesis of venous malformations. In a previous study, we have created a genetically modified mouse model that displays mutant *PIK3CA*<sup>H1047R</sup>-driven congenital vascular malformations with features pathognomonic for human VMs.

Miransertib (ARQ 092) is a potent and selective allosteric pan-AKT inhibitor that potently inhibits AKT1, 2, and 3 isoforms. Biochemical and cellular studies showed that miransertib suppresses AKT activity by inhibiting membrane-bound active form of AKT and preventing activation of inactive form of AKT. Miransertib is currently in clinical studies in patients with PIK3CA-Related Overgrowth Spectrum and Proteus syndrome. To assess the efficacy of miransertib in *PIK3CA*-driven vascular malformations, we treated postnatal mice at the same time of *PIK3CA*<sup>H1047R</sup> induction in the endothelium and analysed the retinal vasculature. By measuring the area of isolectin-B4 staining (IB4 labels the plasma membrane of endothelial cells) and the number of Erg-positive cells (marker for endothelial cells), we showed that miransertib treatment reduced the vascular density and the number of endothelial cells driven by *PIK3CA* mutation. Mechanistically, miransertib suppressed the hyperproliferation of endothelial cells driven by *PIK3CA*<sup>H1047R</sup> mutation.

In conclusion, treatment of miransertib prevents *PIK3CA*<sup>H1047R</sup>-driven vascular malformations in postnatal mouse retinas. The results provide a rationale for treating vascular malformation driven by activating mutation of *PIK3CA* in patients.



## PgmNr 2703: Variation of alpha-1 antitrypsin levels with pulmonary and cardiometabolic traits in the multi-ethnic study of atherosclerosis.

### Authors:

Y. Ma <sup>1</sup>; C.J. Williams <sup>2</sup>; S.J. Bielinski <sup>3</sup>; J.I. Rotter <sup>4</sup>; M. Brantly <sup>2</sup>; S.S. Rich <sup>1</sup>; C.P. Hersh <sup>5</sup>; R.G. Barr <sup>6</sup>; A. Manichaikul <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA.; 2) Department of Medicine, University of Florida, Gainesville, FL, USA.; 3) Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA.; 4) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA.; 5) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA.; 6) Department of Medicine, College of Physicians and Surgeons and Department of Epidemiology, Mailman School of Public Health, Columbia University, NY, USA.

### Introduction / Objectives

Alpha-1 Antitrypsin Deficiency (AATD), most commonly caused by the missense variant rs28929474 in the *SERPINA1* encoding AAT, is the most well-documented genetic cause of emphysema and chronic obstructive pulmonary disease (COPD). Research in diverse populations on the relationship of AAT levels with pulmonary and cardiometabolic traits is limited. Therefore, we sought to fill this gap using the Multi-Ethnic Study of Atherosclerosis (MESA).

### Methods

MESA recruited participants ages 45-84 years old who were free of clinical cardiovascular disease. The Proteinase inhibitor (PI) type and AAT levels were measured in plasma by isoelectric for 2,255 African, 1,865 Hispanic, and 500 non-Hispanic White. AAT levels were compared across race/ethnicity and stratified by PI type, with 93% PiMM (no altered AAT), 5% PiMS (one altered *SERPINA1* allele, minimal risk for AATD), and 1% PiMZ (heterozygote carrier for AATD). We examined the relationship of AAT levels with cardiometabolic and pulmonary traits measured by cardiac CT scans and spirometry. Analyses were performed by linear (or logistic) mixed models, adjusted for age, sex, study site, race/ethnicity, principal components of ancestry, Pi type, and other confounders (e.g., smoking history). Analyses were (1) pooled across all participants and (2) stratified by race/ethnicity and smoking history (current/former/never).

### Results

African Americans had lower AAT levels than both Hispanics ( $\Delta = 3.6$  mg/dL,  $P = 0.003$ ) and whites ( $\Delta = 5.4$  mg/dL,  $P < 0.001$ ). Further analysis showed that the proportion of African ancestry was related to AAT levels among Hispanics ( $\beta = -9.622$ ,  $P = 0.006$ ). Systolic blood pressure, BMI, hypertension, presence of calcium, and Agatston calcium score were positively associated with AAT levels ( $P < 0.01$ ), and fasting glucose showed an inverse relationship with AAT levels ( $P = 0.005$ ), in pooled analysis. Lower AAT levels were associated with higher percent emphysema ( $P = 0.027$ ) and percent emphysema in the lower lobes ( $P = 0.006$ ) among current smokers only.

### Conclusions



AAT levels vary across race/ethnicity and were significantly associated with some cardiovascular risk factors, while a relationship with percent emphysema was observed in current smokers only. These results reinforce the need to evaluate protein biomarkers across race/ethnic groups and may help inform future efforts to develop proteins encoded by rare or Mendelian disease variants as potential disease biomarkers.

## PgmNr 2704: HeartCare: Improving patient care through comprehensive cardiovascular genetic testing.

### Authors:

D. Murdock<sup>1</sup>; E. Venner<sup>1</sup>; D. Muzny<sup>1</sup>; M. Murugan<sup>1</sup>; X. Jia<sup>2</sup>; J. Hu<sup>1</sup>; K. Walker<sup>1</sup>; G. Metcalf<sup>1</sup>; C. Kovar<sup>1</sup>; M.C. Gingras<sup>1</sup>; V. Korchina<sup>1</sup>; C. Howard<sup>3</sup>; K. Collins<sup>4</sup>; D. Riconda<sup>5</sup>; S. Pereira<sup>6</sup>; Z. Huda<sup>2</sup>; A. Nafegar<sup>2</sup>; P. Marino<sup>2</sup>; A. Civitello<sup>2</sup>; T. Hadley<sup>2</sup>; L. Leiber<sup>3</sup>; A. McGuire<sup>6</sup>; X. Wehrens<sup>2</sup>; A. Balasubramanyam<sup>2</sup>; C. Amos<sup>2</sup>; E. Boerwinkle<sup>7</sup>; D. Hostin<sup>4</sup>; C. Ballantyne<sup>2</sup>; R. Gibbs<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Medicine, Baylor College of Medicine, Houston, TX; 3) Information Technology, Baylor College of Medicine, Houston, TX; 4) Precision Medicine Alliance, CommonSpirit Health, Englewood, CO; 5) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 6) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 7) The University of Texas Health Science Center at Houston, TX

---

Cardiovascular disease (CVD) is the leading cause of mortality in the United States, leading to 1 in every 4 deaths. The role of inherited susceptibility to CVD is well established, from rare monogenic disorders to pharmacogenetics and more recently studied polygenic traits. Thus, precision medicine efforts focusing on CVD genetics have the potential to advance adult clinical genetics in a similar fashion that exome sequencing did for pediatrics.

Within the Baylor College of Medicine Human Genome Sequencing Centers-Clinical Laboratory (HGSC-CL), we developed a 158-gene NGS panel, ('HeartCare'), designed to assay genetic risk associated with CVD in four distinct categories: 1) High-penetrance, monogenic conditions including cardiomyopathies, aortopathies, arrhythmias, and dyslipidemias, 2) coronary artery disease polygenic risk score (PRS), 3) *LPA* common variants associated with increased CVD risk, and 4) pharmacogenetic variants contributing to simvastatin-induced myopathy and warfarin metabolism.

For a pilot of 5,000 participants, individuals are enrolled at multiple clinical sites, samples analyzed in a CAP/CLIA-certified laboratory and results returned to the ordering physician and uploaded to the electronic medical record. To date, 126 individuals have been enrolled, primarily from a cardiology dyslipidemia clinic. Of those with results, approximately 7% had a pathogenic/likely pathogenic variant in the high-penetrance, monogenic category, all in the *LDLR* gene consistent with a diagnosis of familial hypercholesterolemia. Nearly 9% of patients were in the highest CVD risk group (top 5%) according to the PRS while 30% carried common *LPA* variants associated with increased Lp(a) lipoprotein and CVD risk in certain populations. Lastly, approximately half harbored variants that increase their risk of simvastatin-induced myopathy and/or affect warfarin metabolism. Management changes and additional lab follow-up have already resulted from these findings with genetic counseling instituted to facilitate familial cascade testing and explore behavior changes. Our results to date support the use of genetic information in routine cardiovascular health management.

## PgmNr 2705: MEK inhibitor treatment in a preterm newborn with Noonan syndrome and severe obstructive hypertrophic cardiomyopathy.

### Authors:

A. Mussa <sup>1</sup>; D. Carli <sup>1</sup>; E. Giorgio <sup>2</sup>; A.M. Villar <sup>3</sup>; C. Carbonara <sup>4</sup>; P. Galletto <sup>4</sup>; G. Botta <sup>5</sup>; G. Andelfinger <sup>6</sup>; M. Tartaglia <sup>7</sup>; E. Medico <sup>8,9</sup>; A. Brusco <sup>2,10</sup>; G.B. Ferrero <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Public Health and Pediatric Sciences, University of Torino, Italy, Torino, Italy; 2) Department of Medical Sciences, University of Torino, Italy; 3) Pediatric Cardiology, Regina Margherita Children Hospital, Città della Salute e della Scienza, Torino, Italy; 4) Neonatal Intensive Care Unit, S.Anna Hospital, Città della Salute e della Scienza, Torino, Italy; 5) Department of Pathology, S.Anna Hospital, Città della Salute e della Scienza, Torino, Italy; 6) Cardiovascular Genetics, CHU Sainte Justine, Department of Pediatrics, Université de Montréal, 3175, Chemin Côte Sainte Catherine, Montréal, QC, Canada; 7) Genetics and Rare Diseases Research Division, Ospedale Pediatrico Bambino Gesù, IRCCS, Rome, Italy; 8) Laboratory of Oncogenomics, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy; 9) Department of Oncology, University of Torino, Italy; 10) Medical Genetics Unit, Città della Salute e della Scienza, Torino, Italy

A female infant was born at 33+6 weeks of gestation complicated by polyhydramnios and fetal hypertrophic cardiomyopathy (HCM). At birth, excessive nuchal skinfolds and facial features allowed a clinical diagnosis of Noonan syndrome (NS). She required resuscitation and mechanical ventilation for 10 days. At the 2<sup>nd</sup> day of life, she developed a grade II cerebral ventricular haemorrhage. Echocardiography confirmed severe obstructive HCM and dysplastic pulmonary valve. She was found to carry a heterozygous *de novo* p.Ser257Leu pathogenic variant in *RAF1*, typically associated with severe NS phenotype, progressive HCM and pulmonary hypertension. In the subsequent 40 days she developed severe congestive heart failure (CHF) that, in spite of furosemide administration, required mechanical ventilation and amine administration. With no treatment options for the near-terminal CHF, Trametinib – a selective reversible inhibitor of MEK1/2 activity approved for selective treatment of cancers with RAS/MAPK pathway activation – was introduced (0.022 mg/kg/day, day 0). A prompt improvement in clinical condition was observed, allowing to withdraw inotropes at day 5 and mechanical ventilation at day 12; nt-pro-BNP decreased from 30,805 pg/ml (pre-treatment) to 2,355 (day 14), consistent with HCM stabilization. During the next month, treatment was continued at the same dosage with drug level monitoring without side effects. Clinical conditions allowed a progressive reduction of non-invasive respiratory support. Cardiac ultrasound monitoring showed a tendency to HCM improvement but revealed the sudden appearance (day 23) of pulmonary artery dilation (18 mm). At 3 months of age, she underwent ventriculoperitoneal shunt device placement for obstructive posthaemorrhagic hydrocephalus and showed, after surgery, a rapidly worsening of CHF needing mechanical ventilation and maximal CHF supportive therapy, with an increase in nt-pro-BNP (47,000 pg/ml). The infant died at 3 months and a half of age (day 58) from untreatable CHF. Autopsy showed an obstructive HCM with obstructed ventricles, severe dilation of pulmonary artery trunk and branches with complete disarray of both gross and fine pulmonary vascular anatomy. Global RNA sequencing of white blood cells before and during treatment was performed to define transcriptional

effects of MEK inhibition. The latter might represent a promising treatment option for HCM in the RASopathies.

## PgmNr 2706: Comparing polygenic risk scores between the healthy elderly and the UK Biobank.

### Authors:

M. Riaz <sup>1</sup>; M. Pinese <sup>2</sup>; R. Woods <sup>1</sup>; I. Winship <sup>3</sup>; P. James <sup>3</sup>; R. Shah <sup>4</sup>; D. Thomas <sup>2</sup>; A. Murray <sup>5</sup>; J. McNeil <sup>1</sup>; P. Lacaze <sup>1</sup>; ASPREE Investigator Group, ASPREE Healthy Ageing Biobank

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology and Preventive Medicine, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia; 2) Garvan Institute of Medical Research, Sydney, Australia; 3) Genetic Medicine and Family Cancer Clinic, Royal Melbourne Hospital, Melbourne, Australia; Department of Medicine, University of Melbourne, Melbourne, Australia; 4) Department of Family Medicine and Rush Alzheimer's Disease Center, 2468 Rush University Medical Center, Chicago, IL, USA; 5) Berman Center for Outcomes and Clinical Research, Hennepin Healthcare Research Institute Minneapolis, MN, USA.

**Background:** ASPirin in Reducing Events in the Elderly (ASPREE) is a primary prevention trial of daily low-dose aspirin in 19,114 healthy elderly individuals. At enrolment (median age 74 years), participants did not have a history of cardiovascular disease (CVD), life-threatening illness or evidence of dementia. Primary trial outcomes were published in three NEJM papers in 2018 (McNeil et al), and include CVD events, cancers, dementia and death over median 4.7 years follow-up. ASPREE is now extended to an observational study of healthy ageing, with longitudinal clinical outcome and deep phenotype data, comparable to the UK Biobank. ASPREE has genotyped 15,000 participants (Axiom PMDA) and we hypothesize depletion of common and rare damaging alleles in the healthy elderly, and enrichment of protective alleles, compared with the UK Biobank.

**Aim:** To assess whether the distribution of polygenic risk scores (PRS) varies between the healthy elderly and the UK Biobank, across a range of disease and clinical phenotypes, and whether the depletion of damaging common alleles contributes to healthy ageing.

**Methods:** GWAS-based QC using Plink v2 on both datasets, ASPREE has 15000 samples and UKBB ~500,000 samples (UKBB Project: 47061), SNPs and samples passing 98% genotyping rates, imputed variant R<sup>2</sup> >0.3, Hardy-Weinberg equilibrium, LD threshold of r<sup>2</sup> <0.2 across 500 kb, minor allele frequency (MAF) <0.05, unrelated individuals and European ancestry. Plink v2 to undertake the PRS analysis on imputed genotype data from UK-Biobank and ASPREE using summary GWAS data for each clinical outcome as reference data PRS distribution between the two populations compared using Plink. Disease outcomes include CVD, cancers (breast, colorectal, prostate), Alzheimer's disease (AD), T2D, and depression.

**Results:** We observe a significant depletion of PRS in the healthy elderly compared to the UK Biobank, suggesting a unique genetic architecture of healthy ageing. The most significant depletion was in PRS for cardiovascular disease, including early-onset coronary artery disease and atrial fibrillation. Depletion was also found for cancer (breast and prostate), Alzheimer's disease and 'short lifespan' PRS.

**Conclusion:** The healthy elderly have, on average, depletion of common damaging variants, and enrichment of common protective alleles, across a range of disease phenotypes compared to the UK Biobank population. This, in part, may contribute to the complex phenotype of healthy ageing.

## PgmNr 2707: Anti-inflammatory statins: An insight into the relationship between *UGT1A* and statin metabolism.

### Authors:

S. Yee; C. Drum

[View Session](#) [Add to Schedule](#)

**Affiliation:** Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

---

**Background** – HMG-CoA reductase inhibitors, or statins, are one of the most prescribed drugs for the prevention of cardiovascular disease and act by reducing plasma LDL cholesterol. In Singapore, 47% of adult patients use either atorvastatin or simvastatin. Statin medications are known for their various lipid independent effects and these effects can contribute to clinical decision making regarding the prescription of statins.

**Methods** – 1,500 subjects on atorvastatin and simvastatin were recruited at two outpatient heart clinics, with each clinic contributing half of the study population. Two blood draws, three hours apart, are done per patient per visit via venipuncture. Blood drawn is used for genomic analysis then spun down to plasma for clinical biomarker measurements via COBAS c111 and drug and metabolite measurements via LC-MS/MS.

**Preliminary Results** – To discover personalized predictors of inflammation reduction, we studied 1,500 Singaporeans recruited from restructured hospitals over 3 years. Mass spectrometry was performed on all samples to quantify individual profiles of parent drug and active metabolites. Genome-wide genotyping was performed on all samples (~647,000 SNPs) to determine common variants associated with statin metabolism and clinical outcomes. Clinical case annotation was provided from case report forms and electronic medical records. Plasma atorvastatin and metabolite levels were quantified and analysed. Our data demonstrates that despite individualized variance in statin metabolism, there is a strong correlation between the ratios of drug metabolite to parent drug and hsCRP ( $p < 0.001$ ), whereas quantified levels of the parent drug have no correlation. Genetic determinants of altered statin metabolism also identify patients with decreased levels of CRP. Coding variants known to reduce function of the statin metabolizing enzyme, UDP glucuronosyltransferase 1A (*UGT1A1*), result in a unique statin phenotype ( $p < 10^{-14}$ ) and result in low CRP values when compared with controls ( $p < 0.05$ ).

**Conclusion** – We have used community derived statin measurements to discover new variants that could be involved in statin metabolism. Using this method, we could expand our methods to look at the other pleiotropic effects of statins and also to discover new variants involved in the metabolism of other drugs

## PgmNr 2708: Cardiovascular pathophysiology in the KATP channelopathy Cantú syndrome: Mechanisms and pharmacotherapy.

### Authors:

C. McClenaghan<sup>1,2</sup>; Y. Huang<sup>1,2</sup>; T. Harter<sup>1,2</sup>; G. van Haaften<sup>3</sup>; D.K. Grange<sup>2,4</sup>; M.S. Remedi<sup>2,5</sup>; C.G. Nichols<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for the Investigation of Membrane Excitability Diseases, Washington University School of Medicine, St. Louis, MO; 2) Cell Biology and Physiology, Washington University School of Medicine, St. Louis, MO; 3) Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands.; 4) Pediatrics, Washington University School of Medicine, St. Louis, MO; 5) Medicine, Washington University School of Medicine, St. Louis, MO

---

Cantú syndrome (CS) is an autosomal-dominant, multi-organ disorder, characterized by hypertrichosis and craniofacial dysmorphology, which arises from gain-of-function mutations in the genes encoding the cardiovascular KATP channel subunits Kir6.1 and SUR2 (*KCNJ8* and *ABCC9*, respectively). Multiple cardiovascular abnormalities are reported in CS including dramatic cardiomegaly, alongside vascular dilation and low systemic blood pressure. How KATP dysfunction results in these complex abnormalities is not fully understood and there are no known therapies for the disorder. Using “Cantú mouse” models, in which disease-causing Kir6.1[V65M] and SUR2[A478V] substitutions were introduced, we have identified the pathophysiological mechanisms underlying CV remodeling and tested the efficacy of the KATP inhibitor glibenclamide in reversing abnormalities.

Cantú mice exhibit markedly decreased systemic vascular resistance (SVR) and high-output cardiac hypertrophy alongside decreased exercise tolerance – suggestive of decreased cardiac reserve and mirroring clinical observations. Expression of transgenic, dominant-negative Kir6.1 subunits specifically in smooth muscle reverses cardiac hypertrophy in Cantú mice, suggesting that cardiac remodeling occurs secondary to vascular dilation and chronically lowered SVR. We demonstrate that renin-angiotensin signaling (RAS) is upregulated in Cantú mice and drives cardiac hypertrophy. Finally, we show that chronic administration of the KATP channel inhibitor glibenclamide results in normalization of vascular and cardiac abnormalities in Cantú mice.

These results demonstrate that KATP over-activity results in decreased electrical excitability in vascular smooth muscle (VSM), upregulation of RAS, and maladaptive cardiac remodeling. We propose that the complex CV phenotypes observed in CS may represent a defining example of the systemic consequences of VSM hypo-excitability, emphasizing how the study of rare, genetically defined diseases can reveal novel insights into broad physiology. Demonstration of the *in vivo* efficacy of glibenclamide in targeting cardiovascular KATP channels suggests that the drug can be clinically repurposed for the treatment of CS specifically - and potentially for diverse cardiovascular conditions arising from decreased smooth muscle excitability in general.

## PgmNr 2709: Large scale proteome-wide imputation and association analysis identifies disease biomarkers and candidate drug targets.

### Authors:

B.P. Prins; J. Danesh; A.S. Butterworth

[View Session](#) [Add to Schedule](#)

**Affiliation:** MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, 2 Worts' Causeway, Cambridge CB1 8RN, UK

---

Imputing molecular data into large cohorts without such measurements is a powerful and affordable way to identify and prioritize molecular markers associated with disease risk.

Previous studies have focused on imputation of RNA expression data by calculating risk scores using machine learning approaches, followed by Transcriptome Wide Association Studies (TWAS). However, variation in expression levels does not always translate to variation in protein levels, the latter appearing to be far more conserved. This phenotypic buffering is the result of (post-) translational regulation of gene expression, and may result in sub-optimal prioritization of risk-associated biomarkers. Furthermore, proteins are the most common targets of drugs, therefore identification of protein disease biomarkers may be more informative.

Here we perform the largest imputed proteome-wide association study (PWAS) to date to identify proteins and novel genomic regions for 20 complex diseases and traits for up to 1 million samples. We constructed protein expression models by performing elastic net regression on 2,869 autosomally coded plasma proteins (SomaLogic) in 3,301 individuals from the INTERVAL study. Ten-fold nested cross-validations were performed to compute the coefficient of determination  $R^2$  to assess model performance.

For 626 probes (584 unique proteins) we were able to build models with a minimal prediction performance ( $R^2_{cv}$ ) of at least 0.01, with the largest  $R^2_{cv}$  for CLEC12A (0.77). Next, we used our prediction models to identify previously unknown protein-disease associations with a set of 20 traits and common complex diseases for which GWAS summary statistics were publicly available. We identified 51 previously unknown associations with biological relevance. For osteoarthritis as an example, we identified 7 novel protein associations, including VIT and CRELD1, both, (extra-) cellular matrix proteins, as well as ITIH1 and HP, both of which are implicated in multiple inflammatory diseases.

In summary, we have created a powerful new paradigm for allowing imputation of expensive plasma protein measurements into large-scale biobanks and subsequent association analyses, and highlighted its potential to identify new biomarkers and drug targets for complex diseases.



## PgmNr 2710: Epigenetic mapping of African American hepatocyte methylation quantitative trait loci (meQTLs) in Clopidogrel response: An ACCOuNT Study.

### Authors:

C.S. Park; T. De; Y. Zhong; C. Alarcon; J.M. Avitia; M.A. Perera; ACCOuNT

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pharmacology, Northwestern University Feinberg School of Medicine, Chicago, Illinois.

Genetic variation may give rise to different patterns of gene expression and DNA methylation. However, the combination of genetic variability and epigenetic modifications that control gene expression and gene-drug interactions is understudied in minority populations. In particular, DNA methylation in the liver may affect gene expression of important hemostatic factors which may in turn affect response to cardiovascular medication. Moreover, the greater degree of genetic variation and differences in gene expression seen in African Americans (AAs) may uncover novel hepatocyte meQTLs. Here we investigate the role of hepatocyte meQTLs in explaining variability in the anti-platelet medication, clopidogrel.

In this study, we used genotype and methylation array data from 44 AA primary hepatocyte cultures to first map meQTLs in AA hepatocytes and then to investigate these potential regulatory variants for association with clopidogrel response. We identified 404 cis-meQTLs mapped to 142 unique CpG sites using a window size of  $\pm 500$  kb of the CpG (FDR < 0.05). These 404 hepatocyte meQTLs were analyzed for their association with P2Y<sub>12</sub> reaction units (PRU) in clopidogrel-treated AA patients. PRU is a clinical measure of response and has previously been shown to vary widely in AAs. After adjusting for age and first PC, we found a meQTL (rs115789536,  $\beta = 0.0257$ , FDR = 0.0298), which increased the methylation at a nearby CpG site (cg07110418), was associated with decreased on-treatment PRU ( $\beta = -1.387$ ,  $p = 0.01$ ).

rs115789536 is predominantly observed in populations with African ancestry (minor allele frequency = 7%) and is a non-coding variant in Prostaglandin I<sub>2</sub> Synthase (*PTGIS*), a PharmGKB very important pharmacogene (VIP). Previous studies have shown that SNPs within *PTGIS* were associated with myocardial infarction risk, systolic hypertension and cerebral infarction. We previously found *PTGIS* gene expression to be negatively associated with increased percentage of inferred African ancestry.

Our data suggests that altered gene expression of *PTGIS* in the liver due to the meQTL rs115789536 may lead to increased response to clopidogrel. This analysis provides evidence for a novel role of *PTGIS* in platelet aggregation and potentially in response to clopidogrel as a result of altered methylation. Functional validation is needed to validate the effect of rs115789536 in both altered methylation and gene expression of *PTGIS*.

## PgmNr 2711: Development of a gene-diet interaction-based dietary fat response score for cardiometabolic risk factors.

**Authors:**

K. Westerman; L. Parnell; J.M. Ordovás

[View Session](#) [Add to Schedule](#)

**Affiliation:** Nutrition and Genomics Laboratory, JM-USDA Human Nutrition Research Center on Aging, Boston, MA

---

Many gene-diet interactions (GDI) have been uncovered for cardiometabolic risk factors (CRFs), but personalized nutrition recommendations will require the incorporation of an ensemble of variants across the genome in predicting diet response. Dietary trials typically lack the sample sizes required to identify genome-wide signals of diet response. Using the Women's Health Initiative (WHI) dataset, we set out to replicate previously-identified GDIs and develop genome-wide dietary scores for CRF response to dietary fat response. First, known GDIs for dietary fat were collected from the CardioGxE database for each of six CRFs: body mass index, systolic blood pressure, LDL-cholesterol, HDL-cholesterol, triglycerides, and fasting glucose. These GDIs were tested for nominal replication ( $p < 0.05$ ) with respect to food frequency-based binary dietary fat percentage in women from WHI not participating in the dietary modification (DM) trial. We found that only a small percentage of these GDIs could be replicated in this almost 10,000-person dataset, suggesting that these literature-reported interactions are population-specific or reflect confounding factors. Next, GDIs were calculated in this same cohort for suggestive SNPs for each risk factor (main-effect GWAS  $p < 1e-5$ ) and aggregated using the pruning and thresholding method into a series of "fat response scores". These scores were tested for prediction of changes in the associated CRF after 3 years in WHI women participating in the fat reduction-focused DM trial. Only the HDL responder score was meaningfully predictive, predicting a 3.1 mg/dL lower change in HDL-C per score standard deviation in DM trial participants ( $p = 0.01$ ). Finally, this HDL response score was assessed with respect to incident coronary heart disease events in DM trial participants, but failed to discriminate between incident cases and non-cases.

## PgmNr 2712: Transcriptome sequencing of patients with hypertrophic cardiomyopathy reveals deep intronic variants amenable to antisense oligonucleotide therapy.

### Authors:

M. Holliday<sup>1,2</sup>; E.S. Singer<sup>1</sup>; S.B. Ross<sup>1,2</sup>; S. Lim<sup>1</sup>; J. Ingles<sup>1,2,3</sup>; S. Lal<sup>2,3</sup>; C. Semsarian<sup>1,2,3</sup>; R.D. Bagnall<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Agnes Ginges Centre for Molecular Cardiology at Centenary Institute, The University of Sydney, Sydney, NSW, Australia; 2) Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia; 3) Department of Cardiology, Royal Prince Alfred Hospital, Sydney, NSW, Australia

Hypertrophic cardiomyopathy (HCM) is an inherited heart disease with a prevalence of 1 in 500 in the general population. DNA-based genetic testing fails to identify a genetic cause of disease in up to 50% of families with HCM. RNA sequencing of diseased tissue is a new genetic testing approach that could increase the diagnostic yield, especially for recently described deep intronic variants, which are often overlooked. However, there is limited access to heart tissue for most patients. We have developed a novel genetic testing approach involving sequencing RNA extracted from induced pluripotent stem cell cardiomyocytes (iPSC-CMs) derived from the patient's fresh blood. We have also used the iPSC-CMs to explore antisense oligonucleotide (AO) inhibition of RNA splicing errors that are caused by deep intronic variants in *MYBPC3*, the most common HCM gene.

We derived iPSC-CMs from (a) fresh blood of two patients with HCM caused by deep intronic variants in *MYBPC3*, (b) a patient with familial HCM where no genetic cause was identified after genome sequencing, and (c) two people without HCM. Transcriptome sequencing was performed on iPSC-CMs and analysed for RNA splicing errors. Splicing errors were confirmed with RT-PCR analysis of RNA extracted from iPSC-CMs, and the primary heart tissue of one patient. We designed a series of AOs to inhibit RNA splicing errors in iPSC-CMs of patients with deep intronic variants.

Transcriptome sequencing of iPSC-CMs identified known exon-gain RNA splicing errors in two people with deep intronic variants in *MYBPC3* (c.1090+453C>T and c.1224-52G>A). In a patient with a previously unresolved genetic cause of HCM, transcriptome sequencing of iPSC-CMs revealed gain of two novel exons due to an underlying *MYBPC3* c.1928-569G>T intronic variant. The same RNA splicing error was confirmed in primary heart tissue of the patient. We used a gradient of AOs (0.5-20  $\mu$ M) to determine the optimal concentration for inhibition of RNA splicing errors. Treatment of iPSC-CMs derived from all 3 patients with deep intronic variants using a series of AOs inhibited splicing errors, whereas a control AO did not inhibit splicing.

Our novel genetic testing approach involving sequencing of iPSC-CM RNA can provide additional genetic diagnoses over current DNA based testing, including genome sequencing. Patient derived iPSC-CMs provide a unique opportunity to explore and optimise the design of AOs for patient-specific inhibition of RNA splicing errors.

## PgmNr 2713: Actionable cardiac genetic variants and variants of uncertain significance in diverse biobank participants.

### Authors:

T.D. Pottinger <sup>1</sup>; M.J. Puckelwartz <sup>1,2</sup>; L.L. Pesce <sup>3</sup>; A. Robinson <sup>1</sup>; S. Kearns <sup>1</sup>; J.A. Pacheco <sup>1</sup>; L.J. Rasmussen-Torvik <sup>4</sup>; M.E. Smith <sup>1</sup>; R. Chisholm <sup>1,5</sup>; E.M. McNally <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL.; 2) Department of Pharmacology, Northwestern University Feinberg School of Medicine, Chicago, IL.; 3) Computation Institute, University of Chicago, Chicago, IL.; 4) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL.; 5) Department of Cell and Molecular Biology, Northwestern University Feinberg School of Medicine, Chicago, IL.

---

**Rationale:** Genetic profiling of medical biobanks offers the opportunity for risk identification and risk reduction. However, many biobanks lack sufficient population diversity to adequately represent the US population. We now analyzed whole genome sequence of 900 diverse medical biobank participants from a single US medical center.

**Methods:** We evaluated variation among participants of different ancestry. Nonsynonymous genetic variants, as reported in ClinVar were compared across racial/ethnic groups. Echocardiographic measures, obtained from the electronic health record, were correlated with genetic variation in the medically actionable genes for cardiac conditions in order to ascertain whether variation correlates with clinical phenotypes.

**Results and conclusions:** Biobank subjects were equally split representing European, African, Hispanic, and mixed race/ethnicities. Compared to those of European descent, genetic diversity was greater among those of African, Hispanic, and mixed race/ethnicity. When evaluating all nonsynonymous protein coding variants, African ancestry participants were more likely to have variants of uncertain significance (VUS) compared to other groups ( $p < 0.05$ ) and were also more likely to have variants not previously reported in ClinVar ( $p < 0.05$ ). We evaluated the American College of Medical Genetics and Genomics (ACMG) medically actionable list of 59 genes across races/ethnicities focusing genes for cardiovascular conditions (30 of 59 genes). We identified 19 individuals with pathogenic/likely pathogenic variants in the 30 cardiac medically actionable genes. African ancestry participants had more uncertain variants in the cardiac medically actionable genes. Longitudinal measures of left ventricle size, corrected for body surface area, from approximately 400 biobank participants (1,723 patient years) correlated with genetic findings. The presence of one or more uncertain variants in the actionable cardiac genes correlated with increased left ventricular internal diameter in diastole ( $p < 0.05$ ) and in systole ( $p < 0.01$ ), and these analyses identified excess variation in *MYBPC3*. We hypothesize subsets of the uncertain variants confer risk. Furthermore, we note that uncertain variants are not normally returned to participants or providers. These data suggest that in some settings, sharing this information with participants and providers offers opportunity for risk reduction.

## PgmNr 2714: Whole exome sequencing and monogenic hypertension in a multiethnic cohort of 28,000 individuals.

### Authors:

I. Paranjpe; K. Johnson; S. Coca; L. Chan; J.C. He; B. Murphy; J. Narula; R. Do; G. Nadkarni

[View Session](#) [Add to Schedule](#)

**Affiliation:** Icahn School of Medicine, New York, NEW YORK.

Hypertension (HTN) is a major cause of morbidity and mortality and a subset may be caused by monogenic mutations which influence treatment decisions and prognosis. We identified the prevalence of these secondary HTN mutations in a multiethnic cohort and determined their association with blood pressure, receiving a diagnosis of secondary HTN, and adverse cardiovascular outcomes.

In 27,972 individuals from the Mount Sinai BioMe Biobank (8304 European, 6993 African, 9985 Hispanic, and 2690 Other ancestry), we selected ClinVar mutations pathogenic for secondary hypertension using whole exome sequencing.

3,125 individuals (11.2%) had pathogenic mutations with a majority in genes associated with catecholamine excess (*SDHD*, *SDHB*, *TMEM127*, *RET*; n=2210) and sodium handling/hyperaldosteronism (*SCNN1G*/*CYP11B1*; n=524). (Table 1) In a linear model adjusted for sex, age, BMI, and 10 genetic principal components (PCs), individuals with mutations had elevated mean arterial pressure (b=0.7±0.6 mmHg; p=0.03) and increased odds of coronary artery disease (CAD) (adjusted OR=1.12, 95% CI 1.0 to 1.3, p=0.04). Individuals with mutations in sodium handling/hyperaldosteronism genes had increased odds of congestive heart failure (CHF) (adjusted OR = 1.44, 95% CI 1.04 to 1.97; p=0.02) adjusted for age, sex, 10 genetic PCs, and systolic blood pressure. The majority of individuals with mutations were diagnosed with essential HTN (54.2%) and only 7.8% had appropriate diagnoses of secondary HTN or received appropriate biochemical evaluation. (Table 2)

Individuals with pathogenic mutations had higher blood pressures, elevated risk for CAD/CHF and the majority were not appropriately evaluated or diagnosed. These results suggest whole exome sequencing may have utility in hypertension diagnosis and management.

**Table 1**

Gene	Condition	Number of Individuals with Variant
<i>SDHD</i>	Pheochromocytoma	1219 (3.9%)
<i>SDHB</i>	Pheochromocytoma	611 (2%)
<i>GCGR</i>	Diabetes mellitus type 2	515 (1.7%)
<i>SCNN1G</i>	Pseudoprietary hyperaldosteronism	393 (1.3%)
<i>TMEM127</i>	Pheochromocytoma	196 (0.6%)

<i>RET</i>	Pheochromocytoma	184 (0.6%)
<i>GDNF</i>	Pheochromocytoma	168 (0.5%)
<i>SERPINA6</i>	Corticosteroid-binding globulin deficiency	131 (0.4%)
<i>VHL</i>	Von Hippel-Lindau syndrome	45 (0.1%)
<i>CYP11B1</i>	Hyperaldosteronism; Congenital adrenal hyperplasia	33 (0.1%)
Other		170 (0.5%)

**Table 2**

	<b>No Variant</b>	<b>With Variant</b>
<b>No Hypertension Diagnosis</b>	1181 (37.8%)	9843 (39.6%)
<b>Essential Hypertension Diagnosis</b>	1695 (54.2%)	13088 (52.7%)
<b>Secondary Hypertension Diagnosis or Lab Evaluation</b>	249 (8.0%)	1916 (7.7%)

## PgmNr 2715: A genetic model of ivabradine treatment recapitulates challenging results from randomized clinical trials.

### Authors:

M.A. Legault<sup>1,2,3</sup>; J. Sandoval<sup>2,3</sup>; S. Provost<sup>2,3</sup>; A. Barhdadi<sup>2,3</sup>; L.P. Lemieux Perreault<sup>2,3</sup>; S. de Denus<sup>1,2</sup>; J.C. Tardif<sup>2,4</sup>; M.P. Dubé<sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Département de biochimie et médecine moléculaire, Université de Montréal, Montreal, Canada; 2) Montreal Heart Institute, Montreal, Canada; 3) Université de Montréal Beaulieu-Saucier Pharmacogenomics Centre, Montreal, Canada; 4) Département de médecine, Université de Montréal, Montreal, Canada

---

Naturally occurring genetic variants are being used as tools for drug target discovery and to guide clinical trial design. Ivabradine is a heart rate lowering drug that reduces cardiovascular outcomes in patients with heart failure (SHIFT trial), but it does not reduce cardiovascular outcomes in patients with stable coronary artery disease (SIGNIFY trial), presenting a challenging model for a proof of concept demonstration of the value of genetics to support drug trial development. Ivabradine also increases the risk of atrial fibrillation (AF), which can lead to cardioembolic stroke.

Using the UK Biobank data and summary statistics from large GWAS, we evaluated the effect of a heart-rate lowering polymorphism at the *HCN4* locus encoding the drug target of ivabradine. Genetic association recapitulated the increase in AF risk (OR=1.09, 95%CI: 1.06-1.13,  $p=9.3 \times 10^{-9}$ ) in the UK Biobank and cardioembolic stroke (OR=1.08, 95%CI: 1.03-1.13,  $p=0.0015$ ) in MEGASTROKE. Using a competing risk model accounting for AF with the UK Biobank prospective data, we show that the heart rate reducing *HCN4* variant reduced incident heart failure (HR=0.90, 95%CI: 0.83-0.98,  $p=0.013$ ), recapitulating results of the SHIFT trial. We used bi-directional Mendelian randomization to show that AF and heart failure are mutually causal of each other, supporting the competing risks approach. The effect of the *HCN4* variant on AF exceeds the effect estimated by Mendelian randomization with genome-wide variants, which suggests a pleiotropic effect of *HCN4* on heart failure. The *HCN4* variant did not associate with coronary artery disease in CARDIOGRAMplusC4D, recapitulating results from the SIGNIFY trial.

Our approach supports the usefulness of human genetics data to guide drug development and trial design even in more complex clinical settings.

## PgmNr 2716: Inter-individual genetic variation leads to differences in exosomal miRNA profiles in response to kinase inhibitors and the development of cardiotoxicity.

### Authors:

A. Turner <sup>1</sup>; A. Matter <sup>1</sup>; P. Aggarwal <sup>1</sup>; B. Olson <sup>1</sup>; M. White <sup>2</sup>; U. Broeckel <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Pediatrics, Children's Research Institute and GSPMC, Medical College of Wisconsin, Milwaukee, WI; 2) Arkansas Colleges of Health Education, Fort Smith, AR

---

The use of targeted oncologic treatments, such as kinase inhibitors (KIs), have greatly improved therapeutic outcome and significantly reduced mortality. However, for many patients, drug-related adverse effects (AEs) including cardiotoxicity occur. Human induced pluripotent stem cell-derived cardiomyocytes (hiPSC-CMs) allow for the study of inter-individual response to KIs, the identification of biomarkers for AEs and the potential for a personalized approach to drug selection and dosing. Recent studies have found circulating exosomal miRNAs (miRNA<sup>EXO</sup>) to be promising biomarkers for drug response. The aim of this study was to comprehensively analyze inter-individual exosomal miRNA expression in hiPSC-CMs in response to KIs to identify exosomal miRNA profiles associated with cardiotoxicity. For toxicity modeling, we utilized six individual hiPSC-CM lines. Each line was treated in triplicate with three KIs, sunitinib (SUN), vandetanib (VAN) and gefitinib (GEF). Physiological measurements of beat rate, cell index and ATP viability were used to test for CM toxicity (20% change +/- from the control value). Exosomes were isolated from cell conditioned media and miRNA libraries were prepared. Next-Gen sequencing and differential expression analysis (DESeq2) identified miRNAs associated with KI response and cardiotoxicity. For VAN and GEF, we identified significantly differentially expressed miRNA<sup>EXO</sup> (n=35 and n=3, adj. p-value ≤ 0.05) in response to each KI common in all cell lines, establishing a miRNA<sup>EXO</sup> profile exclusive to each drug. Interestingly, with SUN and VAN, we observed miRNA<sup>EXO</sup> profiles unique only to cell lines that developed a cardiotoxic drug response. Differential expression analysis of the miRNA<sup>EXO</sup> comparing the cell lines that developed cardiotoxicity to those that were resistant identified 13 miRNAs with significant differential expression for SUN and 11 miRNAs for VAN (adj. p-value ≤ 0.05). In summary, our methodology of combining functional CM analysis in response to KIs with exosomal miRNA expression allows for the identification of potential biomarkers of cardiotoxicity. Our results highlight the inter-individual response to KIs and the need for toxicity screening in multiple, genetically diverse cell lines. Further studies will identify the functional effects of these differentially expressed miRNAs and further determine their utility as a predictive pharmacogenomic biomarker for individualized drug dosing and response.



## **PgmNr 2717: An intergenic variant near *C12orf42*, *LINC02401* and *STAB2* is associated with acute coronary syndrome among Filipinos.**

### **Authors:**

E. Cutiongco de la Paz <sup>1,2</sup>; J. Nevado <sup>1</sup>; E.J. Llanes <sup>3</sup>; J.D. Magno <sup>3</sup>; D. Ona <sup>3</sup>; F.E. Punzalan <sup>3</sup>; P.F. Reganit <sup>3</sup>; L.E. Santos <sup>3</sup>; R.H. Tiongo II <sup>3</sup>; J.A. Aherrera <sup>3</sup>; L. Abrahan IV <sup>3</sup>; C. Agustin <sup>3</sup>; A.Y. Aman <sup>1,2</sup>; A.J. Bejarin <sup>1,2</sup>; R. Sy <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) National Institutes of Health, University of the Philippines Manila, Manila, Philippines; 2) Philippine Genome Center, University of the Philippines; 3) Section of Adult Cardiology, Department of Medicine, College of Medicine, Philippine General Hospital, University of the Philippines Manila

---

Acute coronary syndrome (ACS) pertains to a group of conditions consistent with the clinical manifestations of acute myocardial ischemia or infarction secondary to the decrease in blood flow to the coronary arteries. Because ACS is a life-threatening disease, prevention is of utmost concern especially among patients who have a genetic predisposition. Early detection of these markers may improve the risk prediction of ACS by considering both traditional risk factors and the presence of these variants. While various genes have been studied, Filipinos are underrepresented in most studies looking at the genetic determinants of acute coronary syndrome (ACS). This study sought to determine the association of previously identified genetic variants linked with the development of ACS among Filipinos. A total of 123 adults unrelated Filipinos with ACS were compared with 228 controls in an age- and sex- matched case-control study that looked at 81 genetic variants associated with ACS and similar phenotypes. Genotyping using DNA from blood samples was done using customized Illumina Goldengate microarray chips. Candidate variants and clinical data were correlated with the incidence of ACS using chi-square and logistic regression analysis. An intergenic variant located upstream of *C12orf42* (~22kb), downstream of *LINC02401* (~28kb), and of *STAB2* (~68kb) exhibited robust association with development of ACS among Filipinos (TT vs CC: adjusted OR 2.33,  $p = 0.022$ ), as adjusted for presence of hypertension, borderline to high LDL and history of smoking. The variant is also found 110 bp downstream from a DNase I hypersensitivity site, which is a general feature of cis-regulatory functions, involving enhancers, promoters, insulators, boundary elements, and locus control regions. Previous genome-wide association studies among European and American populations showed, however, that the C allele was associated with restenosis after percutaneous coronary intervention, coronary events and all-cause increased mortality. Validation studies will be done prior to concluding definite association between this marker and coronary artery disease among Filipinos.

## PgmNr 2718: Transcriptome-wide association study of novel copy number variant: associated with venous thromboembolism in African Americans.

### Authors:

H. Zhang; L. Nahwahli; W. Hernandez; M. Perera

[View Session](#) [Add to Schedule](#)

**Affiliation:** Northwestern University, Chicago, Illinois.

Venous thromboembolism (VTE) is a disorder consisting of deep vein thrombosis and pulmonary embolism and the third leading cause of cardiovascular-related mortality in the United States. VTE is multifactorial disease including environmental and genetic risk factors. African Americans (AAs) have a 30 to 60% higher incidence compared to other ethnicities. Although several risk and genetic factors have been investigated, the concrete mechanism underlying the ethnic disparities remains unclear. Through the first genome-wide analysis of copy number variation (CNV), we identified a deletion of approximately 18 kb on Chromosome 10. This CNV region (CNVR) was validate using a custom 8x15K oligonucleotide array (Agilent) and was significantly associated with decreased risk of VTE in AAs. Previous GWAS identified a SNP near this CNVR (rs2420915) which was associated to fibrinogen level, lending support for the importance of this region in coagulation.

To identify other potential gene expression traits linked to this CNVR, we performed transcriptome-wide association study (TWAS), which imputes gene expression in relevant tissues and tests the association of the CNVR to imputed gene expression. Our gene expression imputation was the method employed in PredXcan, however, we used previously published genotype weights trained in AAs from Multi-Ethnic Study of Atherosclerosis (MESA) in monocytes. We identified a subset of well-predicted genes and integrated gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotations as features to characterize these genes. We obtained 10 optimal GO terms ( $p < 0.05$ ) and 1 KEGG sphingolipid metabolic pathway ( $p = 0.0018$ ) that characterized these genes well. Interestingly, Phospholipase D1 (PLD1,  $p = 0.0147$ ), Sphingomyelin Phosphodiesterase 2 (SMPD2,  $p = 0.0147$ ), and Sphingosine Kinase (SPHK1,  $p = 0.0373$ ) regulate the lipid metabolism and significantly associated with the deletion of CNVR.

Previous studies have shown that plasma sphingolipids regulate blood coagulation system and related to VTE and stroke. Through this functional annotation analysis, our data suggested that the sphingolipid metabolic pathway might related to the VTE underlying the mechanism of deletion of the CNVR to decrease the risk of VTE in AAs.

## PgmNr 2719: Genome-wide association study of on-aspirin platelet aggregation identifies *ACYP2* and other loci as novel pharmacogenetic candidates.

### Authors:

A.D. Johnson <sup>1</sup>; M.-H. Chen <sup>1</sup>; B.A.T. Rodriguez <sup>1</sup>; D. Chasman <sup>2</sup>; K. Hall <sup>3</sup>; Z.E. Schneider <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Population Sciences Branch, NHLBI's Framingham Heart Study, Framingham, Massachusetts.; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; 3) Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

---

Aspirin is often used in primary and secondary prevention of cardiovascular disease (CVD). The efficacy of aspirin and other anti-platelet treatments is linked to their ability to suppress platelet activation, a key component in thrombosis. The concept of aspirin resistance is controversial, with unknown prevalence and many studies finding on re-testing that suspected aspirin resistance was due to non-compliance or self-report error. Thus, the extent of differences in aspirin efficacy in the population due to genetic or environmental effects remains to be determined. To study the genetic factors influencing on-aspirin platelet aggregation in the European ancestry Offspring cohort of the Framingham Heart Study, we used a gold standard platelet activation assay (light transmission aggregometry) after addition of arachidonic acid as a sensitive determinant of aspirin takers. Then we conducted WGS-imputed GWAS analysis in a subset of individuals who were taking aspirin at the time of platelet activation measurement. We focused on QTL analysis of platelet activation via 3 agonist pathways (ADP, epinephrine and collagen) that partially escape aspirin inactivation. From these analyses we found 2 genome-wide significant loci ( $P < 5E-8$ ), and 2 borderline loci ( $P < 1E-7$ ) for platelet activation. All four loci (*ABAT*, *MMD2*, *CSH1*, *ACYP2*) had a minor allele associated with decreased epinephrine-induced platelet activation, indicating those individuals had greater platelet suppression on aspirin treatment. We attempted replication in a smaller, diverse sample – the Framingham Omni cohort – and a Welsh population sample – the Caerphilly Prospective Study in Men – and found evidence for association in the same direction in aspirin takers in the region of *ACYP2*. In interaction models including non-aspirin takers there was a significant interaction with the *ACYP2* peak SNP ( $P < 0.03$  interaction). This sentinel SNP is ~6kb upstream of the consensus start site and overlaps an area of epigenetic activity in megakaryocytes, the precursor cell for platelets. Acylphosphatase-2 (*ACYP2*) hydrolyzes calcium channel intermediates, and calcium signaling plays a critical role in platelet activation, particularly for epinephrine-based stimulation. Efforts are currently underway to evaluate whether the identified SNPs contribute to future risk of CVD or bleeding in the Women's Genome Health Study, a randomized trial of aspirin, vitamin E and placebo with >20 years follow-up.

## PgmNr 2720: A genomics-based approach identifies gender-specific modifiable risk markers of left ventricular mass variability within a population.

### Authors:

J.D. Mosley<sup>1,2</sup>; R.T. Levinson<sup>1</sup>; T.L. Edwards<sup>1,3</sup>; J.N. Hellwege<sup>1,3,4</sup>; M.M. Shuey<sup>1</sup>; C.M. Shaffer<sup>1</sup>; M. Shi<sup>1</sup>; E.L. Brittain<sup>1</sup>; W.K. Chung<sup>5</sup>; A.M. Arruda-Olson<sup>6</sup>; M.S. Williams<sup>7</sup>; D.R. Crosslin<sup>8</sup>; J.C. Denny<sup>1,2</sup>; C.M. Stein<sup>1,9</sup>; D.M. Roden<sup>1,2,9</sup>; Q.S. Wells<sup>1,9</sup>; eMERGE contributors, VA Million Veteran Program

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee.; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN; 3) Vanderbilt Epidemiology Center, Vanderbilt University Medical Center, Nashville, TN; 4) Tennessee Valley Healthcare System (626), Vanderbilt University, Nashville, TN; 5) Departments of Pediatrics and Medicine, Columbia University Medical Center, New York, NY; 6) Department of Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 7) Genomic Medicine Institute, Geisinger, Danville, PA; 8) Departments of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA; 9) Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN

**Background:** Left ventricular (LV) mass is a prognostic biomarker of incident heart disease and all-cause mortality. Over time, LV mass may increase due to exposure to risk factors such as high blood pressure. Treating these risk factors using medications or other interventions attenuates or reverses these LV mass increases. We hypothesized that probing the genetic architecture of an integrative phenotype like LV mass, which captures the duration and severity of exposure to risk factors over time, would identify insufficiently treated cardiac risk mechanisms in a target population.

**Methods:** We developed a polygenic single nucleotide polymorphism predictor of LV mass variation in 7,601 individuals with transthoracic echocardiography measurements collected as part of routine clinical care at Vanderbilt University Medical Center. We then tested for associations between this predictor and 894 clinical diagnoses measured in 58,838 genotyped individuals in the Electronic Medical Records and Genomics (eMERGE) network. We used polygenic risk (PRS) scores for the candidate diagnoses in conjunction with inverse-weighted meta-analyses to further define the etiological relationships between the diagnoses and LV mass.

**Results:** There were 29 clinical phenotypes associated with the LV mass genetic predictor at FDR  $q < 0.05$ . Higher LV mass was associated with modifiable cardiac risk factors including obesity, hypertension and type 2 diabetes as well as conditions associated with pathological LV mass changes including heart failure and atrial fibrillation. Analyses using polygenic predictors confirmed significant associations between higher LV mass and body mass index (BMI) (each unit increase in BMI was associated with a 0.11 [0.09 - 0.13] unit increase in log[LV mass],  $p = 6.4 \times 10^{-25}$ ) and systolic blood pressure ( $\beta = 0.003$  [0.001 - 0.004],  $p = 0.002$ ). The magnitude of the body mass-related genetic effects was larger in women ( $p = 0.002$ ). In men, there was an association with atherosclerotic disease including myocardial infarction ( $\beta = 0.003$  [0.001 - 0.004],  $p = 0.002$ ).

**Conclusions:** We show that a polygenic predictor of a integrative phenotype which stably captures an individual's cumulative exposure to risk can be used to identify disease mechanisms contributing to morbidity within a target population. These results can guide precision treatment and prevention strategies directed toward the target population.

## PgmNr 2721: Association of polygenic risk scores for body mass index and systolic blood pressure in a pediatric cohort requiring surgery for congenital heart defects.

### Authors:

J.H. Breeyear<sup>1</sup>; J.M. Keaton<sup>1,2,3</sup>; E.S. Torstenson<sup>1,2</sup>; A.H. Smith<sup>4,5</sup>; S.L. Van Driest<sup>4</sup>; J.G. Weiner<sup>5</sup>; P.J. Kannankeril<sup>5</sup>; T.L. Edwards<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, Tennessee.; 2) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; 3) Institute for Medicine and Public Health, Vanderbilt University Medical Center, Nashville, Tennessee; 4) Departments of Pediatrics and Medicine, Vanderbilt University Medical Center, Nashville, Tennessee; 5) Thomas P. Graham Jr. Division of Cardiology, Department of Pediatrics, Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, Tennessee

---

Congenital malformations are the leading cause of infant mortality in the United States. Congenital heart disease (CHD) represents the most prominent congenital malformation, affecting 40,000 US births per year. Approximately half of these children require surgical intervention(s), many of which experience postoperative complications. To date, no studies have evaluated associations of polygenic risk scores (PRS) for anthropometric traits with postoperative outcomes in a pediatric CHD cohort. Further, the utility of PRS to predict outcomes such as obesity in pediatric subjects is largely unknown. We used high-quality imputed genotypes from pediatric participants requiring surgery for CHD (mean age at surgery  $\pm$  SD =  $3.38 \pm 5.39$  years,  $n = 1978$  subjects). Base data for a body mass index (BMI) PRS used the Genetic Investigation of ANthropometric Traits (GIANT) consortium GWAS transehtnic 2015 BMI data ( $n_{\text{max}} = 322,154$  subjects). Base data for a systolic blood pressure (SBP) PRS used recently published GWAS SBP data ( $n_{\text{max}} = 760,226$  subjects). Target data were pruned for linkage disequilibrium at an  $r^2$  threshold of 0.1 at a maximum distance of 250 kilobases and scores were calculated in PLINK. Associations of PRS for BMI and SBP with BMI and length of hospital stay following surgery (LOS) were modeled using linear regression in R and adjusted for age, sex, and 10 PCs. The SBP model was also adjusted for BMI. BMI PRS was associated with BMI (PRS  $P$  value threshold = 0.001,  $n$  SNPs = 1,603,  $\beta \pm \text{SE} = 0.58 \pm 0.15 \text{ kg/m}^2$ ,  $P = 8.5 \times 10^{-5}$ ) and LOS (PRS  $P$  value threshold = 0.05,  $n$  SNPs = 16,244,  $\beta \pm \text{SE} = 1.67 \pm 0.83$  days,  $P = 0.045$ ). The SBP PRS was associated with LOS (PRS  $P$  value threshold = 0.0001,  $n$  SNPs = 2,985,  $\beta \pm \text{SE} = -0.45 \pm 0.15$  days,  $P = 3.02 \times 10^{-3}$ ). These results demonstrate the ability of PRS developed in adults to predict pediatric traits and outcomes. Specifically, the association of the BMI PRS with BMI provides proof of principle for the use of adult-derived PRSs in younger populations. PRS associations in pediatric participants may identify individuals at higher risk of negative cardiometabolic outcomes.

## PgmNr 2722: An externally validated artificial intelligence driven integrated genetic-epigenetic DNA test for 3-year risk for incident coronary heart disease.

### Authors:

M. Dogan<sup>1,2,5</sup>; S. Knight<sup>3</sup>; A. Lendasse<sup>4</sup>; K. Knowlton<sup>3</sup>; R. Philibert<sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical Engineering, University of Iowa, Iowa City, IA; 2) Cardio Diagnostics LLC, Coralville, IA; 3) Intermountain Heart Institute, Intermountain Medical Center, Salt Lake City, UT; 4) Information and Logistics Technology Department, University of Houston, Houston, TX; 5) Psychiatry, University of Iowa, Iowa City, IA

---

Coronary heart disease (CHD) is the most common type of heart disease, accounting for over 350,000 deaths and 750,000 heart attacks annually. CHD associated morbidity and mortality are largely preventable. Current primary prevention measures include the use of risk estimators that aggregate multiple conventional risk factors. Here, we describe a novel, simpler and more sensitive DNA-based method that we have developed and externally validated. It is an integrated genetic-epigenetic risk estimator for predicting the 3-year risk for incident CHD. The implementation of this Precision Medicine tool only requires DNA from a small amount of blood and its predictive power is driven by Artificial Intelligence (AI). For its development, we utilized genome-wide DNA methylation (DNAm) and SNP data from the Framingham Heart Study (FHS) Offspring cohort. This data was split into training (34/1172 incident CHD cases) and test (8/512 incident CHD cases) sets. The external validation was performed in partnership with Intermountain Healthcare (IM), which was split into validation (23/80 incident CHD cases) and test (21/79 incident CHD cases) sets. Data mining, feature selection and model development were performed on the FHS training set and validated on the IM validation set. The final incident CHD prediction model (ensemble of SVM, Random Forest and Logistic Regression) consisted of 6 loci (3 DNAm and 3 SNPs). It performed with a sensitivity, specificity and AUC of 0.75, 0.72 and 0.74, respectively, in the FHS test set and 0.71, 0.69 and 0.70, respectively, in the IM test set. The performance of our approach was compared to that of the Framingham Risk Score (FRS) and the Pooled Cohort Equation (PCE). The sensitivity, and specificity in the FHS cohort was 0.15 and 0.94, respectively, for FRS, and 0.41 and 0.80, respectively, for PCE. Similarly, the sensitivity and specificity in the IM cohort was 0.31 and 0.89, respectively, for FRS, and 0.56 and 0.63, respectively, for PCE. The superior sensitivity of our tool suggests that its use in a clinical setting could better identify patients at risk for incident CHD that may benefit from primary prevention interventions. Also, because DNAm signatures are dynamic and map to actionable risk factors, they may be used to personalize interventions and continuously monitor risk for incident CHD. This tool is being translated into a Laboratory Developed Test and being extended to include more ethnically diverse cohorts.



## PgmNr 2723: **SOX7** ablation in endocardium results in downregulation of **WNT4** and **BMP2** and abnormal endocardial cushion development.

### Authors:

A. Hernandez-Garcia <sup>1</sup>; M. Wat <sup>1</sup>; C. Ljungberg <sup>3</sup>; C.A. Shaw <sup>1</sup>; M.E. Dickinson <sup>2</sup>; A. Renwick <sup>1</sup>; D.A. Scott <sup>1, 2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Molecular Physiology and Biophysics, Baylor College of Medicine, Houston, Texas;; 3) In situ Hybridization Core, Jan and Duncan Neurological Research Institute. Texas Children's Hospital and Baylor College of Medicine, Houston, Texas.

---

### Background

*SOX7* is located on chromosome 8p23.1 in a critical region that is commonly deleted in individuals with complex cardiovascular malformations. *SOX7* is highly expressed in vascular endothelium and endocardium but its role in these cells remains unclear.

### Hypothesis

*SOX7*-deficiency may play a critical role in endothelial-to-mesenchymal transition (EMT) during the endocardial cushions (EC) formation, modulating the expression of *Wnt4* and *Bmp2*.

### Material and Methods

To explore the role of *SOX7* in cardiovascular development, we created standard and conditional *Sox7* knockout mice. Embryos were harvested from E8.5, to E15.5 for phenotypic, histology, immunohistochemistry, collagen gel, RNA-In Situ Hybridization (RNA-ISH), bioinformatics and RNA-seq analysis.

### Results

- *Sox7*<sup>-/-</sup> embryos die around E11.5 with signs of heart failure, pericardial edema and failure of vasculature remodeling.
- The same phenotype was observed when *Sox7* was ablated in endothelial cells using a Tie2-Cre. This suggests that *SOX7* plays a critical role in vasculogenesis.
- One out of two *Sox7*<sup>flox/flox</sup>;Tie2-Cre embryos that escaped early lethality had a ventricular septal defect.
- *Sox7*<sup>-/-</sup> embryos at E9.5 had hypocellular EC with severely reduced numbers of mesenchymal cells.
- In atrioventricular (AV) canal explant studies, *SOX7* deficiency leads to a severe reduction in EMT.
- *Wnt4*, *Pik3r3* and *Bmp2* transcript levels were downregulated in RNA-seq studies on E9.5 hearts harvested from *Sox7*<sup>-/-</sup> embryos. These genes have been shown to positively regulate EMT.
- RNA-ISH showed undetectable levels of *Wnt4* transcripts in the endocardium of *Sox7*<sup>-/-</sup> embryos and markedly reduced *Bmp2* expression in myocardium of null embryos.
- *SOX7* plays a dual role in the endocardium of the AV canal, separated from its role in endothelial progenitors.



- In Silico analysis suggests multiple SOX family binding sites exist in predicted promoters of *Wnt4* and *PiK3r3*.

### **Conclusions**

Since SOX7 is a transcription factor, we conclude that SOX7 promotes endocardial cushion development regulating the expression of *Wnt4*, which is then expressed in endocardium, and promotes EMT by acting in a paracrine manner to increase the expression of BMP2 in the contiguous myocardium. BMP2 subsequently, promote EMT to ensure the AV cushion formation. Conversely, decreased expression of SOX7 results in *Wnt4* and *Bmp2* downregulation and hypocellular EC, and may contribute to the congenital heart defects seen in individuals with recurrent 8p23.1 microdeletions.

## PgmNr 2724: *Ascc2* as a novel regulator of cardiac progenitor specification.

### Authors:

N. Li-Villarreal; L. Wong; V. Piazza; C. Hsu; M. Dickinson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Molecular physiology & biophysics, Baylor College of Medicine, Houston, Texas.

---

The cardiovascular system is the first functional organ system to form in vertebrate embryos and is essential for subsequent development. In humans, congenital heart defects are among the most common birth defects, often resulting in lethality. Congenital muscular dystrophies (CMDs) are one group of early onset congenital disorders that can present with cardiac and respiratory failure. Recently, human loss of function alleles of Activating signal co-integrator (ASC) complex members, TRIP4 and ASCC1 were shown to lead to CMD with heart and respiratory failure; indicating a novel role for the ASC complex in cardiovascular development and disease. The ASC complex consists of Trip4, *Activating signal cointegrator 1 complex subunit 1* (Ascc1), Ascc2, and Ascc3. This complex has been shown to interact with Serum response factor (SRF) and LIM domain protein, CSRP1, which regulates cardiac mesoderm migration. Here, we begin to dissect Ascc2 function during mouse embryogenesis. We found that *Ascc2* is expressed broadly in early post implantation embryos and like *in vitro*, interacts with other Asc complex members. Interestingly, *Ascc2* expression is only restricted anteriorly with highest expression in the developing heart at embryonic day (E) 8.5. Utilizing *Ascc2*<sup>Tm1d(KOMP)Wtsi</sup>, we have determined that a homozygous mutation in *Ascc2* results in lethality by E9.5. At E8.5, mutant embryos have under-developed hearts, and in some cases, complete absence of heart tube formation. Despite the disruption of cardiac development in *Ascc2* mutants, early cardiac mesoderm markers showed a significant increase in genes required for cardiac progenitor specification and migration, but no change in endothelial cell markers. *Nkx2.5* conditional knock out of *Ascc2* mice are subviable as adults with morphological and function cardiac defects. These data suggest that *Ascc2* is an essential regulator of early cardiac specification and is required for normal heart development.

## PgmNr 2725: Deletions in the *ANGIOTENSINOGEN* gene among preeclamptic/eclamptic women in Calabar, Nigeria.

### Authors:

M. Kooffreh <sup>1</sup>; M. Ekott <sup>2</sup>; C. Sunday <sup>1</sup>; E. Okhume <sup>1</sup>; M. Ozoje <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Biotechnology, University of Calabar, Calabar, Nigeria; 2) University of Calabar Teaching Hospital, Calabar, Nigeria; 3) Animal Science and Genetics Department, Federal University of Agriculture, Abeokuta, Nigeria

---

Preeclampsia is a disorder of human pregnancy characterized by the onset of high blood pressure and often a significant amount of protein in the urine. Eclampsia is characterized by the syndrome of severe preeclampsia with addition of convulsion, seizures, hemorrhage, and coma leading to fatal maternal and neonatal outcomes. The prevalence of pre-eclampsia/eclampsia varies greatly in different ethnic and geographical regions. Mutations of the *ANGIOTENSINOGEN* (*AGT*) gene have been implicated in the development of preeclampsia/eclampsia in some ethnic groups but no such studies have been reported in Calabar, Nigeria. This was a pilot study to investigate the mutations in *EXON 2* of the *AGT* gene among preeclamptic/eclamptic women in Calabar, Nigeria. A total of fifty-eight pregnant women comprising 19 pre-eclamptic, 19 eclamptic and 20 apparently healthy volunteer pregnant women as controls were recruited into the study from the University of Calabar Teaching Hospital, Calabar. Ethical approval and informed consent was obtained from the UCTH Ethics committee and the women respectively. Demographic information was obtained and analyzed using student t-test, 2- 3mls of blood was also collected from all the women. DNA was extracted and PCRs performed and the amplicons were sequenced for the *AGT* mutations. The chromatogram was decoded into nucleotide sequences using chromaPro. Multiple sequence alignment was performed for all the sequenced products. This mutation study identified a deletion of guanine at position 26 in 19(100%) pre-eclamptic women and 16(84.2%) eclamptic women, 3 other deletions 12 del A, 2-5 del CCTT, 2-6 del CCTTT were observed among the 19 eclamptic women. But these mutations were not observed among the control pregnant women in the study population.

**Keywords:** *ANGIOTENSINOGEN*, Preeclamptic/Eclamptic women, Deletions, Calabar.

## PgmNr 2726: Identification of novel variants and transcriptomic alterations in pediatric restrictive cardiomyopathy via exome and RNA sequencing.

### Authors:

A. Phatak <sup>1</sup>; C. Sucharov <sup>2,3</sup>; S. Miyamoto <sup>2,3</sup>; S. Ware <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Indiana University School of Medicine, Indianapolis, Indiana.; 2) University of Colorado Denver, CO; 3) Children's Hospital Colorado, CO; 4) Riley Hospital for Children Indianapolis, IN

---

Restrictive cardiomyopathy (RCM) is a rare myocardial disorder characterized by normal chamber size and thickness with impaired ventricular filling due to diastolic dysfunction. Pediatric RCM has limited therapeutic approaches leaving cardiac transplant as the primary treatment. Although pathogenic variants in sarcomeric genes can cause RCM, the underlying etiology is frequently unknown. To investigate causation and pathogenesis of RCM we performed exome, mRNA- and small RNA-sequencing (RNA-seq) analysis using explanted left ventricular heart tissue from pediatric RCM transplant subjects (n=14) and compared them to pediatric controls (n=4). From exome data, *BAG3* (p.Pro209Leu) and *FLNC* (p.Tyr2563Cys) pathogenic variants were identified in three RCM patients and a fourth patient had a 21bp in-frame insertion in *FLNC* in a domain important for targeting to the Z-line. Multiple novel or rare variants of uncertain significance (MAF <0.0001 and CADD scores ≥20) in cardiomyopathy genes were identified in remaining samples highlighting the genetic heterogeneity and opportunity for novel gene discovery. Using RNA-seq data, we identified 1922 differentially expressed genes (upregulated; n=820 and downregulated; n=1102) which predicted enrichment of ERK/MAPK signaling, p38 MAPK signaling, actin cytoskeleton and adherens junction signaling (IPA™) pathways in RCM. Because aberrant splicing occurs in cardiomyopathy, we analyzed alternative splicing and differential exon usage by rMATS and DEXseq computational tools. We observed sarcomere and cytoskeleton genes, cardiac stress signaling and spliceosomal components alternatively spliced or differentially expressed in RCM. Genes involved in mRNA splicing were also identified as differentially spliced. We therefore performed rMAPS analysis for 115 RNA binding proteins (RBPs) to identify RBP motifs enriched in observed splicing events in RCM and identified *SRSF2*, *SRSF7*, *DAZAP1*, *QKI*, *RBM5* and *RBM24* as likely mediators of observed dysregulated splicing in RCM. These findings suggest that RBP-mediated disruption of splicing is a common pathogenetic mechanism irrespective of underlying genetic cause in RCM. These findings indicate that despite genetic heterogeneity, similar patterns of transcriptome dysregulation and splicing abnormalities may characterize disease pathogenesis in RCM, suggesting an opportunity for development of therapy.

## PgmNr 2727: Absence of aortic disease in *Acta2*<sup>R149C/+</sup> mice is associated with defective release of mutant smooth muscle $\alpha$ -actin from the CCT complex.

### Authors:

K. Kaw<sup>1</sup>; J. Chen<sup>1</sup>; A. Chattopadhyay<sup>1</sup>; X.Y. Duan<sup>1</sup>; Z. Zhou<sup>1</sup>; C.S. Kwartler<sup>1</sup>; S. Ma<sup>1</sup>; K. Kamm<sup>2</sup>; J.T. Stull<sup>2</sup>; K. Trybus<sup>3</sup>; D.M. Milewicz<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Texas Health Science Center McGovern Medical School, Houston, Texas.; 2) Departments of Physiology, University of Texas Southwestern, Medical Center, Dallas, Texas; 3) Department of Molecular Physiology and Biophysics, University of Vermont, Burlington VT

**Rationale:** ACTA2 encodes the smooth muscle cell (SMC) specific isoform of actin, SM  $\alpha$ -actin. Mutations in ACTA2 predispose to various vascular diseases, including thoracic aortic aneurysms and dissections (TAAD), early onset coronary artery disease (CAD), and premature stroke. The most common mutation is ACTA2, p.Arg149Cys, which accounts for 25% of all ACTA2 mutations. ACTA2 p.Arg149Cys mutation carriers present with either TAAD or CAD but rarely have both diseases. The factors dictating whether a mutation carrier has TAAD or CAD are not known.

**Objective:** The purpose of this study is to identify the molecular mechanisms dictating whether an individual with ACTA2 p.Arg149Cys has TAAD or CAD.

**Methods and Results:** We used CRISPR/Cas9 technology to generate an *Acta2* R149C mutant mouse, *Acta2*<sup>R149C/+</sup>, in a C57BL6 background. Aortas from *Acta2*<sup>R149C/+</sup> mice have decreased contraction in response to agonist stimulation compared to wildtype (WT) mice. The mutant mice do not have premature deaths when followed for 48 months and do not have thoracic aortic aneurysms as assessed by echocardiography, even when fed a high salt diet and L-NG-nitroarginine methyl ester to increase blood pressure. When crossed into an *Apoe*<sup>-/-</sup> background and fed a high fat diet, the *Apoe*<sup>-/-</sup> *Acta2*<sup>R149C/+</sup> mouse has increased atherosclerotic plaque burden in the aorta ( $p < 0.01$ ). *In vitro* studies expressing WT and ACTA2 mutants in *Acta2*<sup>-/-</sup> smooth muscle cells (SMCs) found that no SM  $\alpha$ -actin was present when mutations associated with early onset CAD are expressed despite qPCR analyses showing the mutant transcript is present at the same level as the WT transcript. Two dimensional gel electrophoresis studies confirmed reduced levels of the mutant compared to WT SM  $\alpha$ -actin in the cytoplasm in *Acta2*<sup>R149C/+</sup> SMCs. Actins are folded in the chaperonin containing t-complex polypeptide (CCT) complex, raising the possibility that the mutant actin may be retained in this complex. Immunoprecipitation of the CCT complex using CCT1 antibody followed by immunoblot analyses for SM  $\alpha$ -actin demonstrated increased SM  $\alpha$ -actin associated with the CCT complex in the *Acta2*<sup>R149C/+</sup> SMCs.

**Conclusions:** These data indicate that the *Acta2*<sup>R149C/+</sup> mice do not develop aortic disease but rather have increased atherosclerotic burden when bred into a hyperlipidemic background. Our data suggests that decreased levels of mutant versus WT SM  $\alpha$ -actin in the cytoplasm may be responsible for the lack of aortic disease in this mouse model.

## PgmNr 2728: Mutations in the prostaglandin I2 receptor gene as potential causes of fibromuscular dysplasia.

### Authors:

A. Georges <sup>1</sup>; J. Albuissou <sup>2</sup>; D. Dupré <sup>1</sup>; V. D'Escamard <sup>4</sup>; J. Kovacic <sup>4</sup>; A. Januszewicz <sup>5</sup>; P.-F. Plouin <sup>3</sup>; X. Jeunemaitre <sup>1,2</sup>; N. Bouatia-Naji <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Université de Paris, PARCC, INSERM, Paris, FRANCE; 2) AP-HP, Hôpital Européen Georges Pompidou, Département de Génétique, Centre de Référence des Maladies Vasculaires Rares, Paris, France; 3) AP-HP, Hôpital Européen Georges Pompidou, Hypertension Unit, Paris, France; 4) Zena and Michael A Wiener Cardiovascular Institute and Marie-Josée and Henry R. Kravis Center for Cardiovascular Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 5) Department of Hypertension, Institute of Cardiology, Warsaw, Poland

**Background:** Fibromuscular Dysplasia (FMD) is a vascular disease characterized by non-atherosclerotic stenosis of small and medium-size arteries. FMD predominantly affects early middle-aged women and is independent of typical cardiovascular risk factors, such as age and dyslipidemia. Observation of FMD familial cases and clinical similarity with connective tissue disorders suggest inherited or spontaneous protein-coding mutations could cause FMD. However, previous re-sequencing of candidate genes was inconclusive.

**Methods and results:** We performed exome sequencing in 30 individuals including familial cases (N=21 sibs), 2 trios (N=6) and sporadic cases (N=3). Filtering for shared loss-of-function rare mutations (LoFs) among at least 2 families was negative, as was the trios filtering. We curated the results for predicted LoFs shared among sibs from the same family in genes with documented roles in vascular and smooth muscle cell biology. We identified in two sisters a shared rare LoF (rs199560500, p.Gln163Ter) in the sequence of prostaglandin I2 receptor gene (*PTGIR*), a key player in pulmonary arterial hypertension and vascular remodeling. Targeted amplicon-NGS sequencing of *PTGIR* and 19 functionally connected genes in 384 unrelated FMD patients revealed 1 additional LoF in *PTGIR* (rs199560500). Sanger sequencing of *PTGIR* in a total of 1,011 FMD patients identified *PTGIR* nonsense variants in 3 additional patients (1 with rs199560500, p.Gln163Ter, 2 with rs754755149, p.Pro17ArgfsTer6). We estimate the frequency of *PTGIR* LoF alleles to 0.25% in our FMD cohort, compared to a 0.048% estimate from Non-Finnish European populations (gnomAD database). We also identified six uncharacterized missense variants, each in one patient.

Using transient overexpression in HEK293 (human epithelial kidney) cells, we tested the ability of mutant proteins to increase cAMP concentration in response to iloprost, a prostacyclin analog. We confirmed that rs199560500 and rs754755149 nonsense variants result in fully non-functional *PTGIR* protein. All missense variants were sensitive to iloprost and mostly indistinguishable from wild-type *PTGIR*. Western Blot showed a reduced protein expression of two of the variants (rs1397542892, p.Leu67Pro and rs199969416, p.Arg137Cys), suggesting a moderate hypomorphic phenotype.

**Conclusions:** Our study shows that rare *PTGIR* mutations are a putative cause of FMD and should be taken into consideration for further clinical and biological studies.

## PgmNr 2729: High-throughput patch clamping identifies dozens of novel Brugada syndrome-associated *SCN5A* variants.

### Authors:

A. Glazer; Y. Wada; L. Hall; D. Roden

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt University Medical Center, Nashville, Tennessee.

---

The accurate classification of variants in Mendelian disease genes is a major challenge for the successful implementation of personalized medicine. Variants in *SCN5A* are a major cause of inherited arrhythmia syndromes, including partial or complete loss of function *SCN5A* variants, which can cause Brugada Syndrome. The pathogenicity of *SCN5A* variants is often unknown or disputed; indeed, most observed *SCN5A* variants are classified as a variant of uncertain significance (VUS). The widely adopted American College of Medical Genetics and Genomics (ACMG) classification scheme includes *in vitro* functional data as a strong criterion for variant classification. Our aim was to determine the *in vitro* function of dozens of variants in *SCN5A* to improve variant classification. After a literature review of 1,712 reported *SCN5A* variants, we selected 58 variants that had not been functionally characterized (10 suspected benign and 48 suspected Brugada-associated). As a control, we also selected 12 well-characterized variants that had a range of functional defects. We generated each variant with Quikchange mutagenesis and stably expressed each variant in HEK293 cells. For each cell line, we used high-throughput automated patch clamping to measure 5 important functional parameters: peak current, late current, activation and inactivation kinetics, and recovery from inactivation. All 12 previously studied variants had functional properties similar to previously published manual patch clamp results. 9 of the 10 suspected benign variants had wildtype-like channel function. 34 of the 48 suspected Brugada-associated variants (71%) had partial or total loss of channel function. Suspected Brugada Syndrome variants that had been observed in more patients were more likely to result in loss of channel function ( $p < 0.01$ ). Applying ACMG classification criteria, 48 of 58 previously unstudied variants were reclassified with the patch clamp data, including the reclassification of 27 variants from VUS to likely pathogenic and 9 variants from VUS to benign. These results demonstrate that high-throughput *in vitro* functional characterization can identify dozens of novel *SCN5A* variants with loss of function features consistent with Brugada Syndrome. This method can reclassify *SCN5A* variants at scale and help “de-VUS” this important gene.

## PgmNr 2730: PACSIN3 is a modifier gene for hypertrophic cardiomyopathy.

### Authors:

M.E. Lindholm; H. Zhu; D. Waggott; S. Abramowitz; Y. Huang; E.A. Ashley; M.T. Wheeler

[View Session](#) [Add to Schedule](#)

**Affiliation:** Cardiovascular medicine, Stanford University, Palo Alto, California.

---

Hypertrophic cardiomyopathy (HCM) is characterized by increased chamber size and reduced ejection fraction. Inherited HCM is commonly associated with mutations in sarcomeric genes, including the gene for  $\beta$ -cardiac myosin (MYH7). High variability in disease presentation and penetrance in patients with MYH7 mutations has caused some to question the single variant nature of this disease. We previously identified 96 genes with potential for modifying the HCM phenotype through a combination of MYH7 patient genome, public transcriptome, and HCM model system RNAseq analysis.

One novel potential modifier we identified is PACSIN3, which has not been previously associated with cardiac disease. PACSINs are intracellular adapter proteins that are involved in endocytosis and actin reorganization, and PACSIN3 is the muscle-enriched isoform.

To investigate the role of PACSIN3 in HCM, we analyzed PACSIN3 expression in septal myectomy tissue of a separate cohort of 39 HCM patients. PACSIN3 expression was significantly higher in HCM patients compared to 10 healthy control septae ( $p < 0.01$ ). NPPB, a marker for cellular hypertrophy, was also higher in HCM ( $p < 0.02$ ), as expected. RNA sequencing analysis of PACSIN3 knockdown (KD) in neonatal rat ventricular myocytes (NRVMs) showed downregulation of the oxidative phosphorylation pathway ( $N=4$ ,  $FDR < 1e-4$ ). PACSIN3 KD led to higher NPPB expression, indicating hypertrophy, an effect that was exacerbated by phenylephrine treatment (PE). Immunofluorescent imaging also showed that PACSIN3 KD in combination with PE increased NRVM cell size compared to PE alone. In epithelial cells, PACSIN3 was shown to be required for ectodomain shedding of HB-EGF, which stimulates cellular hypertrophy through activation of the EGF receptor. Treating NRVMs with HB-EGF in combination with PACSIN3 KD increased NPPB expression to a similar extent as control, likely bypassing the effect of PACSIN3 KD.

In conclusion, we identified PACSIN3 as a novel modifier of cardiac hypertrophy. PACSIN3 is elevated in cardiac tissue of HCM patients, while reduced levels in cardiac myocytes leads to higher NPPB expression, indicative of acute induction of hypertrophy, an effect that is further exacerbated by a hypertrophic stimulus. Induction of PACSIN3 in HCM may present a compensatory protective mechanism.



## PgmNr 2731: A deep mutational scan of *KCNE1* via a high-throughput cell-surface trafficking assay can identify functionally deleterious variants.

### Authors:

A. Muhammad <sup>1,2</sup>; A.M. Glazer <sup>2,3</sup>; K.A. Matreyek <sup>4</sup>; M.A. Blair <sup>3</sup>; L. Short <sup>3</sup>; D.M. Fowler <sup>4</sup>; D.M. Roden <sup>2,3,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Medical-Scientist Training Program, Vanderbilt University Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Genome Sciences, University of Washington, Seattle, WA; 5) Department of Pharmacology, Vanderbilt University Medical Center, Nashville, TN; 6) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN

---

With the increase in clinical genetic testing, precision medicine faces a new challenge: rapid discovery of VUSs requires accurately classifying Mendelian disease gene variants at scale. One such gene is *KCNE1*, which encodes an accessory subunit for  $I_{Ks}$ , a potassium channel important for cardiac repolarization. Loss of function (LoF) variants in *KCNE1* can lead to type 5 long QT syndrome, a disorder associated with susceptibility to potentially fatal arrhythmias. Previous data suggest that LoF mutations in *KCNE1* either disrupt interactions with other  $I_{Ks}$  subunits, encoded by *KCNQ1*, or reduce trafficking of  $I_{Ks}$  channels to the cell surface. In this study, we conducted a deep mutational scan of *KCNE1* by generating all possible amino acid variants in the 127-amino acid long protein and assessed their trafficking in a high-throughput, multiplexed fashion. We generated a comprehensive library of coding variants in *KCNE1*, expressed them in HEK-293 cells engineered to incorporate a single variant per cell, and then tested this library with a flow-cytometry based assay to identify variants that disrupt *KCNE1* cell surface trafficking. The library tested contained 59.4% of the 2540 possible variants, and showed a bimodal distribution of cell surface *KCNE1* staining, with peaks of the distribution corresponding to wild-type and a known trafficking-deficient variant, respectively. This work successfully deployed a high-throughput trafficking assay to comprehensively interrogate large numbers of protein variants for a clinically important ion channel subunit. Our results will provide the functional information needed to supplement the American College of Medical Genetics and Genomics functional assay criterion by identifying pathogenic *KCNE1* variants, and facilitate clinical interpretation of novel variants and VUSs.

# PgmNr 2732: Cardiac $\beta$ -myosin heavy chain ( $\beta$ -MYH7) head motor domain mutations and understanding their mechanisms leading to cardiomyopathies: a structural point of view.

## Authors:

D. Rani; K. Thangaraj

[View Session](#)
[Add to Schedule](#)

**Affiliation:** Human Genetics, CCMB, UPPAL ROAD,, Hyderabad, Telangana, India

---

## Abstract

### Cardiac $\beta$ -myosin heavy chain ( $\beta$ -MYH7) head motor domain mutations and understanding their mechanisms leading to cardiomyopathies: a structural point of view

**Deepa Selvi Rani and K Thangaraj**

Mutations in cardiac  $\beta$ -myosin heavy chain gene ( $\beta$ -MYH7) are the major cause of inherited cardiomyopathies and sudden cardiac deaths (SCD), yet the molecular mechanisms for the disease have not been fully understood. Cardiac head motor domain of  $\beta$ -MYH7 is a fundamental force-generating element for muscle contraction. However, not many studies on  $\beta$ -MYH7 are carried out in Indian hypertrophic (HCM) and dilated (DCM) cardiomyopathies. Therefore, we sequenced the  $\beta$ -MYH7 in 101 HCM and 147 DCM against 207 controls and found 45 variations (including 29 novel), of which 8 missense mutations were detected in myosin head motor domain, besides 3 splice-site variations, 3 frame-shift mutations and 5 allelic-heterogeneities. When compared this data with our previous data set, we identified a rare digenic mutations in  $\beta$ -MYH7 and *TNNT2* genes, and a compound mutation in  $\beta$ -MYH7, *TMP1* and *ACTC* genes. In addition, we performed PCA for population stratification using a panel of 47 ancestry-informative markers along with Hap Map data sets to show that the cases and controls are from same ethnic background. Further, homology models were built for 8 missense mutations observed in the head motor domain of  $\beta$ -MYH7, and compared each with native structure by using several structural parameters. We found that each homology model perturbs the structure differently, and followed a unique mechanistic pathway. Interestingly, of the 8 homology models analyzed, three mutants (p.Leu510, p.His723 and p.Met431) showed comparatively high RMS deviations of  $\sim 3.4$ ,  $\sim 3.8$ , and  $\sim 3.9$  respectively. Non-bonding interactions showed that these deviations might be due to an additional aberrant hydrophobic interaction within the hydrophobic core (Leu510 with Trp508; His723 with Pro727; Met431 with Leu352) and that led to drastic deviations in their structures. Therefore, structural deviations may partake in cardiac remodeling and may alter the morphology of heart into HCM/DCM. Certainly, identifying the causative mutations and their structural mechanisms would help for pharmacogenomic studies with special emphasis laid on personalized medicine.

## PgmNr 2733: Towards variant effect mapping of coronary artery disease genes.

### Authors:

D. Tabet<sup>1, 2, 3</sup>; A.G. Cote<sup>1, 2, 3</sup>; R. Jiang<sup>1, 2, 3</sup>; J. Knapp<sup>1, 2, 3</sup>; N. Kishore<sup>1, 2, 3</sup>; J. Weile<sup>1, 2, 3, 4</sup>; S. Sun<sup>1, 2, 3, 4, 5</sup>; M. Marinella<sup>1, 2, 3</sup>; Y. Wu<sup>1, 2, 3, 4</sup>; H. Ennajdaoui<sup>1, 2, 3</sup>; I. Fotiadou<sup>1, 2, 3</sup>; D.M. Fowler<sup>6, 7</sup>; F.P. Roth<sup>1, 2, 3, 4, 8, 9</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; 2) The Donnelly Centre, University of Toronto, Toronto, ON, Canada; 3) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 4) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 5) Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden; 6) Department of Genome Sciences, University of Washington, Seattle, WA, USA; 7) Fred Hutchinson Research Center, Seattle, WA, USA; 8) Canadian Institute for Advanced Research, Toronto, ON, Canada; 9) Corresponding author

In the genomics era, the diagnosis and treatment of disease is becoming increasingly individualized based on a person's unique genetic signature. However, understanding the effects of rare variation, a significant contributor to human disease, remains a largely unresolved issue. Over 100,000 exomes have been gathered in GnomAD and over 4 million coding variants have been discovered to date, ~99% of which are rare. Yet, of these, just 2% have a clinical interpretation in ClinVar, the bulk of which are interpreted as variants of unknown significance. As such, we aim to develop and apply high-throughput sub-functional assays that couple a genotype to a selectable phenotype, thereby enabling the analysis of all possible coding sequence variations on a protein's function. To this end, our group employs a framework termed Deep Mutational Scanning (Matreyek *et al.*, 2017) to interrogate the function of all possible coding variation in a disease gene, thereby allowing us to understand the effects of rare, often novel, genetic variation. I am applying this method to study genes related to Coronary Artery Disease (CAD), the prevailing cause of mortality worldwide.

To do so, an input library of random codon-mutagenized coding sequences is subjected to a sub-functional assay to test the effects of each variant on protein function. The resulting post-selection library is deep sequenced using TileSeq and unmeasured variant effects are imputed using machine learning. My current focus is the study of the CHD-related genes LDLR, LDLRAP1, LRP6 and PCSK9, which function in the regulation of circulating low density lipoprotein (LDL) levels, an important marker of CAD risk. To generate variant effect maps for these genes, LDL uptake assays are adapted to high-throughput fluorescence activated cell sorting via a fluorescently labelled LDL molecule. Following CRISPR mediated knockout of the relevant gene, we observed a measurable decrease in LDL uptake, which is to be rescued by the integration of the wild-type sequence into an engineered landing pad constitutively expressed in the cell's genome. Variations of this method will be expanded to screen thousands of variants *en masse* for a number of CAD-related genes, the final maps of which can then be used to interpret the effects of previously unknown variation, potentially improving diagnosis and treatment of CAD.

## PgmNr 2734: Contribution of CPS1 gene to atherosclerosis in mice and humans.

### Authors:

N.C. Woodward <sup>1</sup>; J. Gukasyan <sup>1</sup>; P. Huang <sup>1</sup>; W.S. Schwartzman <sup>1</sup>; Q. Jia <sup>1</sup>; Y. Han <sup>1</sup>; R. Zhu <sup>1</sup>; W.H.W. Tang <sup>2</sup>; Z. Wang <sup>2</sup>; S.L. Hazen <sup>2</sup>; W.J. Mack <sup>1</sup>; H.N. Hodis <sup>3</sup>; J.A. Hartiala <sup>1</sup>; H. Allayee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Preventative Medicine, University of Southern California, Los Angeles, CA; 2) Department of Cellular and Molecular Medicine, Cleveland Clinic, Cleveland, OH; 3) Keck School of Medicine, University of Southern California, Los Angeles, CA

---

We recently identified a variant (rs715) in the gene encoding carbamoyl phosphatase 1 (CPS1), the first committed and rate-limiting step of the urea cycle, as demonstrating sexually dimorphic associations with multiple metabolites in the pathway leading from choline to urea, particularly increased circulating glycine levels in women. Most notably, rs715 showed a striking female-specific association with decreased risk of coronary artery disease (CAD). To replicate and further investigate these findings, we examined the association of rs715 with CAD in UK Biobank and subclinical atherosclerosis in women participating in five clinical trials (ELITE, EPAT, VEAPS, WELLHART, and WISH), as well as through genetic and dietary perturbations in mouse models. We corroborated the female-specific association of rs715 with CAD in UK Biobank (OR=0.960, SE=0.016, p=0.014). In a meta-analysis of women from the five clinical trials, rs715 was associated with significantly reduced carotid artery intima-media thickness (CIMT) in a linear regression using log-transformed CIMT (beta=-0.046, SE=0.011, p=1.6x10<sup>-5</sup>, n=1430). In a mouse model of CPS1 deficiency, heterozygous knockout (CPS1<sup>+/-</sup>) mice on an AAV-PCSK9-induced hyperlipidemic background exhibited sex-specific attenuation of atherosclerosis. CPS1<sup>+/-</sup> mice had 25% reduced lesion formation (p=0.0061, n=15-18) in the aortic root and *en face* stained aorta (p=0.0118, n=4-6), and 40% reduced necrotic core area (p=0.0007, n=15-20), but these effects were primarily observed in male. As a complementary approach, we placed ApoE<sup>-/-</sup> mice on a 2.5% glycine-supplemented diet or a control diet (0.3% glycine) for 16 weeks. Glycine supplementation had no effect on aortic lesions, suggesting the cardioprotective association of decreased CPS1 activity is not due to glycine. However, our results provide further evidence for the species-sex-specific importance of CPS1 in atherosclerosis but the causal cardioprotective metabolite or biological mechanism(s) remains to be determined.

## PgmNr 2735: Gene-environment interactions in iPSC-derived cardiomyocytes.

### Authors:

A.S. Findley <sup>1</sup>; A.L. Richards <sup>1</sup>; K. Rhodes <sup>2</sup>; M.C. Ward <sup>2</sup>; C.A. Kalita <sup>1</sup>; A. Alazizi <sup>1</sup>; X. Wen <sup>3</sup>; D.E. Lanfear <sup>4</sup>; R. Pique-Regi <sup>1</sup>; Y. Gilad <sup>2</sup>; F. Luca <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan.; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 4) Center for Individualized and Genomic Medicine Research, Henry Ford Hospital, Detroit, MI

---

Cardiomyopathy is a disease of the heart muscle that leads to arrhythmias and heart failure. As the number of people diagnosed with heart failure increases, it is imperative to gain a better understanding of the genetic and environmental factors leading to cardiomyopathy. We have established a cell culture model to explore gene-environment interactions (GxE) in cardiomyocytes (CMs). We have reprogrammed lymphoblastoid cell lines (LCLs) from 6 individuals into induced pluripotent stem cells (iPSCs), which have been further differentiated into CMs. A key feature of this system is the ability to expose these cell types to the same panel of treatments to investigate cell type specific response and treatment-by-cell type interactions. To this end, we have exposed all cell lines from the 3 cell types to a panel of 28 treatments, including hormones, vitamins, environmental contaminants, and common drugs. Twelve treatments induced a significant response for hundreds of genes in at least one cell type. We identified 4,835 genes expressed in all cell types and that show evidence of treatment-cell type interactions as measured by likelihood ratio test in DESeq2 (FDR < 10%). In particular, 96% of differentially expressed genes in CMs in response to treatment with insulin are not differentially expressed in LCLs or iPSCs. These unique insulin-CM genes are enriched for cholesterol and ADP metabolic processes. To dissect the sources of gene expression variation across individuals, we used a random effects model accounting for interactions with cell-type and treatment, while taking advantage of experimental replication and accounting for other technical variables. We estimate that 64% of interindividual variation is explained by interactions with cell-type, while 34% of variation is due to interactions with treatments (potential GxE). To identify SNPs which could be responsible for these interactions in the CMs, we performed allele-specific expression (ASE) analysis. We identified 3,933 unique SNPs across all cell types displaying ASE, corresponding to 1,993 genes. When contrasting each treatment to the appropriate control, we identified 62 instances of conditional ASE (cASE) at 43 unique SNPs in 38 genes (FDR < 20%). One of these genes is *TSPAN17*, which plays a role in cardiovascular development. This and other examples show that genetic interactions with environmental exposures (treatments) have a strong contribution to inter-individual variation in health and disease.

## PgmNr 2736: MPRA-based *cis*-regulatory variant screen of QT interval GWAS variants in cardiac open chromatin elements.

### Authors:

A. Kapoor <sup>1</sup>; D. Lee <sup>2</sup>; A. Chakravarti <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX 77030; 2) Center for Human Genetics and Genomics, New York University School of Medicine, New York, NY 10016

---

Sequence variation in multiple *cis*-regulatory elements (CREs) modulating a target gene's expression is the leading hypothesis for GWAS signals in the noncoding genome. However, we require convincing experimental data to identify these causal *cis*-regulatory variants. Here, we perform a massively-parallel reporter assay (MPRA)-based screen to identify association-based CRE sequence variants at all QT interval GWAS loci. All genome-wide significant variants from the QTIGC study and common variants from the 1000 Genomes Project (EUR MAF>1%) in high LD ( $r^2>0.9$ ) with QT GWAS sentinel variants were selected ( $n=4,647$ ). These were further filtered by overlap with potential cardiac CREs for assay design ( $n=927$ ). As a control, we added all variants at the *SCN5A* GWAS locus regardless of CRE overlap ( $n=138$ ). 200-base long oligos with a 20-base 5' flank, variant-centered 129-base target sequence, 18-base spacer with EcoRI and SbfI site, 13-base unique barcode and a 20-base 3' flank were designed. Barcode sequences were designed to contain all four nucleotides and GC-content of 0.4-0.6, and further avoid microRNA seed binding sites,  $\geq 3$ -base homopolymers, the two restriction sites, and to have a minimum edit distance of two between any pair of barcodes. All sequences with additional EcoRI and SbfI sites were discarded. The final design included 1,018 variants with both alleles, each with 50 different barcodes for a total of 101,800 oligos across two equally-sized pools. Each oligo pool was cloned into lentiMPRA vector, pLS-mP, lacking the minP-eGFP reporter cassette, to generate a pre-reporter library from over 5 million clones to preserve library complexity. Paired-end sequencing was performed to check the quality of synthesized oligos and representation of individual barcodes. The minP-eGFP cassette was then cloned into the EcoRI-SbfI site of the two pre-reporter libraries to generate the final reporter plasmid libraries from over 5 million clones each. The two libraries are transfected into the mouse cardiomyocyte cell line HL1 in 10 parallel replicates, each with 5 million cells and cells harvested two days post-transfection to generate input plasmid- and reporter cDNA-based Illumina libraries to sequence the barcodes and assess allele-specific reporter activity. We expect our MPRA screen to not only identify causal variants underlying QT interval GWAS signals but also to serve as an experimental validation of putative CRE and predicted variant effect on CRE function.

## PgmNr 2737: Investigating clinical, socio-demographic, and genetic associations of 26 targeted kidney dysfunction-related variants among sickle cell patients from Cameroon.

### Authors:

V.J. Ngo Bitoungui <sup>1,2</sup>; S. Belinga <sup>3</sup>; T.M. Khuthala <sup>2</sup>; T. Masekoameng <sup>2</sup>; V. Nembaware <sup>2</sup>; F. Ngo Sack <sup>4,5</sup>; C.B. Chetcha <sup>6</sup>; G.K. Mazandu <sup>2,7</sup>; A. Wonkam <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biochemistry / West African Centre for Cell Biology of Infectious Pathogens, University of Ghana, Legon; 2) Pathology / Human genetics, University of Cape Town, Cape Town, Western Cape, South Africa; 3) Centre Pasteur of Cameroon, Yaoundé, Cameroon; 4) Clinical Biology / School of Health Sciences, Catholic University of central Africa, Yaoundé, Cameroon; 5) Hematology/ Faculty of Medicine and Pharmaceutical Sciences, University of Douala, Cameroon; 6) Hematology/ Faculty of Medicine and Biomedical Sciences, University of Yaoundé? 1, Cameroon; 7) 7African Institute for Mathematical Sciences, Muizenberg, Cape Town 7945, South Africa

---

**Background:** Renal failure is one of the leading causes of morbidity and mortality among sickle cell disease (SCD) patients. Several genetic factors are associated to renal failure, however there are limited studies in this field particularly for SCD patients in Africa.

**Objective:** To investigate the prevalence of 26 genetic variants, clinical and socio-demographic variables associated with kidney dysfunction for a group of SCD patients in Cameroon.

**Methods:** SCD patients were recruited in Yaoundé and Douala main hospitals. Socio-demographic and clinical features, anthropometric variables, blood pressure, hematological indices, creatinine, and crude albuminuria were measured. Estimated glomerular filtration rates (eGFR) were calculated using the CKD-EPI-creatinine and Schwartz formula. Kidney dysfunction was defined as an eGFR < 90 ml/min per 1.73 m<sup>2</sup>. Participants were genotyped for 26 kidney dysfunction-related gene-variants. Linear regression frameworks were designed in R to elucidate associations between gene variants and eGFR / microalbuminuria.

**Results:** A total of 413 SCD steady-state patients were studied. The median age was 15.5 years range (2 – 58) with 51% male. Sixty-one percent of these patients had micro-albuminuria, 2.4% proteinuria, 36% with glomerular hyperfiltration and 28% renal dysfunctions. Twenty-seven percent had both micro-albuminuria and kidney dysfunctions. In a simple regression model, considering only genetic factors, 22 and 23 variants out of 26 were significantly associated with micro-albuminuria and eGFR respectively. However, in a multi-variable model when combining clinical and socio-demographic factors with genetic variants, only the *Signal Induced Proliferation Associated 1 Like 3* (*SIPA1L3-rs11666497*) and *Uromodulin* gene variant (*UMOD-rs4293393*) were found to be significantly associated with eGFR with P-values of 0.04359 and 0.04091, respectively.

**Conclusion:** These results confirmed a high proportion of micro-albuminuria, among Cameroonian children and adolescents living with SCD. The study identified *rs4293393* in *Uromodulin* and



*rs11666497* in *SIPA1L3*, as key risk factors for chronic kidney diseases in SCD patients. These results demonstrate the need for further research to evaluate the feasibility of using multifactorial models (genetic, clinical and sociodemographic) for kidney dysfunction risk assessment in African populations.



## PgmNr 2738: Novel *de novo* *CUL3* mutation results in altered function of Cullin-RING E3 ubiquitin ligase and associated with a rare form of hypertension.

### Authors:

P. Sharma <sup>1,3</sup>; H.E. Chatrathi <sup>1,3</sup>; L. Wolfe <sup>1</sup>; C.C. Lau <sup>1</sup>; B.N. Pusey <sup>1</sup>; W.A. Gahl <sup>1,2</sup>; D.R. Adams <sup>1,2</sup>; M.C.V. Malicdan <sup>1,2</sup>; Undiagnosed Diseases Network

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NIH Undiagnosed Diseases Program, Common Fund, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 2) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 3) These authors contributed equally to this work

---

Hypertension is a principal risk factor for cardiovascular diseases such as heart failure and stroke. Pseudohypoaldosteronism type II (OMIM 614496, PHA2E) is a rare Mendelian form of hypertension caused by the imbalance between renal salt reabsorption and K<sup>+</sup> and H<sup>+</sup> excretion resulting in hyperkalemia and metabolic acidosis. Rare mutations in *CUL3* have been associated with PHA2E. *CUL3* (Cullin-3), a critical subunit of the multi-subunit CUL3-Ring-Ligase (CRL3) ubiquitin ligase complex, acts as a scaffold protein. It recruits ubiquitin-loaded E2 (Ub-E2) conjugating enzyme and adaptors in order to facilitate ubiquitination of substrates including With-no-lysine (K) kinases (WNKs), which control potassium homeostasis and blood pressure by regulating the activity of Na<sup>+</sup>Cl<sup>-</sup> cotransporter (NCC) along the distal convoluted tubule of the kidney. The activation of CRL3 to ubiquitinate substrates requires flexibility in the structure, which is introduced by the covalent attachment of Nedd8 (neddylation). CRL3 activation is readily reversed by the action of the COP9 signalosome (CSN) complex. Here we report a child with a novel *de novo* heterozygous mutation (NM\_001257198.1: c.1435\_1446del; p.F480\_M483del) in *CUL3* who presented with elevated serum potassium without renal tubular acidosis, short stature, global developmental delay, and frequent pulmonary and gastrointestinal infections. Extensive clinical and biochemical phenotyping also revealed glycosylation abnormalities. Functional analysis utilizing patient-derived dermal fibroblasts demonstrated increased neddylation and reduced levels of *CUL3* protein. Transient over-expression analysis in human embryonic kidney (HEK) cells suggested altered recycling of mutant *CUL3*, mediated by the catalytically active subunit of CSN. Consequently, the ubiquitination of WNK1 is increased, resulting to reduced steady-state levels. We thus hypothesize that pF480\_M483del mutation affects the function of CRL complex leading to altered ubiquitination of substrates. These findings support the role of the pF480\_M483del mutation in *CUL3* in hyperkalemia and hypertension.

## PgmNr 2739: 100,000 Genomes Project identification and validation of a novel pathogenic variant in *GDF2* (BMP9) responsible for hereditary hemorrhagic telangiectasia and pulmonary arteriovenous malformations.

### Authors:

S. Balachandar <sup>1</sup>; A. Shimonty <sup>1</sup>; M. Sroya <sup>2</sup>; M. Mohammed <sup>2</sup>; M. Alikian <sup>2</sup>; E. Curetean <sup>2</sup>; G. Rose <sup>2</sup>; E. Thomas <sup>2</sup>; J. Kilner <sup>3</sup>; K. Kerr <sup>4</sup>; V. McConnell <sup>4</sup>; S. McKee <sup>4</sup>; F. Boardman-Pretty <sup>5</sup>; A. Devereau <sup>5</sup>; E. McDonagh <sup>5</sup>; R. Scott <sup>5</sup>; T. Fowler <sup>5</sup>; M. Caulfield <sup>5</sup>; E.W. Alton <sup>6</sup>; T. Ferguson <sup>2</sup>; J. Redhead <sup>2</sup>; A.J. McKnight <sup>3</sup>; G.A. Thomas <sup>6</sup>; M.A. Aldred <sup>1</sup>; C.L. Shovlin <sup>6</sup>; Genomics England and Respiratory GeCIP

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Indiana University School of Medicine, Indianapolis, Indiana.; 2) Imperial College Healthcare NHS Trust, London, UK; 3) Queen's University of Belfast, UK; 4) Belfast Health and Social Care Trust, UK; 5) Genomics England, UK; 6) Imperial College London, UK

---

Hereditary hemorrhagic telangiectasia (HHT) is a multisystemic vascular dysplasia inherited as an autosomal dominant trait, and characterized by arteriovenous malformations (AVMs), mucocutaneous telangiectasia and nosebleeds. HHT is caused by a single pathogenic null allele in *ACVRL1*, *ENG*, or *SMAD4*. These genes encode proteins mediating bone morphogenetic protein (BMP)9 signaling. Although three variants identified in the *GDF2* gene encoding BMP9 were reported to cause a vascular anomaly syndrome similar to HHT, none meet ACMG-AMP Guideline criteria for pathogenicity, and no *GDF2* variants have been identified to date in *ENG/ACVRL1/SMAD4*-negative HHT families across the VASCERN HHT European Reference Network. Genomics England protocols were followed for the nomination of HHT, generation of a data model based on human phenotype ontology terms, development of a PanelApp HHT gene panel, and recruitment of HHT families from NHS Genomic Medicine Centres across the UK. Whole genome sequencing and tiering protocols identified a novel, heterozygous *GDF2* sequence variant in all three affected members of an HHT family. All had nosebleeds and typical HHT telangiectasia, and one (with severe pulmonary AVMs) had previously screened negative for *ENG*, *ACVRL1* and *SMAD4*. No other *GDF2* variants were identified in more than 100 *ENG/ACVRL1/SMAD4* negative probands with HHT or pulmonary AVMs. As part of the Respiratory GeCIP (Clinical Interpretation Partnership) subdomain established specifically to analyse HHT WGS data, the variant was introduced into a human BMP9 expression construct by site-directed mutagenesis and transfected into 293T cells. Conditioned media was collected for ELISA and western blotting. BMP9 is synthesized as a pro-protein that then dimerizes and is cleaved to produce the mature secreted protein. Western blotting showed the mutant construct expressed the pro-protein but lacked mature BMP9 dimer, suggesting the mutation disrupts correct cleavage of the protein. ELISA confirmed that mature BMP9 was undetectable in the conditioned media from the mutant construct compared to WT. Furthermore, conditioned media from the mutant construct failed to upregulate BMP-induced target genes in cultured endothelial cells. Analysis of circulating BMP9 levels in plasma from affected and unaffected family members is currently underway. In conclusion, we propose that heterozygous inheritance of this *GDF2* variant is a rare cause of HHT associated with pulmonary AVMs.

## PgmNr 2740: Mitonuclear discordance in 22q11 deletion syndrome admixed patients.

**Authors:**

B. Rebolledo-Jaramillo; G. Encina; A. Cuiza; G. Repetto

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Genetics and Genomics, Universidad del Desarrollo, Santiago, Chile

---

Chromosome 22q11.2 deletion syndrome (22q11DS) has an incidence of 1 in 4,000 and it accounts for ~2% of congenital heart disease (CHD) cases. Most patients share the same deletion, but show wide variation in phenotypic manifestations. For example, CHD is present in 50-60% of 22q11DS cases. There are six nuclear-encoded mitochondrial genes within the deleted region. Mitochondria have their own genome (mtDNA) containing only 37 genes. The remaining mitochondrial genes are nuclear-encoded, prompting the cell to coordinate both genomes to guarantee proper function. Differences in the genetic background for the mtDNA and nuclear genes (mitonuclear discordance, MND), obtained from crossing model organisms, have been shown to impact mitochondrial function. However, whether these incompatibilities naturally exist and contribute to phenotypic variation in humans is currently unknown. Admixed latinos present a great opportunity to study MND, since the mtDNA is predominantly of Native American (NAT) origin, and their nuclear genome reflects a gradient of Native American, European (EUR) and African (AFR) components.

We are interested in the genetic determinants of CHD in 22q11DS patients, so we evaluated the contribution of MND to the cardiac phenotype. We compared 145 22q11DS patients genotyped with the Affymetrix SNP 6.0 platform: 81 cases of CHD and 64 controls. We calculated the mtDNA haplogroup and the global Native American, European, and African ancestry of each patient. We defined global MND as the proportion of the nuclear ancestry not represented by the mtDNA ancestry. We found no statistical evidence to relate Native American haplogroups or global mitonuclear discordance to CHD in our cohort. We concluded that CHD might not be related to global mitochondrial genomic variation in 22q11DS admixed Chileans.

Funded by FONDECYT Chile grant 3170280 and 1171014.

# PgmNr 2741: RNA-seq isoform reconstruction and long read sequencing identifies smoking-induced alternative polyadenylation site usage in 1,221 smokers from the COPDGene Study.

## Authors:

P. Castaldi <sup>1</sup>; Z. Xu <sup>1</sup>; E. Tseng <sup>2</sup>; K. Tuttle <sup>3</sup>; M. Parker <sup>4</sup>; A. Saferali <sup>1</sup>; J. Yun <sup>1</sup>; R. Chase <sup>1</sup>; S. Lee <sup>1</sup>; M. Cho <sup>1</sup>; E. Silverman <sup>1</sup>; A. Laederach <sup>5</sup>; C. Hersh <sup>1</sup>; D. Hill <sup>6</sup>; G. Sheynkman <sup>6</sup>; M. Vidal <sup>6</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Brigham and Women's Hospital, Boston, MA; 2) Pacific Biosciences, Menlo Park, CA; 3) Northeastern University, Boston, MA; 4) Alnylam Pharmaceuticals, Cambridge, MA; 5) University of North Carolina, Chapel Hill, North Carolina; 6) Dana Farber Cancer Center, Boston, MA

## Background

Current cigarette smoking (CS) has profound transcriptomic effects and is a risk factor for many diseases. Cells use post-transcriptional RNA processing mechanisms to respond rapidly to environmental stresses, such as oxidative stress caused by CS. However, post-transcriptional changes induced by CS exposure have not been studied on a genome-wide level.

## Methods

To assess changes in isoform usage induced by CS, we analyzed blood RNA-seq data from 1,221 current and former smokers in the COPDGene Study. We estimated isoform abundance using the pseudoalignment method implemented in the Salmon program, and we also performed analysis to detect differential exon usage using limma. For 125 genes with significant evidence of differential expression or usage, confirmation of these results was performed via PacBio long read RNA sequencing (Iso-Seq) from three COPDGene current and former smokers.

## Results

We identified 1,062 isoforms in 891 genes with evidence of differential usage and 10% FDR in current versus former smokers (i.e. usage = significantly different fold change relative to all other expressed isoforms of a gene). 426 isoforms showed evidence of increased usage as opposed to 636 with decreased usage. We observed a systematic tendency for up-used isoforms to be the longest isoform for a given gene, and this length difference was largely accounted for by the 3' untranslated regions (UTRs) which were significantly longer in up-used isoforms (median length = 1125 nucleotides) than both down-used and non DIU isoforms (median lengths = 617 and 666, respectively,  $p < 0.05$  for both). Exon-level confirmed these results and demonstrated that for genes where CS induced 3' UTR lengthening, the effect was to decrease overall expression levels. We also observed that up-used 3' UTRs were enriched for both polyadenylation and micro-RNA binding sites. To validate a subset of these findings, we used targeted long-read sequencing (ORF Capture-Seq). With a single Sequel 1M SMRT cell, we obtained 488 on-target transcripts representing 73 of 125 targeted genes. We observed 138 annotated and 215 novel isoforms, with 83% concordance between the dominant isoform in long read and short read data from the same samples.

## Conclusions

CS exposure causes widespread lengthening of the 3' UTR via usage of alternative polyadenylation sites, resulting in acquisition of novel micro-RNA binding sites with a tendency towards lower overall

expression of these genes.

## PgmNr 2742: Integrated Hi-C, ATAC-Seq, and RNA-seq studies in bronchial epithelial cells (BECs) identifies both known and novel asthma candidate genes.

### Authors:

B. Helling; D. Sobreira; G. Hansen; N. Sakabe; K. Luo; C. Billstrand; R. Nicolae; B. Laxman; D. Nicolae; S. White; M. Nobrega; C. Ober

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of Chicago, Chicago, Illinois.

---

Asthma is a complex disease with an estimated 50% heritability. Although GWASs have identified over 60 asthma-associated loci, most SNPs at these loci are in non-coding regions and the relevant gene(s) remain unclear. Because BECs are a central cell type in asthma pathogenesis, we focused our studies in these cells to: 1) Identify differentially expressed genes (DEGs) and differentially accessible chromatin (DACs) between cells from asthmatic and non-asthmatic individuals, 2) Map functional interactions between DEGs and DACs using promoter capture Hi-C, and 3) Assess transcriptional and epigenetic patterns at known asthma loci and identify novel biological pathways. BECs were obtained from donor lungs from the Gift of Hope Regional Organ Bank of Illinois that were not used for transplant. Cultured cells from 18 donors (9 with asthma, 9 without asthma) were used for ATAC-seq and RNA-seq, and freshly isolated BECs pooled from 8 donors were used for promoter capture Hi-C. We first identified 1,907 DEGs and 2,655 DACs at an FDR  $\leq 0.05$ . Next, we mapped all predicted Hi-C chromatin interactions (loops) between gene promoters and DACs. Overall, there were 9,473 physical interactions between gene promoters and DACs, of which chromatin accessibility and transcript levels were significantly correlated for 1,015 DAC-promoter pairs, corresponding to 443 unique genes (Spearman FDR  $\leq 0.05$ ). Of these unique genes, 124 were also differentially expressed between asthmatics and non-asthmatics and formed 261 significantly correlated DAC-DEG pairs (Spearman FDR  $\leq 0.05$ ). These included genes at known asthma loci (e.g. *CCL20* on chromosome 2q36.21 and *TSLP* on chromosome 5q22.1), as well as genes not previously implicated in asthma (e.g., *FAM131C* on chromosome 1p36.13 and *IL4I1* on chromosome 19q13.33). Gene ontology (GO) analysis was performed using ToppGene's ToppFun-functional analysis tool to determine the functional categories enriched in these 124 genes. The top GO terms included fluid shear stress, cell migration and motility and response to vitamin E. In summary, combining multi-omic data from RNA-Seq, ATAC-Seq and Hi-C from a relevant asthma cell type yielded a set of asthma candidate genes and genomic regions implicated in their regulation, and provides a strategy for assigning genes with function in BECs to GWAS loci.

## PgmNr 2743: A meta-analysis of gene expression profiles in hypertensive African American women and mice.

### Authors:

W. Camejo Nunez <sup>1</sup>; N. Noren Hooten <sup>2</sup>; S. Kumar <sup>3</sup>; M. Evans <sup>2</sup>; D. Dluzen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biology, Morgan State University, Baltimore, MD; 2) Laboratory of Epidemiology and Population Science, National Institute on Aging, Baltimore, MD; 3) Department of Biomedical Engineering, Emory University, Atlanta, GA

---

Hypertension is a major risk factor for cardiovascular and cerebrovascular diseases, including heart failure, atherosclerosis, and stroke. Over 85 million people in the United States live with hypertension and African Americans (AAs) have the highest prevalence. Nearly 50% of AA women will be diagnosed with hypertension at some point in their life, however, few studies examine how hypertension develops and affects this population. Our goal is to understand how differential gene expression (DGE) contributes to hypertension health disparities in AA women. DGE is considered an important contributing factor to the predisposition or pathology of diseases. Previously, our lab found that mRNAs are differentially-expressed in AA women with hypertension and differentially-expressed microRNAs (miRNAs) play a regulatory role governing these gene expression patterns. For this follow-up analysis, we hypothesized that comparing our previous datasets of gene expression in AA women with hypertension with gene expression datasets in different animal models of hypertension, might identify an overlap of mRNA or miRNA genes, suggesting their importance for hypertension. For this meta-analysis, we compared a microarray examining DGE in peripheral blood mononuclear cells isolated from AA and white women with hypertension (n=6/group) and compared this with a microarray of DGE in endothelial cells, isolated from the left coronary artery of mice with induced shear stress (n=6/group). We found that DGE of 13 mRNAs overlapped between human and mouse, including *PPP1R16B*, *GADD45A*, *ADAM15*, *ICAM15*, *SNN*, *BIRC3*, *FBXO34*, *ICK*, *STX6*, *HMGB2*, *CALML4*, *SIVA1*, and *EHBP1L1*. We also identified an overlap of 4 miRNAs, miR-19b-1-5p, miR-103a-2-5p, miR-20a-5p, and miR-30c-5p, between our human and mouse datasets. We used cut-off values of  $P < 0.05$ , an absolute value in fold-change  $\geq 1.5$ -fold, and an FDR or 0.3 to identify significant genes. We predicted miRNA binding within the 3' untranslated region (UTR) of each mRNA identified using the TargetScan database (v7.2). These four miRNAs are predicted to target *FBXO34*, *ICK*, *STX6*, and *CALML4* in human and mice and may contribute to their differential expression. We intend to validate the expression of these mRNA-miRNA pairs in hypertension physiology. In conclusion, we identified 13 mRNAs and 4 miRNAs that overlap between human and mouse which could play an important role in hypertension in AA women.



## PgmNr 2744: Unraveling regulation of *FOXF1* expression and lung transcriptome perturbations in patients with ACDMPV.

### Authors:

P. Szafranski <sup>1</sup>; T. Gambin <sup>1,2</sup>; J.A. Karolak <sup>1,3</sup>; D. Lanza <sup>1</sup>; J. Heaney <sup>1</sup>; E. Popek <sup>4</sup>; P. Stankiewicz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College Medicine, Houston, Texas.; 2) Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland.; 3) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poznan, Poland.; 4) Department of Pathology and Immunology, Baylor College of Medicine, Houston, Texas.

---

Alveolar capillary dysplasia with misalignment of pulmonary veins (ACDMPV) is a rare neonatal lethal lung developmental disorder featuring severe respiratory failure and refractory pulmonary hypertension (PAH). The vast majority of ACDMPV cases are sporadic and causative heterozygous SNVs or CNV deletions involving *FOXF1* or its distal lung-specific enhancer on 16q24.1 have been identified in 80-90% of ACDMPV patients. *FOXF1* is a forkhead family transcription factor, expressed in mesoderm-derived cells in response to SHH signaling mediated by *GLI2*. Interestingly, knockout mice for lncRNA *Fendrr*, which maps 1.5 kb upstream to *Foxf1* in the opposite orientation and likely utilizes the same bi-directional promoter, die due to cardiopulmonary anomalies. We studied transcriptomes in fetal lung-derived cell line IMR-90 depleted of *FOXF1*, *FENDRR*, or genes encoding transcription factors predicted by ChIP-seq to bind the *FOXF1* enhancer, using CRISPR/Cas9 or RNAi. We found that *TFAP2C* positively regulates expression of *FOXF1* and that *FENDRR* weakly suppresses expression of *FOXF1*, whereas *FOXF1* strongly promotes *FENDRR* expression. To further expand our previous ACDMPV and *Foxf1*<sup>+/-</sup> mouse transcriptome microarray studies (PMID: 24722050), and to identify additional genes downstream of *FOXF1*, we also performed comparative RNA-seq analyses of lung transcriptomes in ACDMPV neonates with pathogenic *FOXF1* loss-of-function variants and neonatal lung controls. In addition, we compared lung transcriptome perturbations in newborns with ACDMPV vs. acinar dysplasia and congenital alveolar dysplasia due to pathogenic variants in *TBX4* or *FGF10* (PMID: 30639323). We identified 39 genes with a statistically significant deregulation of their expression, including *GREM1* known to play a role in the pathogenesis of PAH and linked to pulmonary fibrosis (PF) and lung adenocarcinoma, *AVPR1A* expressed in vascular smooth muscles, *PDE2A* and *TTN* misregulated in pulmonary edema, *PLA2G4F* and *DIO2* associated with PF, and *CSH2*, *HOXD9*, and 21 other genes linked to lung cancer (abnormal cell proliferation, apoptosis, and epithelial-mesenchyme transition). Overall, our analyses provide insight into complex *FOXF1* regulation, confirm our previous microarray-based findings, and contribute to a better understanding of the differences between SHH-*FOXF1* and *TBX4*-*FGF10* signaling pathways in lung development.



## PgmNr 2745: The influence of poverty on gene expression in African Americans in HANDLS.

### Authors:

N. Arnold <sup>1</sup>; N. Noren Hooten <sup>2</sup>; Y. Zhang <sup>3</sup>; K. Becker <sup>3</sup>; D. Dluzen <sup>1</sup>; M. Evans <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biology, Morgan State University, Baltimore, MD 21212; 2) Laboratory of Epidemiology and Population Sciences National Institute on Aging, National Institutes of Health, Baltimore, MD 21224; 3) Laboratory of Genetics and Genomics National Institute on Aging, National Institutes of Health, Baltimore, MD 21224

---

Socioeconomic status (SES) and other social determinants of health, including access to health care and nutritious food, correlates with health inequalities and contribute to health disparities. For example, African American (AA) men living below poverty in Baltimore City have a higher incidence of cardiovascular-related mortality when compared to either white males or AA females living below poverty. Previous studies in our laboratory and elsewhere suggest that environmental conditions are associated with differentially-expressed genes in white blood cells and this may contribute to the onset of diseases in the immune or cardiovascular systems. Differential gene expression (DGE) patterns observed in our laboratory have been associated with hypertension and cardiovascular disease and correlate with race and gender. However, no studies have investigated how poverty status affects DGE between male and female AAs and whites living in Baltimore City. We examined DGE in 54 AA and white participants of the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) cohort, who are living above or below 125% of the 2004 federal poverty line. We performed a microarray to assess DGE patterns in peripheral blood mononuclear cells (PBMCs) from these participants. We identified genes related to tumor necrosis factor signaling (TNF), antigen processing and presentation, NF-kB signaling, toll-like receptor signaling, and other inflammation-related pathways that were differentially-expressed in our cohort, particular in AA males living below poverty. There were 1,616 genes differentially-expressed between AA men and women living in poverty. African American males living in poverty had 229 differentially expressed genes compared with AA men living above poverty, of which 46 were up-regulated and 183 down-regulated. We performed gene expression validation using RT-qPCR of target genes *CD36*, *KCTD12*, *DUSP2*, and *GIMAP1* to confirm our microarray results and validate the presence of DGE in AA men living in poverty. This study will serve as a link to clarify the social determinants responsible for health disparities and determine if poverty status influences gene expression in the immune system. The identification of the biological mechanisms linking poverty to health disparities can potentially improve health outcomes for populations at a higher risk for mortality or other cardiovascular-related diseases in disparate communities.

## PgmNr 2746: Integrative vascular endothelial cell genomics identify *AIDA* as a coronary artery disease candidate gene.

### Authors:

S. Lalonde<sup>1</sup>; V.-A. Codina-Fauteux<sup>1,2</sup>; S. Méric de Bellefon<sup>1,2</sup>; F. Leblanc<sup>1,2</sup>; M. Beaudoin<sup>1</sup>; M.-M. Simon<sup>3</sup>; R. Dali<sup>3</sup>; T. Kwan<sup>3</sup>; K.S. Lo<sup>1</sup>; T. Pastinen<sup>4</sup>; G. Lettre<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Montreal Heart Institute, Montréal, Québec, H1T 1C8, Canada; 2) Faculté de Médecine, Université de Montréal, Montréal, Québec, H3T 1J4, Canada; 3) McGill University and Genome Québec Innovation Center, Montréal, Québec, H3A 0G1, Canada; 4) Center for Pediatric Genomic Medicine (CPGM), Children's Mercy Kansas City, 2401 Gillham Road, Kansas City, MO 64108

---

**Background.** Genome-wide association studies (GWAS) have identified 100s of loci associated with coronary artery disease (CAD) and blood pressure (BP)/hypertension. Many of these loci are not associated with traditional risk factors, nor include obvious candidate genes, complicating their functional characterization. We hypothesized that many GWAS loci associated with vascular diseases modulate endothelial functions. Endothelial cells play critical roles in regulating vascular homeostasis (e.g. selective barrier, inflammation, hemostasis, vascular tone) and endothelial dysfunction is a hallmark of atherosclerosis and hypertension.

**Results.** We generated an integrated map of gene expression (RNA-sequencing), open chromatin regions (ATAC-sequencing), and 3D interactions (Hi-C) in resting and TNF $\alpha$ -treated human endothelial cells. We showed that genetic variants associated with CAD and BP are enriched in open chromatin regions identified in endothelial cells. We used physical loops identified by Hi-C to link open chromatin peaks that include CAD or BP SNPs with the promoters of genes expressed in endothelial cells. This analysis highlighted 991 combinations of open chromatin regions-gene promoters that map to 38 CAD and 92 BP GWAS loci. At a CAD locus, we validated one of these pairs by engineering a deletion of the TNF $\alpha$ -sensitive regulatory element using CRISPR/Cas9 and measuring the effect on the expression of the novel CAD candidate gene *AIDA*.

**Conclusions.** Our data support an important role played by genetic variants acting in the vascular endothelium to modulate inter-individual risk in CAD and hypertension.

## PgmNr 2747: DNA methylation changes in newborn leucocyte: The mechanisms and the prediction of hypoplastic left heart syndrome.

### Authors:

S. Vishweswaraiah <sup>1</sup>; R. Bahado-Singh <sup>1</sup>; N. K Mishra <sup>2</sup>; C. Guda <sup>2</sup>; U. Radhakrishna <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oakland University William Beaumont School of Medicine, Royal Oak, MICHIGAN.; 2) University of Nebraska Medical Center Omaha, Nebraska

---

**Purpose:** Hypoplastic Left Heart Syndrome (HLHS) is a severe Congenital Heart Defect (CHD characterized by aberrant development of the left ventricle and its outflow tract). It accounts for 1.4 to 3.8% of all CHDs. While volume of blood flow through the left heart and outflow tract is important, the molecular pathogenesis of HLHS remains largely unknown. Epigenetic changes reflect the impact of the environment on the genome and is an important mechanism for the control of transcription. We investigated the epigenetic changes (DNA methylation) associated with non-syndromic HLHS by the analysis of newborn leucocyte DNA. In addition, we evaluated the diagnostic accuracy of methylation of cytosine (CpG) biomarkers for HLHS detection.

**Methods:** Genome wide methylation profiling was conducted using Illumina Methylation assay in 24 HLHS subjects and matched controls using newborn heel-stick blood. Differential methylation analysis of individual CpG loci in cases versus controls was performed using “GenomeStudio” followed by Principal Component Analysis (PCA), heatmap, regression analysis, Area Under the Receiver Operating Characteristic Curve (AUC) using various “R” packages (v3.4). Pathway analysis was performed using Ingenuity Pathway Analysis (IPA) tool.

**Results:** 512 differentially methylated CpGs (1.5-fold or higher change in methylation in cases versus controls and Benjamini-Hochberg FDR  $p < 0.05$ ) in 557 genes were identified. In terms of individual predictive accuracy, 141 CpGs (153 genes) had excellent predictive accuracy ( $AUC \geq 0.90$ ) for HLHS detection. We identified statistically significant CpG methylation differences with HLHS in multiple genes previously known to play a role in heart developmental: *JARID2*, *MEF2A*, *MEF2C*, *RICTOR*, *IRX3* and *EP300*. Significant methylation changes were also identified in novel genes not previously linked to CHD. Known pathways such as TGF- $\beta$ , VEGF and Hypoxia signaling were identified to be dysregulated due to methylation of candidate genes. These pathways were previously established to be critical to cardiovascular morphogenesis.

**Conclusion:** We present the first genome-wide DNA methylation analysis in newborn HLHS subjects. In addition analysis was performed non-invasively using newborn blood. Epigenetic dysregulation was found to play a significant role in HLHS with DNA methylation changes in cardiac developmental genes. Finally, blood epigenetic markers appeared to accurately predict the presence of HLHS in the newborn.

## PgmNr 2748: Epigenome wide association study (EWAS) identifies *KDM4B* as a potential mediator of the link between psychosocial stress and cardiovascular disease.

### Authors:

R. Jiang<sup>1</sup>; L.C. Kwee<sup>2</sup>; J.W. Abdulrahim<sup>2</sup>; A. Singh<sup>1</sup>; B.H. Brummett<sup>1</sup>; M.A. Babyak<sup>1</sup>; I.C. Siegler<sup>1</sup>; R.B. Williams<sup>1</sup>; W.E. Kraus<sup>2,4</sup>; E.R. Hauser<sup>2,3</sup>; S.H. Shah<sup>2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry and Behavioral Sciences, Duke Univ Med Ctr, Durham, NC; 2) Duke Molecular Physiology Institute, Duke Univ Med Ctr, Durham, NC; 3) Department of Biostatistics and Bioinformatics, Duke Univ Med Ctr, Durham, NC; 4) Department of Medicine, Duke Univ Med Ctr, Durham, NC

**Introduction.** Psychosocial stress has been associated with cardiovascular disease (CVD) risk factors and outcomes. Epigenetic changes such as DNA methylation may be a potential mechanism to account for this association, namely that stress induces epigenetic changes which ultimately lead to increased CVD risk. Thus, we performed an epigenome-wide study (EWAS) for stress measures and then determined if differentially methylated CpG sites resulted in increased CVD risk.

**Methods.** The discovery cohort consisted of individuals enrolled in the CATHGEN cardiovascular study (N=845) and a validation cohort consisting of individuals enrolled in the Duke Caregiver study (N=285). Methylation profiling was performed with one of two Illumina methylation arrays: the 450K or EPIC arrays. Stress measure was previously constructed and validated, consisting of a sum of scores (0-5) of chronic psychosocial stress based on the availability of five components in both cohorts. METAL package was used for meta-analysis of the two different arrays within each study, only including CpG sites represented in both arrays. Individual CpG were tested for association with the composite stress measure using linear regression model, adjusted for age, gender, race and batch effects. CpG sites significantly associated with this stress measure in both cohorts were then tested as methylQTL using whole blood gene expression array data in CATHGEN, and further tested for association with incident myocardial infarction (MI) or death using Cox proportional hazards models in CATHGEN.

**Results.** Fifty-two CpG sites were differentially methylated for the composite stress measure ( $p < 0.0001$ ) in CATHGEN; one of these validated in the Caregiver study ( $p < 0.05$  and same direction of effect): cg23839891 (*KDM4B*). This site didn't have a significant methylQTL. Greater methylation at cg23839891 was associated with higher risk of death or MI ( $P = 0.04$ ). *In silico*, this site was shown to be in a DNase I hypersensitivity site in 24 cell types.

**Conclusions.** Using an EWAS approach for a composite measure of chronic psychosocial stress, we have identified a differentially methylated, potentially functionally relevant CpG site in the *KDM4B* gene that also predicts increased risk for incident CVD events. *KDM4B* encodes lysine demethylase 4B which is involved in chromatic regulation and thus is a good biological candidate for linking psychosocial stress and CVD disease.

## PgmNr 2749: WASF2 differential gene expression and its link to hypertension health disparities.

### Authors:

M. Arkorful <sup>1</sup>; A. Hewwit <sup>2</sup>; N. Arnold <sup>1</sup>; N. Noren Hooten <sup>2</sup>; Y. Zhang <sup>3</sup>; M.K. Evans <sup>2</sup>; D. Dluzen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Morgan State University, Baltimore, MD.; 2) Laboratory of Epidemiology and Population Science, National Institute on Aging, 251 Bayview Blvd. Baltimore, MD 21224; 3) Laboratory of Genetics and Genomics, National Institute on Aging, 251 Bayview Blvd. Baltimore, MD 21224

---

The prevalence of hypertension in African Americans (AA) in the US is among the highest in the world, where almost half of all AA women in the US have high blood pressure. Previous data from our laboratory suggest that differential gene expression (DGE) of *WASF2* may have a role in hypertension-related disparities in AA. *WASF2* forms a multiprotein complex that plays a role in the signaling of the actin cytoskeleton, influencing changes in cell shape, motility, or function, and has never been associated with hypertension pathology. The actin cytoskeleton, composed of actin filaments, is responsible for cell migration, and controlling the shape of the cell. The actin cytoskeleton in the endothelium provides a barrier which allows the cells beneath the endothelium to respond to environmental cues and make structural changes. Variation in the gene expression levels of *WASF2* may influence cardiovascular diseases, such as hypertension or atherosclerosis, and may be exacerbated by social determinants of health including factors like socioeconomic status and stress in the environment. We found that DGE of mRNAs and microRNAs (miRNAs) in peripheral blood mononuclear cells (PBMCs) isolated from AA and white women with or without hypertension were involved in hypertension-related pathways that were differentially-expressed by race and may contribute to predisposition to hypertension. Because of its effect on cell shape and function, we hypothesized that DGE of *WASF2* is involved in endothelial dysfunction in hypertension. *In silico* miRNA target prediction suggested that miRNA miR-1253 is involved in the expression of *WASF2* mRNA and protein. miR-1253 mimics were transfected into HUVECs and HAECs and *WASF2* mRNA and protein levels were significantly repressed ( $P < 0.05$ ), confirming miR-1253 regulates *WASF2* expression. Currently, we are performing experiments to confirm the downstream impact of this regulatory pathway on endothelial cell cytoskeleton function, and validate the miR-1253 binding site in the *WASF2* 3' UTR. Discovery of additional miRNAs which are involved in disparate diseases will provide more options for treatment and help to combat hypertension disparities, particularly in communities most at risk.

## PgmNr 2750: Multi-omics analysis identify microRNAs associated with cardiometabolic traits.

### Authors:

M.M.J. Mens <sup>1</sup>; S.C.E. Maas <sup>1,2</sup>; J. Klap <sup>3</sup>; G.J. Weverling <sup>3</sup>; P. Klatser <sup>3</sup>; J.P.J. Brakenhoff <sup>3</sup>; J.B.J. van Meurs <sup>1,4</sup>; A.G. Uitterlinden <sup>1,4</sup>; M.A. Ikram <sup>1</sup>; M. Kavousi <sup>1</sup>; M. Ghanbari <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands; 2) Department of Genetic Identification, Erasmus MC University Medical Center, Rotterdam, the Netherlands; 3) World Without Disease Accelerator, Data Sciences & Prevention Biomarkers, Johnson & Johnson, Leiden, the Netherlands; 4) Department of Internal Medicine, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands

---

**Introduction:** MicroRNAs (miRNAs) are non-coding RNA molecules that regulate gene expression. Extensive research has explored the role of miRNAs in the risk for type 2 diabetes (T2D) and coronary heart disease (CHD) using single omics data, but much less using multi-omics data.

**Methods:** We aimed to link miRNAs with cardiometabolic traits by integrating multiple omics layers, including genomics, epi-genomics and miRomics. First, using publicly available summary statistics from large genome-wide association studies we sought genetic variants in miRNA-related sequences associated with cardiometabolic risk factors and diseases, including lipid and obesity-related traits, glycemic indices, blood pressure, and disease prevalence of T2D and CHD. Second, we used DNA methylation (N=1450) and miRNA expression data (N=2000) measured in blood from the population-based Rotterdam Study to further investigate the link between associated miRNAs and cardiometabolic traits.

**Results:** After correcting for multiple testing, 202 genetic variants annotated to 71 independent miRNAs were associated with the studied cardiometabolic traits. Alterations in DNA methylation levels of CpG sites annotated to 39 of these miRNAs were associated with the same trait. Moreover, we found that plasma expression levels of 6 of the 71 identified miRNAs were also associated with the same trait. Integrating the results of different omics data showed miR-148a, miR-125b and miR-100 to be strongly linked to lipid traits.

**Conclusion:** Our multi-omics analysis demonstrates multiple miRNAs associated with cardiometabolic risk factors that might be important regulators of biological processes involved in T2D and CHD. The identified miRNAs could be considered as candidate biomarkers for early diagnosis and progression of these cardiometabolic diseases.

## PgmNr 2751: Integration of DNA sequencing and transcriptome data for association analyses of rare variants and lipid traits.

### Authors:

T. Yang <sup>1</sup>; C. Wu <sup>2</sup>; P. Wei <sup>3</sup>; W. Pan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Biostatistics, University of Minnesota, Minneapolis, MN.; 2) Department of Statistics, Florida State University, Tallahassee, FL.; 3) Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX

---

Leveraging comprehensive regulatory knowledge, transcriptome-wide association studies utilize GWAS data to showcase their improved statistical power of identifying gene-trait associations and importantly, offer further biological insights. Comparing to common variants, the findings for rare variants are limited and their underlying role in regulating gene expression is not totally clear. To fill in the gaps, we integrate whole genome sequencing data with transcriptome data for rare variants. Using data from Framingham Heart Study, we demonstrate that rare variants play an important and universal role in predicting gene expression, which is not completely due to linkage disequilibrium with the nearby common variants: the variation of gene expression of more than 6,000 genes can be explained by rare variants and around 1,500 of which cannot be explained well by common variants. Moreover, by including rare variants, in addition to common variants, the predictivity for 79% of the examined genes has a trend to increase. Incorporating this functional genomics information, we perform association testing for five lipid traits in two UK10K cohorts by hypothesizing that cis-eQTLs, including RVs, are more likely to be susceptible. We discover two genes, *LDLR* and *TTC22*, that are genome-wide significantly associated with low-density lipoprotein cholesterol among 3,910 subjects and are ignored by analysis of common variants. We further demonstrate that a joint analysis of both common and rare variants identifies association signals that would be missed by testing on either common variants or rare variants alone.



## PgmNr 2752: Investigation of gene-gene interactions in cardiac diseases and fatty acids in LURIC.

### Authors:

J. Zhou <sup>1</sup>; K. Passero <sup>2</sup>; B. Müller-Myhsok <sup>3, 5, 6</sup>; M.E. Kleber <sup>4</sup>; W. Maerz <sup>4, 7, 8</sup>; M.A. Hall <sup>1, 2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Veterinary and Biomedical Sciences, The Pennsylvania State University, University Park, PA.; 2) The Huck Institutes of the Life Science, The Pennsylvania State University, University Park, PA, USA; 3) Statistical Genetics, Max Planck Institute of Psychiatry, Munich, 80804, Germany; 4) Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; 5) Munich Cluster of Systems Biology, SyNergy, 81377 Munich, Germany; 6) Institute of Translational Medicine, University of Liverpool, Liverpool, L69 3BX, United Kingdom; 7) Synlab Holding Deutschland GmbH, SYNLAB Academy, Mannheim and Augsburg, Germany; 8) Medical University of Graz, Clinical Institute of Medical and Chemical Laboratory Diagnostics, Graz, Austria

---

Epistasis analysis elucidates the genetic effects of interactions between multiple loci for understanding complex traits. However, the large computational demands and the high multiple testing burdens impede their discovery. Here, we illustrate the utilization of two methods, main effect filtering based on individual GWAS results, and biological knowledge-based modeling through Biofilter, to reduce the number of testing interactions among Single nucleotide polymorphisms (SNPs) for 15 cardiac-related diseases and 14 fatty acids. We performed interaction analyses using the two filtering methods, adjusting for age, sex, BMI, Waist-Hip ratio, and principal components, among 2,824 samples from Ludwigshafen Risk and Cardiovascular Health Study (LURIC). Using Biofilter, one interaction was nearly Bonferroni significant: an interaction between *XRCC4* (rs7735781) and *XRCC5* (rs10804247) was identified for venous thrombosis with a Bonferroni-adjusted likelihood ratio test (LRT)  $p$ : 0.0627. The protein complex from *XRCC4* and *XRCC5* functions as the DNA ligase in the repair of double-strand DNA breaks and the completion of V(D)J recombination in the development of the lymphocytes. Lack of *XRCC4* expression would lead to the dysfunction in the development and homeostasis of hormones. A total of 429 interactions were identified from the main effect filtering for different diseases (10) and fatty acids (419) at Bonferroni-adjusted LRT  $p < 0.05$ . For diseases, the top interactions between *SNTG1* (rs1383819, rs12547263, and rs2467203) and gene close to rs1493939 or rs1389542 are associated with the history of arterial hypertension. *SNTG1*, a member of the syntrophin family, relates to idiopathic scoliosis which promotes the development of cardiac disease and pulmonary hypertension. For fatty acids, the top interactions between *KCND3* (rs4839193 and rs11102365) and an unknown gene close to *LOC107984022* (rs10829717) with Bonferroni-adjusted LRT  $p$  close to 0.00002 are both associated with linolelaidic acid (C18 2n6tt), an omega-6 trans fatty acid. *KCND3* is a part of the voltage-gated potassium channel which is widely expressed for regulating neurotransmitter functions, heart rate, and muscle contraction. Here, we successfully applied filtering approaches to identify numerous genetic interactions related to cardiac-related diseases and traits. These filtering methods offer the potential to detect epistasis in other common diseases and traits by overcoming the multiple testing burden.



## PgmNr 2753: Mendelian randomization of plasma transcriptome and proteome identify genes associated with blood pressure.

### Authors:

D.A. Baird <sup>1,2</sup>; J.C. Talbott <sup>2</sup>; T.R. Gaunt <sup>1,2</sup>; D.A. Lawlor <sup>1,2</sup>; J. Zheng <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 2) Population Health Science, Bristol Medical School, University of Bristol, Bristol, United Kingdom

Despite Genome Wide Association Studies (GWASs) of blood pressure (BP) traits detecting numerous associations, it remains a challenge to identify the causal biological mechanisms that explain these. Therefore, we carried out Mendelian Randomisation (MR) study to identify molecular (gene expression and protein expression levels) causes of BP variation that could then be used to identify genes involved in the regulation of BP.

We used two-sample MR with summary data from GWAS of plasma expression quantitative trait loci (eQTLs) on blood pressure sourced from the eQTL-Gen Consortium (n = 31,684 samples) and of plasma protein quantitative trait loci (pQTLs) from five recent pQTL GWAS (Zheng et al, n ~3,000 samples) as the genetic instruments. Summary results from a GWAS of SBP and DBP undertaken in the UK Biobank were used as genetic instrument-outcome data (sample 2). MR estimates were calculated either using Wald Ratios (WR) or Inverse Variance Weighted (IVW) method if multiple QTL instruments were available. Top MR findings were determined using a multiple testing threshold of  $P < 3.64 \times 10^{-6}$  (13,725 independent tests). Sharing of the same causal variant between the gene expression and blood pressure traits in the genomic region was verified by colocalization analysis. MR analysis was performed using the TwoSample MR R package and colocalization using the coloc R package.

Of the 65,424 MR tests across 15,353 genes, 339 gene and 131 protein expression associations with DBP or SBP passed the multiple threshold correction. We found two gene-blood pressure associations, *MST1* and *TIE1*, were supported by evidence from both eQTL and pQTL analyses. *MST1* was associated with DBP only (eQTL WR  $\beta = -0.085$ , SE=0.017,  $P = 4.50 \times 10^{-7}$ ; pQTL WR  $\beta = 0.012$ , SE=0.0021,  $P = 3.33 \times 10^{-9}$ ). *TIE1* was associated with both SBP (eQTL WR  $\beta = 0.205$ , SE=0.035,  $P = 4.87 \times 10^{-9}$ ; pQTL WR  $\beta = -0.050$ , SE=0.010,  $P = 5.06 \times 10^{-7}$ ) and DBP (eQTL WR  $\beta = 0.180$ , SE=0.035,  $P = 2.69 \times 10^{-7}$ ; pQTL WR  $\beta = -0.048$ , SE=0.010,  $P = 1.95 \times 10^{-6}$ ). Of these associations, *TIE1* gene expression showed moderate evidence of colocalization with SBP (coloc prob=67%). Current literature on these genes supports a role in cardiovascular disease pathogenesis as *MST1* is known to promote cardiac myocyte apoptosis and *TIE1* to mediate angiogenesis.

Our study suggests that both the gene expression and protein levels of *TIE1* are casually related to BP, although latent horizontal pleiotropy could still confound this association.

## PgmNr 2754: Genome-wide association study of the human lipidome: Towards identifying endophenotypes for coronary artery disease.

### Authors:

G. Cadby<sup>1</sup>; P.E. Melton<sup>1</sup>; J. Hung<sup>2</sup>; J. Beilby<sup>3</sup>; N.A. Mellett<sup>4</sup>; K. Huynh<sup>4</sup>; C. Giles<sup>4</sup>; M.P. Dube<sup>5</sup>; F. van Bockxmeer<sup>6</sup>; G.F. Watts<sup>2,7</sup>; J. Blangero<sup>8</sup>; P.J. Meikle<sup>4</sup>; E.K. Moses<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Genetic Origins of Health and Disease, Curtin University and University of Western Australia, WA, Australia; 2) Medical School, University of Western Australia, WA, Australia; 3) PathWest Laboratory Medicine WA, WA, Australia; 4) Metabolomics Laboratory, Baker Heart and Diabetes Institute, VIC, Australia; 5) Université de Montréal Beaulieu-Saucier Pharmacogenomics Centre, Montreal Heart Institute, Quebec, Canada; 6) School of Surgery, University of Western Australia, WA, Australia; 7) Lipid Disorders Clinic, Department of Cardiology, Royal Perth Hospital, WA, Australia; 8) South Texas Diabetes and Obesity Institute, The University of Texas Rio Grande Valley School of Medicine, Texas, USA

---

**Background:** Coronary artery disease (CAD) is the leading cause of death worldwide. Circulating lipid levels play an important role in CAD and genetic investigations into the human lipidome may provide greater insight into disease risk. We have utilised a large community-based study, the Busselton Family Heart Study (n=4492), with the aim of identifying causal genetic risk variants for CAD, through the integration of lipidomic endophenotypes, genomic variants, and whole-genome sequence (WGS) data.

**Methods:** Targeted lipidomic profiling was performed using electrospray ionization-tandem mass spectrometry to identify 530 lipid species from 33 lipid classes/subclasses. To exploit both known and unknown relatedness, we estimated empirical kinship probabilities using LDAK. The general linear mixed model incorporated in SOLAR was used to estimate narrow-sense heritabilities of lipid species. Using BOLT-LMM, we performed genome-wide association studies (GWAS) of all 530 lipid species on ~14M imputed variants ( $R^2 > 0.3$  & minor allele copies  $\geq 5$ ). All analyses were adjusted by age, sex, and their interactions. GWAS were run both with and without adjustment for plasma lipid concentrations (i.e. HDL-C, total cholesterol and triglycerides).

**Results:** Over 99% of lipid species were significantly heritable ( $h^2$ : 0.07 to 0.51; FDR adjusted p-value:  $1.1 \times 10^{-2}$  to  $1.1 \times 10^{-54}$ ), with a median heritability of 0.30 (IQR=0.09). Of the 530 lipid species, 451 lipid species had at least one variant associated at the genome-wide level of significance ( $p < 5 \times 10^{-8}$ ). We identified over 60,000 significant associations between lipid species and genetic variants from approximately 150 gene regions. Approximately 70% of the observed associations were significant with and without adjustment for plasma lipid concentrations, while 30% were significant in only one analysis. Across all lipid species, 0.5% to 16.5% of the variation in species was attributed to the significantly-associated genotyped variants.

**Conclusion:** This is the largest GWAS of the human lipidome (largest sample size and number of lipid

species) in which we have shown that the plasma lipidome is heritable and have identified new genetic variant-lipid species associations. Future work includes interrogation of the WGS data within these gene regions in an attempt to identify putative causal risk variants for lipidome endophenotypes and CAD.

## PgmNr 2755: Multi-omic analyses implicate sphingolipids as biomarkers of recurrent stroke in African Americans.

### Authors:

N.M. Davis Armstrong <sup>1</sup>; K.J. Spragley <sup>1</sup>; W.-M. Chen <sup>2,3</sup>; F.-C. Hsu <sup>4</sup>; M.S. Brewer <sup>1</sup>; S.R. Williams <sup>5</sup>; M.M. Sale <sup>2,3</sup>; B.B. Worrall <sup>3,5</sup>; K.L. Keene <sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biology, East Carolina University, Greenville, North Carolina.; 2) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 3) Department of Public Health Sciences, University of Virginia, Charlottesville, VA; 4) Department of Biostatistics and Data Science, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 5) Department of Neurology, University of Virginia, Charlottesville, VA, USA; 6) Center for Health Disparities, Brody School of Medicine, East Carolina University, Greenville, NC

African Americans endure a nearly two-fold greater risk of suffering a stroke and are 2-3 times more likely to die from stroke compared to Caucasians. Among individuals who survive an initial stroke, African Americans have a greater risk of recurrent stroke and vascular events, which are deadlier and more disabling than initial events. Metabolomics provide insight of physiological states of systems and represent potential biomarkers that could influence or predict disease and/or recurrence. Combining metabolomics data with traditional genetic and epigenetic methods could help better understand disease etiology and progression for complex diseases with both heritable and environmental risk factors.

We conducted an integrative, multi-omic study on 922 plasma metabolites, 473,864 DNA methylation loci, and 556 SNPs from 50 African American participants in the Vitamin Intervention for Stroke Prevention (VISP) clinical trial. Welch's *t*-tests, survival and matched-pairs analyses were performed on metabolite profiles. Network analyses using Weighted Gene Co-expression Network Analysis (WGCNA) tested correlations between metabolites, methylation loci, and clinical traits. Epigenetic associations were identified using regression and WGCNA, while SNP-based tests were used to detect metabolite differences by genotype.

Sixteen metabolites, including cotinine ( $p=9.25e-13$ ), N-delta-acetylornithine ( $p=1.97e-05$ ), and sphingomyelin (d18:2/23:0, d18:1/23:1, d17:1/24:1) ( $p=3.00e-04$ ) were identified in *t*-tests of recurrent stroke outcome or baseline smoking status. Tricosanoyl sphingomyelin (d18:1/23:0) exceeded statistical significance in Cox Proportional Hazards survival analyses with time to recurrent stroke ( $p=1.56e-05$ ). Network analyses identified moderate correlations between sphingolipids and clinical traits including days to recurrent stroke ( $p=0.007$ ). Integrative analyses between genetic variants in sphingolipid pathway genes identified 11 SNPs nominally associated with metabolite levels in a one-way ANOVA, while epigenomic analyses identified xenobiotics, predominately smoking metabolites and pharmaceutical drugs, to be associated with methylation profiles.

Taken together, our results suggest that metabolites, specifically those associated with sphingolipid metabolism, are potential plasma biomarkers for stroke recurrence in African Americans. Furthermore, genetic variation and DNA methylation may play a role in the regulation of these

metabolites.

## PgmNr 2756: Genome-wide association analysis of 100,000 Chinese individuals identifies novel loci for blood pressure traits.

### Authors:

W. Gan<sup>1,2</sup>; K. Lin<sup>1</sup>; I.Y. Millwood<sup>1,3</sup>; Y. Guo<sup>4</sup>; C. Yu<sup>5</sup>; J. Lv<sup>5</sup>; R.J. Clarke<sup>1</sup>; L. Li<sup>5</sup>; Z. Chen<sup>1</sup>; R.G. Walters<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, U.K.; 2) Big Data Institute, University of Oxford, Oxford, U.K.; 3) Medical Research Council Population Health Research Unit at the University of Oxford, Oxford, UK; 4) Chinese Academy of Medical Sciences, Dong Cheng District, Beijing, China; 5) School of Public Health, Peking University Health Science Center, Beijing, China

### Abstract

**Background:** High blood pressure (BP) is a leading global health risk factor and genome-wide association studies (GWAS), conducted predominantly in populations of European descent, have identified >1000 loci associated with blood pressure traits. Major ancestry groups exhibit differences in patterns of linkage disequilibrium and allele frequencies, and may harbor population-specific association signals. Thus, GWAS in non-European populations have the potential to identify additional loci.

**Methods:** We performed GWAS studies for systolic, diastolic, pulse and mean arterial blood pressure (SBP, DBP, PP, and MAP) in up to 100,473 Chinese participants aged 30-79 years from the China Kadoorie Biobank (CKB). We used linear mixed models testing an additive genetic model (BOLT-LMM v2.3.2) for each of the four blood pressure traits for ~10 million genotyped and imputed genetic variants (MAF>0.005, info>0.3) with adjustment for age, age-squared, sex, recruitment region, anti-hypertensive medication, genotyping array version, mean outdoor temperature, and body mass index. SNP based heritability was estimated using linkage disequilibrium (LD) score regression.

**Results:** The estimated SNP-wide heritability ( $h^2$ ) of the blood pressure traits in CKB was 0.213, 0.217, 0.171 and 0.225 for SBP, DBP, PP, and MAP, respectively. Low genomic control factors ( $\lambda_{GC} = 1.121-1.168$ ) and LD score regression intercepts (0.935-0.949) indicated little inflation and were consistent with the polygenic nature of BP traits. We identified a total of 135 independent loci associated with one or more the blood pressure traits at  $p < 5 \times 10^{-8}$ . Among them, 78 (5 novel) were associated with SBP, 89 (3 novel) with DBP, 44 (1 novel) with PP, and 98 (4 novel) with MAP. Thus, this CKB study confirms the transferability between ancestry groups of many associations, while also extending the list of East Asian-specific blood pressure-associated variants.

**Conclusions:** This study demonstrates that further GWAS analyses of blood pressure in diverse populations, including the Chinese population, can provide important additional information about the genetic determinants of blood pressure and other complex traits.

## PgmNr 2757: *TBX2* gene polymorphisms as factors for predisposing coronary artery diseases in South Indian population.

### Authors:

M. Ganesan <sup>1</sup>; S.K. Katkam <sup>2</sup>; V.K. Kutala <sup>2</sup>; S. Arora <sup>3</sup>; Z. Syeda <sup>3</sup>; A. Neyha Zainab <sup>3</sup>; Q. Hasan <sup>3</sup>; U.K. Hosad <sup>4</sup>; L. Loret <sup>4</sup>; T. Kumarasamy <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) CSIR-Centre for Cellular and Molecular Biology, Hyderabad, Telangana, India; 2) Department of Clinical Pharmacology and Therapeutics, Nizam's Institute of Medical Sciences (NIMS), Hyderabad, India; 3) Department of Genetics and Molecular Medicine, Kamineni Hospitals, Hyderabad, India.; 4) Yashoda Super Specialty Hospitals, Hyderabad India

---

**Background:** Coronary artery disease (CAD) has been the leading cause of mortality and morbidity worldwide. It is projected that the annual number of deaths due to CVD will increase from 17 million in 2008 to 25 million in 2030 (World Health Organisation, 2012). Although it is a multifactorial disease, several studies have demonstrated that genetic factors play a major role in CAD.

*TBX2* transcriptional program plays central role for coronary smooth muscle formation and function, and mutations in *TBX2* are known to play an important role in congenital heart disease, they may also result in malformed coronary vasculature leading to increased susceptibility to CAD.

**Objectives:** Previous studies from our lab on various Indian populations suggest that the Indians are not only unique in their origin but are unique in the genetic etiology of various diseases. Considering the crucial role of *TBX2* in the myocardium formation, we evaluated the genetic variants of the *TBX2* gene to identify association if any with risk of CAD in patients from South India.

**Methods and Results:** We have sequenced all the exons of *TBX2* gene along with intronic boundaries in 700 South Indian CAD subjects along with 400 age and ethnically matched controls. We found two novel missense variants in one of the patients (R350G and R 350L) in exon 5. Another polymorphism (rs761618292) has an A allele frequency of 63% in contrast to world population where it is 0.008%. We also observed different allelic frequencies with respect to other polymorphisms in the *TBX2* gene like rs56298202 C>G, rs35619711 C>G, rs1057987 C>T which are very different from the other populations reported. However the *TBX2* gene polymorphisms did not show any significant association between the patient group and the control group in our study.

**Conclusions:** This is a study carried out for the first time in a cohort of patients from South India on *TBX2* gene. Our results suggest that although there are novel variations in *TBX2* gene they do not contribute to CAD susceptibility among South Indian Population. The frequency of variations observed in *TBX2* gene in our study are so different with respect to other population once again reiterates the fact that Indians as a group are unique in their genetic etiology and manifestation of disease pattern in general.

## **PgmNr 2758: Probands' variants of uncertain significance in established dilated cardiomyopathy genes may influence left ventricular size in family members: Preliminary data from the DCM Precision Medicine Study.**

### **Authors:**

D.D. Kinnamon; J.O. Mead; A. Morales; E. Jordan; L. Salyer; J.R. Cowan; R.E. Hershberger; on behalf of the DCM Consortium and DCM Precision Medicine Study

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Human Genetics, Department of Internal Medicine, The Ohio State University Wexner Medical Center, Columbus, Ohio

---

The ongoing DCM Precision Medicine Study is obtaining exome sequence data on probands with idiopathic dilated cardiomyopathy (DCM) and adjudicating variants in 35 established DCM genes using modified ClinGen criteria. In preliminary data from this study, 43% of probands had variants of uncertain significance (VUSs) as their only clinical results. In order to determine whether these variants might be biologically relevant for DCM, we leveraged the study's family-based design to determine their contribution to variation in echocardiographic endophenotypes in family members aged  $\geq 20$  years. From the first three batches of exome sequenced individuals, we identified all families of self-reported white, non-Hispanic probands in which all of the proband's variants had been adjudicated. VUSs, as well as likely pathogenic (LP) and pathogenic (P) variants, identified in probands were Sanger sequenced in all enrolled family members with available DNA and echocardiographic data. The burden of the proband's variants of a particular class (VUS or LP/P) carried by an individual was defined as the number of variant alleles or zero if no variants of that class were identified in the proband. Associations between the burdens of variants of each class and the individual's sex- and height-specific left ventricular end-diastolic dimension z-score (LVEDDz) or left ventricular ejection fraction (LVEF) were estimated from data on 299 or 314 individuals in 97 or 101 families using a linear mixed model with an additive polygenic effect and pedigree-based kinship estimates. The model likelihood was conditioned on the proband's phenotype to account for single ascertainment through the proband. Because the age-dependent penetrance of DCM suggests that burdens of VUS and LP/P alleles may have age-dependent effects, these were allowed to differ between individuals aged  $\leq 40$  years or  $>40$  years. In an individual over 40, carrying an additional allele of the proband's LP/P variants was associated with an increase in LVEDDz of 1.35 (95% CI: 0.39-2.30;  $p=0.006$ ) and a decline in LVEF of 10.6% (95% CI: 4.9%-16.3%;  $p<0.001$ ). Conditional on this individual's LP/P burden, carrying an additional allele of the proband's VUSs was also associated with an increase in LVEDDz of 0.62 (95% CI: -0.01-1.25;  $p=0.05$ ) and a decline in LVEF of 2.9% (95% CI: -1.0%-6.7%;  $p=0.15$ ). Our preliminary results suggest that probands' VUSs in established DCM genes may be biologically relevant for DCM.



# PgmNr 2759: Sarcopenia-related traits and coronary artery disease: A bi-directional two-sample Mendelian randomization study.

## Authors:

H.M. Liu <sup>1,2</sup>; Q. Zhang <sup>4</sup>; W.D. Shen <sup>1,2</sup>; B.Y. L <sup>1,2</sup>; X. Qiu <sup>1,2</sup>; H.M. Xiao <sup>1,2</sup>; H.W. Deng <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Center of System Biology and Data Information, School of Basic Medical Science, Central South University, No. 172, Tongzipo Road, Yuelu District, Changsha 410013, Hunan Province, P.R. China; 2) Center of reproductive health?School of Basic Medical Science, Central South University, No. 172, Tongzipo Road, Yuelu District, Changsha 410013, Hunan Province, P.R. China; 3) Tulane Center of Bioinformatics and Genomics, Department of Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA 70112, USA; 4) College of Public Health, Zhengzhou University, No.100 Kexue Road, High-Tech Development Zone of States, Zhengzhou, People's Republic of China.

## Background

Sarcopenia was defined as age-related loss of muscle mass and strength, which might lead to coronary artery disease (CAD) in traditional epidemiology study. However, this observational result might be confounded by many environmental factors. Therefore, we used bi-directional Mendelian randomization (MR) to assess the effect of sarcopenia-related traits (whole body lean mass and handgrip strength) on CAD.

## Methods

Inverse variance weighting estimates were obtained by applying genetic variants that predict whole body lean mass and handgrip strength, obtained from the UK Biobank GWAS, to the largest available European case-control study of CAD from Coronary ARtery Disease Genome wide Replication And Meta-analysis (CARDIoGRAM) consortium. Conversely, we also applied genetic variants that predict CAD to UK Biobank summary statistics for genetic association with whole body lean mass and handgrip strength. To further demonstrate the validity of the selected SNPs, we included myopia as negative control in our analysis. MR-Egger regression was performed to assess the horizontal pleiotropic pathway between genetic variants and outcome independent of exposure. As sensitivity analyses we used weighted median, Robust Adjusted Profile Score (RAPS), Mendelian Randomization Pleiotropy RESidual Sum and Outlier (MR-PRESSO), removed proxy SNPs and removed pleiotropic SNPs.

## Results

In this study, inverse variance weighted (IVW) analysis (left handgrip strength,  $b(95\% \text{ CI})$ : -0.54(-0.92, -0.17),  $P=0.005$ ; right handgrip strength,  $b(95\% \text{ CI})$ : -0.56(-0.90, -0.22),  $P=0.001$ ) shown that genetically predicted handgrip strength was significantly related to CAD without bias of horizontal pleiotropy (left handgrip strength, MR-Egger intercept= -0.0008363,  $P=0.807$ ; right handgrip strength MR-Egger intercept= 0.0084743,  $P=0.391$ ). The result was robustly validated in all the sensitivity analyses. However, we did not detect the relationship between genetically instrumented whole body lean mass and CAD in our bi-directional MR analysis.

## Conclusion

Genetically predicted handgrip strength is related to lower CAD risk. Handgrip strength could warrant further investigation as a possible target of intervention for CAD prevention.

## PgmNr 2760: Genetic study of coronary artery disease in Taiwan.

### Authors:

K. Liu <sup>1</sup>; I. Hsieh <sup>2</sup>; M. Wen <sup>2</sup>; C. Wang <sup>2</sup>; C. Chen <sup>2</sup>; H. Chuang <sup>1</sup>; Y. Chen <sup>1</sup>; C. Yang <sup>1</sup>; H. Kao <sup>1</sup>; C. Chen <sup>1</sup>; J. Wu <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) 2.Division of Cardiology, Chang Gung Memorial Hospital, Linkou Medical Center, Taoyuan City, Taiwan;; 3) School of Chinese Medicine, China Medical University, Taichung, Taiwan

---

Coronary artery disease (CAD) is the major cause of death in developed countries. Progressive atherosclerosis and CAD will result in angina pectoris, unstable angina and acute myocardial infarction. With the advancement of recent coronary intervention, CAD still comprises the major causes of mortality. Etiologies of CAD are complex and multifactorial, including family history, obesity, hypertension, smoking, diabetes and dyslipidemia. Family history and genetic factors are most important independent factors among those risk factors. Recent genome wide association studies have identified several genetic loci which are associated with CAD development. How these loci affect CAD and atherosclerosis development are still unclear. After these decades of intensive research on atherosclerosis and CAD, clinical management for the CAD patients has evolved dramatically. For further understand specific susceptibility genes of the CAD patients in our population. We performed a genome-wide association study (GWAS) which was conducted on 877 CAD patients and 1722 controls from a Han Chinese population of Taiwan. In this discovery stage, we Identified 4 single nucleotide polymorphisms (SNPs), *rs10738607*, *s10757278*, *rs1333049* and *rs4977574*, clustered with *CDKN2B* gene at *9p21.3* (all  $p < 10^{-9}$ ). One SNP, *rs671*, located at *12q24.12* in *ALDH2* gene ( $p < 10^{-6}$ ). The gene variants of *CDKN2B* and *ALDH2* have been found to be associated with CAD in a meta-analysis study included 48 studies with a total of 60,801 CAD cases and 123,504 controls[1]. Interestingly, we also identified three SNPs clustered with plaque formation gene ( $p < 10^{-5}$ ) which have not been reported in other GWAS studies of CAD before. This gene belongs cadherin superfamily. Recent studies indicated that a number of cadherins play an important role in cardiovascular disease progression including atherosclerosis and restenosis[2]. We speculated that this plaque formation gene might be a susceptibility gene for CAD specifically in Han Chinese population. We are now enrolling another cohort with 800 CAD patients for the replication. We believe our finding will shed light on the pathogenesis of CAD.

1. Miao, X., et al., *Tissue-specific Network Analysis of Genetic Variants Associated with Coronary Artery Disease*. Sci Rep, 2018. **8**(1): p. 11492.
2. Resink, T.J., et al., *Cadherins and cardiovascular disease*. Swiss Med Wkly, 2009. **139**(9-10): p. 122-34.

## PgmNr 2761: Smoothed moving landmark analysis for the age-dependent effects of DNA methylation on the risk of coronary heart disease.

### Authors:

B. Shi <sup>1,2</sup>; Z. Wang <sup>1,3</sup>; X. Huang <sup>1</sup>; P. Wei <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) biostatistics, M.D Anderson Cancer Center, Houston, TX 77030.; 2) Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, TX 77030; 3) The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030

---

Coronary heart disease (CHD) is the most common type of heart disease, which is the leading cause of mortality among adults in the US. Numerous evidence has been accumulated to show that DNA methylation plays important roles in the development of CHD. This effect is believed to vary by age. However, due to the economic burden, in most studies, DNA methylation is measured only once, at a different age for each subject, which prohibits estimating age-dependent effects of methylation on CHD risk by conventional statistical methods. Here we provide two novel approaches to tackle this analytical challenge, namely separate models and super models. By separate models, we divide all subjects into different groups by their age of methylation measurement, and then estimate the methylation effects in these groups separately using landmark analysis. These estimates along the follow-up ages dynamically approximate the effect profile of methylation on CHD onset risk. On the other hand, the super models combine the above separate landmark analyses together and borrow information between adjacent age intervals to construct a smoother and more stable profile of the methylation effects on CHD risk. Simulation studies confirm these advantages of the super models. We apply both approaches to the Framingham Heart Study of 1540 subjects with CHD event data and methylation profiling at a single clinical visit, for illustration and demonstration of the comparisons between them. Our novel analysis reveals that methylation levels at several CpG sites show declining age-dependent effects on the CHD risk.

## **PgmNr 2762: Causality of blood lipoprotein cholesterol on coronary artery disease from multivariate Mendelian randomization analysis.**

**Authors:**

H. Xu; V. George

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dept of Population Health Sciences, Augusta University, Augusta, Georgia.

---

Instrumental variables are critical for valid and efficient Mendelian randomization (MR) analysis. We developed a genetic variant selection method from the empirical selection approach of Do et al. (2013). Using this method, we selected 338 single-nucleotide-polymorphisms (SNPs) from a meta-analysis data set of genome-wide association study (GWAS) of lipid and coronary heart disease (CAD) and 363 SNPs from another joint GWAS meta-analysis data. From the multivariate MR analysis, we found that low density lipoprotein cholesterol (LDL-c) was strongly associated with increasing risk for CAD, high density lipoprotein cholesterol (HDL-c) was strongly associated with decreasing risk for CAD, and triglycerides (TG) were not associated with risk for CAD. We performed a simulation study accounting for the linkage disequilibrium among the selected SNPs and the pleiotropy among lipid components. The results from simulation study confirmed the result from the multivariate MR analysis for the causal relationship of lipoprotein cholesterol and CAD.

## PgmNr 2763: Genome-wide interaction analysis in more than 500K European individuals identifies novel blood pressure genetic markers with an age dependent effect.

### Authors:

C. Gao; A. Marcketta; B. Ye; C. O'Dushlaine; N. Banerjee; J. Overton; J. Reid; A. Baras; M. Ferreira; J. Marchini; G. Abecasis; A. Shuldiner; S. McCarthy; Geisinger-Regeneron DiscovEHR Collaboration

[View Session](#) [Add to Schedule](#)

**Affiliation:** Regeneron Genetics Center, Tarrytown, New York.

---

Hypertension is a common epidemic that is a strong risk factor for many health issues. Epidemiology studies show that aging is one of the strongest risk factors for hypertension, i.e. the prevalence increases from 7.5% (18-39 yrs) to 63.1% (60+ yrs). Genome-wide association studies have identified numerous blood pressure genetic signals, yet little has been done to explore the genetic architecture of aging-related hypertension. In this study, we performed a genome-wide interaction analysis to search for blood pressure genetic signals with age-dependent effects. Four blood pressure phenotypes were included, i.e. systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure (MAP), and pulse pressure (PP). Ten million imputed markers ( $MAF \geq 0.5\%$ ) were analyzed in 539K European individuals from UK biobank ( $N=459,333$ ) and DiscovEHR ( $N=79,761$ ). To search for coding variants with potentially larger genetic effects, 1.6 million exome sequencing markers from 224,249 individuals (141,527 from UKB and 82,722 from DiscovEHR) were also analyzed. Age was modeled as an interaction variable and interaction was tested using PLINK, adjusting for sex, BMI, ten principle components, and cohort specific covariates. After removing the HLA region, 1DF interaction model identified 15 independent signals with genome-wide significance ( $P < 5E-08$ ). A common SNP ( $AAF=45\%$ ), located in the first intron of gene *ELN* (elastin), was strongly associated with PP with an age-dependent effect ( $P_{1DF}=9.76E-10$ ). This SNP has a modest eQTL signal with a non-coding RNA (RP11-731K22.1,  $P=3.60E-05$ ), which is located upstream of *ELN*. The A allele was associated with a 0.09 mmHg lower ( $P=0.14$ ) and 0.50 mmHg higher ( $P=1.08E-07$ ) PP in young ( $<45$  years) and old ( $>65$  years) groups, respectively. Previous studies reported a genome-wide significant association between PP and a *ELN* downstream SNP (7:74076882:A:G). Further analyses suggest that this SNP had no significant interaction with age ( $P_{1DF}=0.70$ ) and both SNPs are in poor LD ( $R^2=0.0034$ ). Taken together, the newly identified *ELN* SNP is associated with PP with an age-dependent effect and is independent from the previously reported signal. In summary, genome-wide interaction analysis in 539K European individuals identified genetic signals with age-dependent effects. This study provides biological insights underlying age-related hypertension and suggests potential mechanisms and therapeutic targets for age-related hypertension.

## PgmNr 2764: Additive and multiplicative gene-lifestyle interactions in coronary artery disease in the UK Biobank.

### Authors:

Y. Huang <sup>1</sup>; Q. Hui <sup>1</sup>; V. Vaccarino <sup>1</sup>; A.A. Quyyumi <sup>2</sup>; M. Gwinn <sup>1</sup>; Y. Hu <sup>3</sup>; Y.V. Sun <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 2) Division of Cardiology, Department of Medicine, School of Medicine, Emory University, Atlanta, GA; 3) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA; 4) Department of Biomedical Informatics, School of Medicine, Emory University, Atlanta, GA

---

Coronary artery disease (CAD) is the pre-eminent cause of death. Both genetic and lifestyle factors such as cigarette smoking and physical activity (PA) contribute to development of CAD. Over 160 loci have been linked with risk of CAD in genome-wide association studies. However, whether genetic risk of CAD can be modified by lifestyle factors remains unclear. We assessed interactions between CAD polygenic risk score (PRS) and lifestyle factors (i.e., smoking and PA) on both additive and multiplicative scales in a European ancestry sample of 307,147 adults from the UK Biobank. Primary incident CAD events were identified using in-patient hospitalization data and national death registry data. Four CAD-PRSs were constructed including: 1) an overall PRS based on 161 published CAD SNPs; 2) a blood pressure (BP) sub-PRS based on 26 SNPs; 3) a lipids sub-PRS based on 17 SNPs; and 4) a body mass index (BMI) sub-PRS based on 16 SNPs. Cox proportional hazards models were used to assess the association of CAD-PRSs with incident CAD as well as the interaction effects. Multiplicative interaction was assessed by the product term in each model, and additive interaction was assessed by calculating relative excess risk due to interaction (RERI). Comparing to the low CAD genetic risk group (quintile 1), those with intermediate genetic risk (quintile 2-4) and high genetic risk (quintile 5) had a 1.43-fold (95% CI: 1.35-1.52) and 2.16-fold (95% CI: 2.02-2.31) increased risk of CAD, respectively. Lipids sub-PRS had the strongest association with incident CAD, compared with BP sub-PRS and BMI sub-PRS. Overall, no significant multiplicative interaction was observed between any pair of CAD PRS and lifestyle factors. For the comprehensive CAD-PRS, synergistic additive interaction was observed with current smoking in both intermediate (RERI: 0.394, 95% bootstrap interval: 0.097, 0.729) and high (RERI: 1.051, 95% bootstrap interval: 0.615, 1.497) genetic risk groups, meaning the CAD risk due to smoking was increased in groups at higher genetic risk. An additive interaction with smoking was also identified for the BP sub-PRS but not lipids or BMI sub-PRSs. No additive interaction was observed between any pair of CAD PRS and PA. In this PRS-based interaction study, we identified synergistic excess risk among people with high lifestyle and genetic risk for CAD, which can provide novel mechanistic insights on CAD risk and may lead to innovative intervention strategy.

## PgmNr 2765: Large-scale multi-ancestry gene-environment interaction screenings point towards different genetic mechanisms by population and exposure.

### Authors:

V. Laville <sup>1</sup>; T. Majarian <sup>2</sup>; Y.J. Sung <sup>3</sup>; M.F. Feitosa <sup>4</sup>; D. Chasman <sup>5</sup>; A.R. Bentley <sup>6</sup>; C.N. Rotimi <sup>6</sup>; L.A. Cupples <sup>7,8</sup>; P.S. de Vries <sup>9</sup>; M.R. Brown <sup>9</sup>; A.C. Morrison <sup>9</sup>; A.T. Kraja <sup>4</sup>; M. Province <sup>4</sup>; K. Schwander <sup>3</sup>; J. Gauderman <sup>10</sup>; D.C. Rao <sup>3</sup>; A. Manning <sup>2,11</sup>; H. Aschard <sup>1,12</sup>; CHARGE Gene-Lifestyle Interaction Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computational Biology - USR 3756 CNRS, Institut Pasteur, 75724 Paris, France; 2) Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA, 02114, USA; 3) Division of Biostatistics, Washington University, St. Louis, MO, 63110, USA; 4) Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, 63108-221, USA; 5) Division of Preventive Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, 2215, USA; 6) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, 20892, USA; 7) Department of Biostatistics, Boston University School of Public Health, Boston, MA, 2118, USA; 8) Framingham Heart Study, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, 20982, USA; 9) Department of Epidemiology, Human Genetics, and Environmental Sciences, The University of Texas School of Public Health, Houston, TX, 77030, USA; 10) Division of Biostatistics, Department of Preventive Medicine, University of Southern California, Los Angeles, CA, 90032, USA; 11) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA; 12) Program in Genetic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, Boston, MA, 02115

---

Improving our understanding of the genetic architecture and the role of gene-by-environment (GxE) interactions in complex multifactorial traits has become a major goal for the genetics community. Recently, the Gene-Lifestyle Interactions Working Group conducted a range of multi-ancestry genome-wide interaction studies (GWIS) focusing on interaction between genetic variants and two exposures namely alcohol consumption and smoking habits on lipids (triglycerides (TG), high and low-density lipoproteins) and blood pressure (diastolic and systolic blood pressure, mean arterial pressure, and pulse pressure) in up to 610,475 individuals from multiple ancestry groups.

In this study, we conducted series of cross-study analyses using the GWIS summary statistics. We observed important similarity and specificity of GxE effects across phenotypes and ancestries. For example, some interactions were found only in the African-ancestry population and involved genetic variants specific to this population. Meanwhile, several Gene-by-Alcohol interactions were identified only in the Asian-ancestry population for variants common in all populations, suggesting a potential contribution of the genetic or environmental background. Comparisons between the different GWIS also stressed the potential importance of specific phenotype-exposure pairs for the detection of new signal, such as lipid-drinking and blood pressure-smoking. We also highlighted small but persistent negative interaction effects between TG associated variants and current smoking.

To explore further the impact of these exposures on genetic mechanisms, we estimated the genetic



heritability for all traits in unexposed and exposed individuals. In each group, we also partitioned heritability by cell-type and looked for heritability enrichment by cell-types. We found substantial differences in both the overall heritability and the enrichment by cell type. These analyses suggest important differences in the genetic responses on phenotypes to the environmental stress. For example, we noted a reduced heritability of most phenotypes among smokers. In lipids traits, this reduction appears to be driven by a smaller contribution of liver-related variants among those individuals.

Our in-depth analyses of the GWIS results emphasize the importance of performing ancestry-specific analyses and suggest the existence of exposure-specific pathways, opening new paths for future in-silico and functional gene-environment interaction studies.

## PgmNr 2766: A whole-genome approach reveals novel genetic and non-genetic variance components modulated by lifestyle for cardiovascular health.

### Authors:

Zhou <sup>1</sup>; J. van der Werf <sup>2</sup>; K. Carson-Chahhoud <sup>1</sup>; G. Ni <sup>3</sup>; J. McGrath <sup>4,7</sup>; E. Hypponen <sup>1,5,6</sup>; H. Lee <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Australian Centre for Precision Health, University of South Australia, Adelaide, South Australia, Australia; 2) School of Environmental and Rural Science, University of New England, Armidale, New South Wales, Australia; 3) Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia.; 4) Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia; 5) South Australian Health and Medical Research Institute, Adelaide, Australia; 6) Population, Policy and Practice, UCL Great Ormond Street Institute of Child Health, London, UK; 7) Queensland Centre for Mental Health Research, The Park Centre for Mental Health, Wacol, Queensland, Australia

---

Both genetic and non-genetic factors can predispose individuals to cardiovascular health risk. Finding ways to alter their predispositions is important for cardiovascular disease (CVD) prevention. Here, using a novel whole-genome framework, we estimated genetic and non-genetic effects on—hence their predispositions to—cardiovascular health risk and determined whether they vary with respect to lifestyle factors. We performed analyses on the Atherosclerosis Risk in Communities Study (ARIC) and validated findings using the UK Biobank (UKBB), where cardiovascular health risk was measured using 23 traits in the ARIC and 8 in the UKBB, including body mass index, resting heart rate, white blood cell count and diastolic blood pressure; and lifestyle choice was captured in aspects such as physical activity, smoking and alcohol intake. In both datasets, we found that physical activity alters genetic effects on white blood cell count and HDL cholesterol level but non-genetic effects on heart rate, BMI and waist-to-hip ratio. Alcohol intake alters both genetic and non-genetic effects on BMI, while smoking alters non-genetic effects on heart rate and white blood cell count. In addition, saturated fat intake modifies genetic effects on BMI. These results highlight the relevance of lifestyle changes for CVD prevention. Using the whole-genome framework, we also stratified individuals into groups, according to their genetic predispositions, and showed that lifestyle effects on cardiovascular health risk could take on drastically different directions across groups, implying the need for individualizing lifestyle changes for CVD prevention, hence the importance of precision medicine. Finally, we showed that neglecting lifestyle modulation of genetic and non-genetic effects, on average, reduced SNP heritability estimates of cardiovascular traits by a small yet significant amount, primarily owing to overestimation of residual variance. Thus, current SNP heritability estimates for cardiovascular traits, which commonly do not consider modulating effects of lifestyles, are likely underestimated.

## PgmNr 2767: Genetic basis of hypertrophic cardiomyopathy in patients from the São Miguel Island (Azores, Portugal).

### Authors:

L. Mota-Vieira <sup>1,2</sup>; M. Baixia <sup>3,4</sup>; T. Pereirinha <sup>1</sup>; S. Castedo <sup>3,4</sup>; C. Machado <sup>5</sup>; L. Oliveira <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Genetics and Pathology Unit, Hospital of Divino Espirito Santo of Ponta Delgada, EPER, Ponta Delgada, Azores Islands, Portugal; 2) BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisboa, Portugal; 3) IPATIMUP - Instituto de Patologia e Imunologia Molecular, Universidade do Porto, Portugal; 4) i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal; 5) Cardiology Department, Hospital of Divino Espirito Santo of Ponta Delgada, EPER, Ponta Delgada, Azores Islands, Portugal

---

**Background:** Hypertrophic cardiomyopathy (HCM) is the most common inherited heart disease (approx. 1 in 500 individuals), characterized by left ventricular hypertrophy ( $\geq 15$  mm) that is not solely explained by abnormal loading conditions. To date, about 1500 mutations in genes encoding sarcomere proteins have been identified in HCM patients, but their mutation spectrum may be variable across different populations. **Objective and method:** The present study aimed to understand the genetic causes of HCM on the Azorean island of São Miguel (Portugal). To that end, we investigated a cohort of 18 unrelated adult patients (12 males, 6 females) who fulfilled criteria for HCM. In brief, we screened 29 cardiomyopathy-associated genes using targeted next-generation sequencing. Variants were filtered and classified according to standards and guidelines of the American College of Medical Genetics and Genomics and to recommendations of the ClinGen Inherited Cardiomyopathy Expert Panel. Pathogenic or likely pathogenic variants were confirmed by Sanger sequencing, *in silico* effect prediction analysis and published data from functional studies.

**Results:** We found that 11 (61.1%) patients harbored causative mutations in 3 genes encoding sarcomere proteins. Among them, 7 patients (pts) had pathogenic variants in *MYH7* (NM\_000257.2: c.1750G>C, p.Gly584Arg in 6 pts; c.1750G>A, p.Gly584Ser in 1 pt); 2 patients had likely pathogenic variants in *MYH7* (c.2389G>A, p.Ala797Thr), 1 in *MYBPC3* (NM\_000256.3: c.1484G>A, p.Arg495Gln) and 1 in *TNNT2* (NM\_001276345.1: c.842A>T, p.Asn281Ile). We also detected two variants of uncertain significance in *JPH2* (NM\_020433.4) gene: 1 patient carried the c.590G>T, p.Arg179Leu, the other one the c.692G>A, p.Arg231Gln. The remaining 5 patients had no mutation in the genes sequenced. All patients were heterozygous, except one (male) who was apparent homozygous for the *TNNT2*: c.842A>T, p.Asn281Ile. To clarify if this variant is homozygous or hemizygous, we are now performing MLPA analysis. Moreover, we will apply an extended NGS gene panel to investigate the genetic causes of HCM in the 7 (38.9%) patients without causative mutations identified.

**Conclusions:** On the São Miguel Island germline mutations in the *MYH7* gene account, at least, for 50% of unrelated patients with HCM. The identification of the causative mutations in HCM is essential since it confirms the diagnosis, enables genetic counseling to families and could allow individual therapy planning.

## PgmNr 2768: Targeted genome sequencing identifies multiple rare variants in Caveolin-1 (CAV1) associated with obstructive sleep apnea trait.

### Authors:

J. Liang <sup>1</sup>; B. Cade <sup>2,3,4</sup>; X. Zhu <sup>1</sup>; S. Redline <sup>2,3,5</sup>; TOPMed Sleep Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, 44106, USA; 2) Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA 02115, USA; 3) Division of Sleep Medicine, Harvard Medical School, Boston, MA, 02115, USA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA; 5) Division of Pulmonary, Critical Care and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215, USA

---

Obstructive sleep apnea (OSA) is a common disorder associated with increased risk for cardiovascular disease, diabetes, and premature mortality. There is strong clinical and epidemiologic evidence supporting the importance of genetic factors influencing OSA, but limited data implicating specific genes. Disease severity is most commonly measured by the apnea-hypopnea index (AHI), which quantifies the number of complete and partial episodes of upper airway obstruction per hour of sleep. We report genetic associations of AHI using high depth genomic sequencing data from the NHLBI Trans-Omics for Precision Medicine (TOPMed) project and focus on the Cleveland Family Study (CFS), a cohort with enrichment for OSA. We identified an associated 19 Mb linkage region on chromosome 7 in European Americans and examined rare variants that contributed to the observed linkage evidence in gene-based analysis. We identified 18 candidate genes ( $P < 0.05$ ) associated with AHI under the linkage peak in CFS. Follow-up analysis of these variants in the 18 candidate genes in 4 independent TOPMed studies identified 21 non-coding rare variants in Caveolin-1 (CAV1) significantly associated with AHI ( $P = 5.4 \times 10^{-7}$ ) in 3,259 European Americans. These non-coding variants significantly contribute to the CAV1 expression level in skeletal muscle ( $P = 3.8 \times 10^{-4}$ ) in GTEx data. Caveolin-1 is a membrane scaffolding protein that is essential in the formation of plasma membrane lipid rafts and mediates cholesterol trafficking and regulates several signaling molecules including TGF-beta, TLR4 and endothelial nitric oxide synthase (eNOS). It is implicated in disorders associated with OSA, including metabolic and endothelial dysfunction, immune regulation and inflammation and lung remodeling. Common CAV1 variants were associated with severe OSA in a small case-control study. Our results suggest that caveolin-1 may play a significant role in OSA pathogenesis.

## PgmNr 2769: Genetic architecture of cardiac repolarization dynamics.

### Authors:

S. van Duijvenboden <sup>1,2</sup>; J. Ramirez <sup>2,1</sup>; M. Orini <sup>2</sup>; A. Tinker <sup>2</sup>; P.D. Lambiase <sup>1</sup>; P.B. Munroe <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Cardiovascular Science, UCL, LONDON, United Kingdom; 2) Clinical Pharmacology, Queen Mary University of London, United Kingdom

---

**BACKGROUND:** Cardiac repolarization is a key cellular process and disruptions are linked with increased cardiovascular risk. The genetic basis of cardiac repolarization has been studied extensively by measuring the QT interval from the electrocardiogram (ECG) under resting conditions. However, repolarization varies with heart rate and abnormalities of these dynamics have shown to be important in modulating cardiovascular risk as well. This work investigates the genetic basis and the underlying biological mechanisms of repolarization dynamics during exercise and recovery alongside resting QT interval.

**OBJECTIVES:** To identify single nucleotide variants (SNVs) associated with ECG markers of repolarization dynamics during exercise and recovery and to study the overlap with resting QT.

**METHODS:** Exercise ECG data from 51,574 subjects of the UK Biobank study were analysed for two established markers of repolarization dynamics: QT dynamics, defined as the slope of the QT/RR profile and the T-wave morphology restitution (TMR), which quantifies the spatio-temporal heterogeneity of ventricular repolarization. Both markers were measured during exercise and recovery. We conducted genome-wide association studies (GWASs) for each trait and resting QT interval. Bioinformatics analyses were performed to annotate SNVs and identify candidate genes.

**RESULTS:** GWASs identified 20 loci for QT dynamics and indicated a heritability up to 10%. There was substantial overlap of shared loci between QT dynamics and resting QT. TMR was less heritable (~5%), with discovery of 12 genetic loci, of which 9 were also associated with resting QT interval and 7 with QT dynamics. Genetic variants were found at long-QT syndrome loci: KCNH2, KCNJ2, SCN5A and KCNQ1. These channel proteins underlie the major repolarising ventricular potassium currents and depolarising sodium current. Variations in these currents might lead to changes in the T-wave morphology and this is consistent with known physiology. Other candidate genes at TMR loci include: NOS1AP, known to modulate QT interval duration and risk of arrhythmias in the long QT syndrome and CAMK2D which regulates calcium dynamics.

**CONCLUSION:** This systematic investigation indicates that repolarization dynamics has a genetic predisposition. The important substantial genetic overlap with resting QT loci, suggests that biological mechanisms governing repolarization during steady state conditions are also important in modulating repolarization dynamics.

## PgmNr 2770: Combined-phenotype study of seven RBC traits in the PAGE study highlights the benefits of ancestrally diverse study populations.

### Authors:

C.J. Hodonsky <sup>1</sup>; A.B. Baldassari <sup>1</sup>; S.A. Bien <sup>2</sup>; H.M. Highland <sup>1</sup>; C.M. Sitlani <sup>3</sup>; M. Graff <sup>1</sup>; G. Wojcik <sup>4</sup>; R. Tao <sup>5</sup>; S. Buyske <sup>6</sup>; L.A. Hindorff <sup>7</sup>; Y. Li <sup>8</sup>; D. Lin <sup>9</sup>; R.J.F. Loos <sup>10</sup>; A.P. Reiner <sup>2,3</sup>; K.E. North <sup>1,8</sup>; C.L. Kooperberg <sup>2</sup>; C.L. Avery <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, UNC - Chapel Hill, Chapel Hill, NC; 2) Public Health, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Biostatistics, University of Washington, Seattle, WA; 4) Stanford University, Stanford, CA; 5) Biostatistics, Vanderbilt University, Nashville, TN; 6) Biostatistics, Rutgers University, Piscataway, NJ; 7) National Human Genome Research Institute, Bethesda, MD; 8) Genetics, UNC - Chapel Hill, Chapel Hill, NC; 9) Biostatistics, UNC - Chapel Hill, Chapel Hill, NC; 10) Icahn School of Medicine at Mount Sinai, New York, NY

Quantitative red blood cell (RBC) traits are highly polygenic clinically relevant traits, with nearly 500 reported GWAS loci. The majority of RBC trait GWAS have been performed in European- or East Asian-ancestry populations, despite evidence that rare or ancestry-specific variants contribute substantially to RBC trait heritability. Recently developed combined-phenotype methods which leverage trait correlation to improve statistical power have not yet been applied to these traits. Here we performed a combined-phenotype analysis of seven correlated quantitative RBC traits in a multi-ethnic study population.

Based on correct type I error rate and ability to evaluate opposite directions of effect by trait, we selected a simulation-based adaptive sum of powered scores (aSPU) method to test overall associations between ~22 million SNPs and univariate summary statistics from GWAS of seven RBC traits. A maximum of 68,633 study participants were included (42% MEGA-genotyped, 24% African American, 30% Hispanic/Latino, 43% European American, 76% female). Step-wise conditional analysis was performed to identify all independent, genome-wide-significant ( $P < 5 \times 10^{-9}$ ) lead SNPs.

Thirty-nine loci contained at least one significant association signal, with lead SNPs at five loci significantly associated with three or more RBC traits. A majority of the lead SNPs were common ( $MAF > 5\%$ ) across all ancestral populations. 19 additional independent association signals were identified at seven loci (*HFE*, *KIT*, *HBS1L/MYB*, *CITED2/FILNC1*, *ABO*, *HBA1/2*, and *PLIN4/5*). For example, the *HBA1/2* locus contained 14 conditionally independent association signals, 11 of which were specific to African and Amerindian ancestries. One variant at this region was common in all ancestries, but demonstrated a narrower locus in African Americans than European Americans or Hispanics/Latinos. GTEx eQTL analysis of all independent lead SNPs (excluding ancestry-specific SNPs not available in the GTEx database) yielded 31 significant associations in relevant tissues, over half of which were not at the gene immediately proximal to the lead SNP.

Conditional analysis revealed a complex LD structure at the *HBA1/2* locus, which also contains known

causal variants for monogenic blood disorders originating in malaria-endemic regions. Our work demonstrates the continued importance of including ancestrally diverse study populations when attempting to identify and characterization complex trait loci.

## PgmNr 2771: Evaluating genetic support for lipid-modifying drug targets.

### Authors:

M. Gordillo-Maranon <sup>1</sup>; C. Finan <sup>1</sup>; A.F. Schmidt <sup>1,2</sup>; A.D. Hingorani <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Cardiovascular Science, University College London, London, United Kingdom; 2) Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, the Netherlands

---

Despite considerable interest including several drug development programs, the role of high-density lipoprotein cholesterol (HDL-C) in coronary heart disease (CHD) is still unclear.

In observational studies, circulating low-density lipoprotein cholesterol (LDL-C) exhibits a positive association with CHD risk. Treatment trials of LDL-lowering drugs (statins, ezetimibe and PCSK9 inhibitors), each targeting different mechanisms (HMGCR, NPC1L1 and PCSK9 respectively), show consistent reductions in CHD risk. Naturally occurring genetic variation in the genes encoding these targets, and at other loci identified by the Global Lipid Genetic Consortium, accurately instrument causal effects of LDL-C on CHD using data from the CardiogramPlusC4D Consortium in Mendelian randomisation analysis.

By contrast, circulating HDL-C exhibits a negative association with CHD risk in observational studies, but treatment trials of several HDL-C raising agents e.g. targeting CETP and GPR109A have been negative, causing uncertainty about whether this association is causal and if the development of drugs that raise HDL-C is a useful therapeutic strategy for CHD prevention. Mendelian randomisation analyses of HDL-C in CHD have also been equivocal.

Here, we use Mendelian randomisation for drug target validation (*cis*-MR) leveraging genetic variants associated with HDL-C, LDL-C and triglycerides (TG) in-and-around loci encoding drugged and potentially druggable targets. We identify a number drug targets that exhibit similar effects to existing lipid-lowering targets. We also demonstrate that prior failure of certain CETP inhibitor drugs are molecule rather than target failures, and that modification of certain drugged/druggable proteins that affect HDL-C and TG metabolism is likely to offer new effective therapeutic mechanisms for CHD prevention.



## PgmNr 2772: Kidney function and blood pressure: A bidirectional Mendelian randomization study.

### Authors:

Z. Yu <sup>1</sup>; J. Coresh <sup>1</sup>; M.E. Grams <sup>1,2</sup>; G. Qi <sup>3</sup>; E. Boerwinkle <sup>4,5</sup>; H. Snieder <sup>6</sup>; A. Teumer <sup>7</sup>; A. Köttgen <sup>1,8</sup>; C. Pattaro <sup>9</sup>; N. Chatterjee <sup>3,10</sup>; A. Tin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Division of Nephrology, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 4) Human Genetics Center, University of Texas Health Science Center, Houston, TX; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 7) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 8) Institute of Genetic Epidemiology, Medical Center and Faculty of Medicine, University of Freiburg, Freiburg, Germany; 9) Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy; 10) Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD

### BACKGROUND

Reduced kidney function has been associated with hypertension and vice versa. It remains unclear if the association is causal. We used summary statistics from large-scale genome-wide association studies (GWAS) to conduct two-sample bidirectional Mendelian randomization (MR) analyses to examine the potential causal relations between kidney function and blood pressure.

### METHODS

Kidney function was represented by: i) estimated glomerular filtration rate (eGFR) using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) serum creatinine equation and ii) blood urea nitrogen (BUN). We used index SNPs from the CKDGen Consortium European-ancestry GWAS results (eGFR, n=567,460; BUN, n=243,031) and the UKB-ICBP Consortium GWAS results of systolic and diastolic blood pressure (SBP and DBP, n=~1 million). Both eGFR and BUN have marker-specific determinants unrelated to kidney function. To reduce bias, we selected SNPs that are likely related to kidney function rather than to the biomarker metabolism by requiring that the index SNPs of eGFR had opposite direction of association with BUN at Bonferroni corrected significance threshold, and vice versa for BUN index SNPs. To address the observed widespread pleiotropy between kidney function and blood pressure, we used pleiotropy-robust MR methods for primary analyses: weighted mode and MRMix.

### RESULTS

Of the reported index SNPs (eGFR: 256, BUN: 75), 40 eGFR and 26 BUN independent SNPs were finally selected as instruments for kidney function. For blood pressure, 265 SBP and 259 DBP independent SNPs were used as instruments without filtering. We observed significant causal effects of reduced kidney function on higher blood pressure (weighted mode estimate per 1 SD lower in ln(eGFR): 0.18 SD higher in SBP,  $p=1.10 \times 10^{-6}$  and 0.16 SD higher in DBP,  $p=3.03 \times 10^{-5}$ ). Similar results were observed for BUN to SBP and DBP while results including all reported index SNPs were not significant. The

causal estimates of blood pressure on kidney function were not significant albeit in the direction observed in epidemiological studies.

## **CONCLUSIONS**

MR analyses, with SNP filtering for kidney effects excluding marker effects, support causal effects of lower kidney function to higher blood pressure while inconclusive on the causal effects of higher blood pressure to lower kidney function. These findings help the interpretation of observational epidemiologic associations.

# PgmNr 2773: Comprehensive investigation of serum biomarkers and their causal role in cardiometabolic disease.

## Authors:

D. Zanetti <sup>1</sup>; S. Gustafsson <sup>2</sup>; E. Ingelsson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Division of Cardiovascular Medicine, CVI Stanford Cardiovascular Institute, Palo Alto, California.; 2) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Sweden.

---

## Background

Serum biomarkers representing glucose homeostasis, lipid metabolism, kidney and liver function, have been associated with cardiometabolic traits in observational studies, but the nature of these associations is incompletely understood.

## Methods

We performed multivariable-adjusted regression models to assess associations of 27 serum biomarkers with 7 cardiovascular traits in up to 469,628 participants of the UK Biobank. Further, we assessed the causal relationships of these biomarkers with the same outcomes using a two-sample Mendelian randomization (MR) approach combining data from UK Biobank with summary statistics from external GWAS meta-analyses. We applied a Bonferroni-corrected alpha threshold of  $2.7 \times 10^{-4}$  (adjusting for 27 biomarkers and 7 outcomes).

## Results

In our observational analyses, we replicated and extended associations implicated in previous, smaller studies. After multiple testing correction, we found a total of 21, 15, 9, 24, 26, 24 and 26 biomarkers strongly associated with atrial fibrillation, coronary artery disease (CAD), ischemic stroke, T2D, systolic blood pressure, body mass index (BMI) and waist-to-hip ratio (WHR); respectively. The MR analyses confirmed strong evidence of previously suggested causal associations for a number of glucose- and lipid-related biomarkers with T2D and CAD. In addition, we detected a causal role of alanine aminotransferase in the development of T2D and hypertension ( $\beta$ , 0.46 and 0.06, per SD change in exposure and outcome;  $P=4.8 \times 10^{-11}$  and  $P=2.7 \times 10^{-5}$ ). We highlighted a causal association of triglyceride levels with WHR ( $\beta$ , 0.18, per SD change in exposure and outcome;  $P=5.0 \times 10^{-15}$ ); and albumin with BMI ( $\beta$ , -0.02, per SD change in exposure and outcome;  $P=1.7 \times 10^{-5}$ ). With 99% statistical power, we excluded a causal effect of at least 0.1 SD change in vitamin D with CAD. Sensitivity analyses including MR-PRESSO provided consistent results and indicated no significant directional horizontal pleiotropy for any of the reported causal associations.

## Conclusion

Our study confirm a causal role of several serum biomarkers related with glucose and lipid metabolism with CAD and T2D. In addition, we highlight previously unknown possible causal links between liver function and T2D and hypertension. This knowledge will provide important insights regarding the etiological understanding of these diseases, likely accelerating new prevention strategies.

## PgmNr 2774: Genetic determinants in relation to metabolism of the Mediterranean diet and association with coronary heart disease risk.

### Authors:

J. Li <sup>1,2</sup>; T. Zhao <sup>1</sup>; W. Chung <sup>1</sup>; C. Turman <sup>1</sup>; M. Guasch-Ferré <sup>1,3</sup>; F. Hu <sup>1,2,3</sup>; L. Liang <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Department of Nutrition, Harvard T.H. Chan School of Public Health, Boston, MA; 3) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 4) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA

---

The Mediterranean diet is beneficial for coronary heart disease (CHD) prevention, possibly through its effect on human metabolism. In our previous study, leveraging metabolomic profiling in multiple cohorts, we identified a plasma metabolic signature (consist of 67 metabolites) that is robustly associated with adherence to the Mediterranean diet and predicts CHD risk. However, a large proportion of the variation in the metabolic signature cannot be explained by dietary intakes. We hereby performed a genome-wide association study (GWAS) for the metabolic signature in 1,925 participants with both metabolomic and genotype data, to identify genetic determinants related to the Mediterranean diet metabolism. Heritability estimation suggested that 13.1% of the total variance in the metabolic signature was explained by genetics ( $P=0.27$ ). We identified two loci associated with the metabolic signature at the genome-wide significance and one at suggestive significance; these included the fatty acid desaturase gene cluster *FADS1-3* (led by rs174541 on Chr11,  $P=3.7\times10^{-13}$ ), a locus near D-amino acid oxidase activator *DAOA* (led by indel chr13:105803942,  $P=7.3\times10^{-11}$ ), and *LINC01187* (led by rs9313494 on Chr5,  $P=1.8\times10^{-7}$ ). We further identified two loci influencing the metabolic signature through their interactive effects with adherence to the Mediterranean diet (*ABCC1-6*, led by rs115523483 on Chr16,  $P=1.9\times10^{-8}$ ; *LOC101928516*, led by rs525923 on Chr6,  $P=3.6\times10^{-8}$ ). Top variants on *FADS1-3* showed significant associations with expressions of *FADS1*, 2, and 3, and *TMEM258* in multiple tissues including brain, liver, muscle, heart, and arteries ( $P<2\times10^{-5}$ ). The lead variant on *LINC01187* was associated with expressions of lincRNA 1187 in lung, adipose tissue, and whole blood ( $P<2\times10^{-3}$ ). Taking advantage of recently published GWAS summary statistics for CHD (122,733 cases, 424,528 controls), a Mendelian randomization analysis showed that the genetic component of the metabolic signature was associated with an 11% reduction in CHD risk (based on 58 variants;  $P<0.001$ ). In conclusion, the metabolic signature of the Mediterranean diet reflects the combined effects of both dietary adherence and metabolic response determined by genetics; and may potentially point to causal metabolic pathways related to CHD development. The efficacy of the Mediterranean diet on CHD prevention may vary depending on individuals' metabolic capacity, especially metabolism of fatty acids and amino acids.

## PgmNr 2775: Electrocardiographic P-wave duration reveals diverse genetic mechanisms of atrial fibrillation.

### Authors:

L. Weng<sup>1,2</sup>; A. Hall<sup>1,2</sup>; S.H. Choi<sup>2</sup>; S. Jurgens<sup>2</sup>; D. Arking<sup>3</sup>; P. Munroe<sup>4,5</sup>; P. Huang<sup>1</sup>; N. Sotoodehnia<sup>6,7,8</sup>; A. Isaacs<sup>9</sup>; H. Lin<sup>10,11</sup>; J. Kanter<sup>12</sup>; M. Müller-Nurasyid<sup>13,14,15,16</sup>; P. van der Harst<sup>17</sup>; J. van Setten<sup>18</sup>; H. Lin<sup>19</sup>; D. Mook-Kanamori<sup>20</sup>; B. Stricker<sup>21</sup>; M. Dörr<sup>22,23</sup>; A. Teumer<sup>23,24</sup>; C. Kooperberg<sup>25</sup>; E. Benjamin<sup>11,26</sup>; K. Lunetta<sup>11,27</sup>; P. Ellinor<sup>1,2,28</sup>; S. Lubitz<sup>1,2,28</sup>; on behalf of the CHARGE Consortium EKG Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; 2) Program in Medical population genetics, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 4) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; 5) National Institute for Health Research, Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, London, UK; 6) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; 7) Cardiology Division, University of Washington, Seattle, WA, USA; 8) Department of Epidemiology, University of Washington, Seattle, WA, USA; 9) CARIM School for Cardiovascular Diseases, Maastricht Center for Systems Biology (MaCSBio), Dept. of Biochemistry, and Dept. of Physiology, Maastricht University, Maastricht, Netherlands.; 10) Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA; 11) National Heart Lung and Blood Institute's and Boston University's Framingham Heart Study, Framingham, MA, USA; 12) Laboratory of Experimental Cardiology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 13) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 14) Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany; 15) Department of Internal Medicine I (Cardiology), Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, Germany; 16) DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany; 17) University of Groningen, University Medical Center Groningen, Department of Cardiology, Groningen, the Netherlands; 18) Dept of Cardiology, Division Heart & Lungs, University Medical Center Utrecht, Utrecht, the Netherlands; 19) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA, USA; 20) Department of Clinical Epidemiology and Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands; 21) Department of Epidemiology Erasmus MC - University Medical Center Rotterdam, Rotterdam, the Netherlands; 22) Department of Internal Medicine B - Cardiology, Pneumology, Infectious Diseases, Intensive Care Medicine; University Medicine Greifswald, Greifswald, Germany; 23) DZHK (German Centre for Cardiovascular Research); partner site Greifswald, Greifswald, Germany.; 24) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany.; 25) Fred Hutchinson Cancer Research Center, Division of Public Health Sciences, Seattle, WA, USA.; 26) Section of Cardiovascular Medicine and Section of Preventive Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA; 27) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA.; 28) Cardiac Arrhythmia Service, Massachusetts General Hospital, Boston, MA, USA

---

**Background:** P-wave duration (**PWD**) is an electrocardiographic (**ECG**) measurement that represents cardiac conduction in the atrium. Both shortened and prolonged PWD have been associated with atrial fibrillation (**AF**). Although PWD is heritable, genetic studies of PWD are limited. To identify novel variants for PWD, our study examined the effects of common and rare coding variants for maximum PWD, using exome chip data.

**Methods:** We included 15 studies comprising 64,452 individuals of four ancestries (56,955 European, 5,681 African, 1,186 Hispanic, and 630 Asian). Nearly 245,000 variants genotyped on the exome chip were tested for associations with maximum PWD on the 12-lead ECG. We performed single variant score tests in each individual study, adjusting for age, sex, RR interval, and study-specific covariates, and accounting for relatedness. Association results were summarized via meta-analysis. We used gene-based burden and SKAT tests to examine associations between low-frequency and rare variants with PWD. We then examined the directionality of PWD associations with AF risk by examining results from a previous genetic analysis of AF. We used Bonferroni corrections for multiple testing.

**Results:** We identified 20 common genetic loci (13 novel) associated with maximum PWD, including several well-known AF loci (*TTN*, *CAND2*, *SCN10A*, *PITX2*, *PLN*, *CAV1*, *SYNPO2L*, *SOX5*, *TBX5*, and *MYH6*). Notably, PWD variants governed diverse biological mechanisms, including cardiac development, cardiac electrophysiology, and muscle function. The top variants at known sarcomeric genes (e.g., *TTN*, *MYH6*) were associated with both increased PWD and increased AF risk. In contrast, top variants at loci involved in transcriptional regulation (*PITX2*) or ion channel function (*SCN10A*) were associated with increased PWD but reduced AF risk. Gene-based SKAT results using variants with  $MAF \leq 5\%$  identified 4 genes related to PWD (*TTN*, *DLEC1*, *TTC21A*, and *RPL3L*).

**Conclusion:** Our findings highlight the shared genetic components of atrial conduction and AF. The different directions of associations between loci with both PWD and AF suggest that AF may result from multiple genetic mechanisms, and PWD may be an endophenotype for these mechanisms. Future studies focused on ECG traits are warranted to determine if different pathophysiologic subtypes of AF can be identified.

## PgmNr 2776: Mendelian randomization reveals a causal role for obesity in stroke that varies across subtypes.

### Authors:

S. Marini<sup>1</sup>; J. Merino<sup>1,2,3</sup>; B.E. Montgomery<sup>1</sup>; R. Malik<sup>4</sup>; C.L. Sudlow<sup>5</sup>; M. Dichgans<sup>4,6,7</sup>; J.C. Florez<sup>1,2,3,8</sup>; J. Rosand<sup>1,8,9,10</sup>; D. Gill<sup>11</sup>; C.D. Anderson<sup>1,8,9,10</sup>; on behalf of the International Stroke Genetics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) center for genomic medicine, Massachusetts general hospital, Boston, Massachusetts.; 2) Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA; 3) Department of Medicine, Harvard Medical School, Boston, MA, USA; 4) Institute for Stroke and Dementia Research (ISD), University Hospital of Ludwig-Maximilians-University (LMU), Munich, Germany; 5) Centre for Medical Informatics, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom; 6) Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; 7) German Centre for Neurodegenerative Diseases (DZNE), Munich, Germany; 8) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 9) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 10) Henry and Allison McCance Center for Brain Health, Massachusetts Gen; 11) Department of Biostatistics and Epidemiology, School of Public Health, Imperial College London, London, UK

**Background and Purpose:** Obesity has a known association with cerebrovascular disease. We aimed to clarify the (i) causality of this association; (ii) differential effects across stroke subtypes; and (iii) the role of hypertension and hyperglycemia in mediating the effect of obesity.

**Methods:** We used summary statistics from genome-wide association studies (GWAS) for body mass index (BMI), waist-to-hip ratio (WHR) and multiple cerebrovascular disease phenotypes. We explored causal associations by employing two-sample Mendelian Randomization (MR) accounting for reciprocal influence between BMI and WHR with multivariable MR. We also assessed what proportion of the association between obesity and cerebrovascular disease was mediated by systolic blood pressure (SBP) and average blood sugar levels (using glycated hemoglobin, HbA1c, as a proxy).

**Results:** Genetic predisposition to higher BMI did not increase risk for cerebrovascular disease. In contrast, WHR was associated with all-cause ischemic stroke, and specifically large artery stroke (Odds Ratio (OR) for 1 log-increment in WHR = 1.75; 95% Confidence Intervals (CI)= 1.44-2.13); and across manifestations of small vessel disease, including small vessel stroke (OR=1.57; CI=1.29-1.91), non-lobar intracerebral hemorrhage (OR= 2.97; CI=1.59-5.57) and white matter hyperintensity (OR=1.1; CI=1.01-1.23). These WHR associations were confirmed when adjusted for BMI. Proportion of the observed effect of WHR mediated by SBP was significant for all-cause ischemic stroke (11.8%), LAS (13.4%) and SVS (11.0%), but no mediation was observed for HbA1c.

**Discussion:** Causal pathological processes, totally or partially independent from hypertension and glucose levels, may link abdominal adiposity with cerebrovascular disease. Potential targets of this cascade elucidated by further research could represent novel therapeutic targets for cerebrovascular disease.



## PgmNr 2777: Calibrated rare variant genetic risk scores for complex disease prediction using large exome sequence repositories.

### Authors:

R. Lali<sup>1</sup>; M. Chong<sup>1,2</sup>; A. Omid<sup>1</sup>; P. Mohammadi-Shemirani<sup>1,3</sup>; A. Le<sup>1,3</sup>; G. Paré<sup>1,2,3,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Population Health Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, 237 Barton Street East, Hamilton, ON L8L 2X2, Canada Health Research Institute, Hamilton, Canada; 2) Department of Biochemistry and Biomedical Sciences, McMaster University, Faculty of Health Sciences, 1280 Main Street West, Hamilton ON L8S 4K1, Canada; 3) Department of Medical Sciences, McMaster University, Faculty of Health Sciences, 1280 Main Street West, Hamilton ON L8S 4K1, Canada; 4) Thrombosis and Atherosclerosis Research Institute, David Braley Cardiac, Vascular and Stroke Research Institute, 237 Barton Street East, Hamilton, ON L8L 2X2, Canada; 5) Department of Pathology and Molecular Medicine, McMaster University, Michael G. DeGroote School of Medicine, 1280 Main Street West, Hamilton ON L8S 4K1, Canada; 6) Department of Clinical Epidemiology & Biostatistics, McMaster University, 1280 Main Street West, Hamilton ON L8S 4K1, Canada

Although rare variants cause Mendelian disorders, known disease-causing variants explain a negligible portion of population risk for complex diseases. We hypothesized that rare variants (minor allele frequency < 0.001) individually confer modest effects, but are collectively numerous and may account for a considerable proportion of complex disease risk. Identifying true rare variant associations is challenging due to the need for large sample sizes, presence of technical artefacts, and heterogeneity in population structure. We propose a novel method that leverages summary-level data from a large public exome sequencing database (gnomAD) as controls and calibrates rare variant burden at the individual and gene levels to circumvent biases in population substructure and mutation burden, respectively. First, we demonstrate the need for rare variant calibration by showing that gnomAD harbours systematic biases compared to gold-standard consensus sequences. Second, we show through simulations that our method leads to substantial improvements in power while controlling for false-positive associations. Third, applying our method to a comparison of healthy control exomes (N = 6,099) to gnomAD (N = 55,880) empirically demonstrates that it adequately controls for unadjusted genomic deflation ( $\lambda_{\text{unadjusted}} = 0.61$ ;  $\lambda_{\text{adjusted}} = 0.96$ ). The method was then applied to a European coronary artery disease (CAD) cohort (N = 5,921) where the *bonafide* CAD gene, *LDLR*, reached exome-wide significance (OR = 2.21;  $P = 2.09 \times 10^{-12}$ ). A rare variant genetic risk score was generated for the top 970 discovery genes (RVGRS970) and found to associate with CAD in two independent case-control samples: 1) UK Biobank Europeans (N = 45,850) (OR = 1.08 per SD; 95% CI, 1.05-1.12;  $P = 2.50 \times 10^{-6}$ ) and 2) Pakistan Risk of Myocardial Infarction South Asians (N = 6,654) (OR = 1.07; 95% CI, 1.02-1.12;  $P = 0.009$ ). Furthermore, RVGRS970 was independent of both common variant genetic risk score and clinical risk factors (OR = 1.08; 95% CI, 1.04-1.12;  $P = 7.90 \times 10^{-6}$ ), and significantly improved classification of CAD events (Net Reclassification Index = 0.0674;  $P = 7.02 \times 10^{-5}$ ) even when excluding *LDLR*. Current polygenic risk scores have only been derived from common genetic variants. Our method improved CAD risk prediction by accounting for the aggregate effect of rare variants through a polygenic model of inheritance, which was shown to



be robust to gene, individual, and population level biases.

## PgmNr 2778: Rare variant position: Predictor of pathogenicity in cardiomyopathy.

**Authors:**

A. Waring; K. Thompson; A. Harper; M. Farrall

[View Session](#) [Add to Schedule](#)

**Affiliation:** Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford

---

In many Mendelian disease genes, variants cluster in specific regions of the gene. For many observed clusters in the disease cohort, there is a corresponding depletion of variants at the same region in controls. A plausible mechanisms for this phenomenon are multiple loss or gain-of-function mutations in an important motif of the protein.

One Mendelian disease that displays this clustering is hypertrophic cardiomyopathy (HCM). In HCM, causal mutations are rare, dominant, observed in a small set of autosomal genes. Only around half of cases have a mutations in a known gene after genetic screening. Two explanations for this shortfall are; the causal gene is not on the gene panel or the causal variant was not detected. The most common way of determining a causal gene or mutation is an increase in the frequency of mutations in an affected group (cases) versus an unaffected group (controls). However, more power can be gained by integrating more features into the analysis such as allele frequency and pathogenicity scores from variant prediction algorithms. Here, we argue that protein position is a powerful and essential feature to consider.

Here we develop a rare variant association test (POSBURDEN) to screen for novel genes and a statistical model (hcmGAM) to interpret observed variants in core HCM genes. In both, position signal is used to improve power over existing methods. Using simulated data we demonstrate that POSBURDEN is powerful, fast and keeps type 1 error at alpha. When applied to our gene panel data, we see the landscape of burden and position signals across known HCM genes. For most known genes signal strength is increased when considering positional information. For the core genes with strong position signal our predictive model hcmGAM performs very well in distinguishing pathogenic and non-pathogenic variants. The model also outputs predicted odds-ratios with 95% confidence intervals that are useful for novel variant interpretation.

## **PgmNr 2779: DataSTAGE: Leveraging cloud technology to accelerate scientific discovery for heart, lung, blood, and sleep research.**

### **Authors:**

S. Ahalt; K.C. Bradford; DataSTAGE Consortium

[View Session](#) [Add to Schedule](#)

**Affiliation:** Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NORTH CAROLINA.

---

Biomedical and genomics research represent an exponentially growing source of data that far outstrips individual researchers' ability to store and analyze. Recognizing this challenge and the need to democratize data and tool access for the scientific community, NHLBI initiated DataSTAGE (Storage, Toolspace, Access and analytics for biG data Empowerment) in 2018. DataSTAGE's goal is to enable heart, lung, blood, and sleep research investigators to find, access, share, store, and compute on large data sets. DataSTAGE is a cloud-based platform providing tools, applications, and workflows to enable these capabilities in secure workspaces. DataSTAGE initially focused on opening access to TOPMed datasets, representing a high value, high return for NHLBI investigators.

DataSTAGE is agile and iterative, defining and supporting initial scientific use cases to accelerate infrastructure development. DataSTAGE focused on two initial use cases: 1) enabling large scale genomic analyses and 2) enabling investigators to conduct Deep Learning analyses on image data. Both require addressing extensive challenges, such as handling big data, computing on a large scale, and enabling easy data access. DataSTAGE addresses these by developing solutions on a cloud-based platform utilizing technology and expertise from multiple teams.

Currently, alpha users log in to the DataSTAGE platform with their eRA Commons ID and import datasets of interest into a secure workspace based on their dbGaP credentials. Users can access several tools to explore datasets and to gain a better understanding of cohort characteristics. Additionally, users can conduct high powered analyses such as Genome-Wide Association Studies (GWAS) or Deep Learning on large scale datasets utilizing cloud resources. Users can publish and share their results on the DataSTAGE platform. Users also have access to extensive IT support for questions or suggestions for improvements.

In summary, DataSTAGE has developed many solutions to address NHLBI's challenges. DataSTAGE utilizes cloud technology to empower big data and its computation. DataSTAGE utilizes federated, centralized storage to facilitate secure data access. DataSTAGE worked with TOPMed investigators to provide state-of-the-art, sophisticated tools for genomics and other analyses. DataSTAGE facilitates team science, collaboration, and the ability for investigators to bring-your-own data to the platform. Thus, DataSTAGE significantly accelerates research discovery.

## PgmNr 2780: Malacoda: A Bayesian framework for NGS high-throughput screens that enables knowledge integration.

### Authors:

A. Ghazi <sup>1</sup>; X. Kong <sup>2</sup>; E. Chen <sup>3</sup>; L. Edelsten <sup>2</sup>; C. Shaw <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX; 2) Cardeza Foundation for Hematological Research, Thomas Jefferson University, Philadelphia, PA; 3) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

---

NGS studies have uncovered an ever-growing catalog of human variation while leaving an enormous gap between observed variation and experimental characterization of variant function. High-throughput screens powered by NGS have greatly increased the speed of variant functionalization, but appropriate and generalized statistical methods to analyze screen data have lagged. In the massively parallel reporter assay (MPRA), short barcodes are counted by sequencing input DNA and output RNA in order to measure the shifts in transcription induced by thousands of genetic variants simultaneously. These counts present many statistical challenges, including over-dispersion, depth dependence, and uncertain DNA concentrations. So far, the statistical methods used have been rudimentary, employing transformations on count level data, disregarding experimental and technical structure, and failing to quantify uncertainty in the statistical model. Analysis is also typically performed in an *ad hoc* manner that has low power, poor specificity, and can be difficult to communicate and reproduce.

We have developed an extensive framework for the analysis of NGS functionalization screens available as an R package called malacoda. Our software implements a generative, fully Bayesian model of screen data. The model uses the negative binomial distribution with gamma priors to model sequencing counts while accounting for effects from input library preparation and sequencing depth. The priors are estimated empirically and can optionally use external annotations such as ENCODE data to obtain informative priors – a transformative capability for data integration. The package also includes quality control and utility functions, including automated barcode counting and visualization methods.

To demonstrate the power of our method, we analyzed published data, created simulations, and performed two MPRA studies to analyze variant-induced regulation of platelet genes. The first MPRA tested 81 variants in *CD36* and identified 10 affecting transcription, including one identified only by malacoda. The second MPRA assessed 2666 variants associated with 184 platelet-related genes, identifying 164 functional variants of which 79 were newly identified by malacoda. Validation experiments have confirmed the function of these variants, demonstrating the method's higher sensitivity and specificity. We have also demonstrated the advantages of our framework on simulated and public MPRA and CRISPR data.

## **PgmNr 2781: The pegasis pipeline for cardiovascular genetic discovery: Open-access pedigree-based shared genomic segment analysis on the American Heart Association Precision Medicine Platform.**

### **Authors:**

S. Knight <sup>1,3</sup>; R.C. McCubrey <sup>1</sup>; R. Sargent <sup>2</sup>; D. Bride <sup>2</sup>; M. Madsen <sup>2</sup>; J. Gardner <sup>2</sup>; J.B. Muhlestein <sup>1,3</sup>; K.U. Knowlton <sup>1</sup>; N.J. Camp <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Intermountain Medical Center, Murray, Utah.; 2) Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah; 3) Internal Medicine, University of Utah, Salt Lake City, Utah

---

The genetics of cardiovascular disease (CVD) is not fully understood and a significant barrier to discovery is heterogeneity. Extended, high-risk pedigree designs can, in part, address this issue by providing a sizable and more homogeneous set of diseased individuals, as well as providing evidence for segregation of susceptibility alleles. However, genetic analyses in large pedigrees can add computational burden and complexity. To provide a more straightforward approach to large pedigree analyses for CVD phenotypes, we have developed a data resource and analysis workflow to facilitate collaborative gene-discovery studies. The Pegasis Pipeline (Pedigree-based Genetic Analyses) is hosted on the American Heart Association Precision Medicine Platform which uses Amazon Web Services. The pipeline consists of two main components: (1) pedigree identification and selection; and (2) shared genomic segment (SGS) analysis. The first component allows the user to define CVD phenotypes and assess the existence of pedigrees. The cardiovascular-based clinical dataset is searched using a Jupyter notebook to identify individuals with a defined phenotype ("cases"), then genealogy and DNA/genetic data from Intermountain Healthcare are used to identify pedigrees with at least four cases with DNA samples. There are a total of 8500 pedigrees defined by the genealogical data and genotype data is already available for a subset with coronary artery disease and atrial fibrillation. Pedigrees can also be drawn as part of the pipeline. The second component allows the user to analyze the genotypes within these pedigrees using SGS. As a pedigree segregation-based method, SGS identifies areas of the genome that cases in the pedigree are sharing and assigns statistical significances to that sharing. The genetic analysis component to the pipeline has a user interface with middleware and postgres database backend and uses the cloud computing environment to dramatically decrease analysis time. The pipeline completes a genomewide SGS pedigree analysis in less than one week. Importantly, the site has received HIPAA compliance certification and provides a secure platform which can be accessed by different institutions. Beta testing of the pipeline using 17 pedigrees with high rates of atrial fibrillation is currently underway. Future collaborative efforts are being developed and we anticipate the pipeline will be ready for external users by 2020.

## PgmNr 2782: Mutational Evidence Visualization (*MutEviz*): Integrated information rich interactive web-based clinico-genomic application of personalized medicine.

### Authors:

S. Krishna Murthy <sup>1</sup>; D. Carey <sup>1</sup>; H. Williams <sup>1,2</sup>; R. Metpally <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Functional Genomics, Geisinger Health System, Danville, Pennsylvania.; 2) Department of Urology, Geisinger, Danville, PA, USA

**Background:** Processing large information with minimal data handling expertise and time constraints require automated succinct visualization in the field of Precision Medicine. Existing tools that follow the standards and guidance by the ACMG for the variant interpretation have discrete functionality and require that end-users rely on multiple software with complex dependencies for analysis and interpretation. We propose to develop a publicly available tool that assists to easily interpret the clinical significance of variants through graphical representation of variants with adequate functional annotation.

**Methods:** A web-based application developed using R Shiny that helps to visualize variants represented upon the 2-dimensional protein structure from Pfam database with defined Uniprot domains markings and functional sites marked on them. Evolutionary conserved regions are plotted using the amino acid conservation scores from Consurf database provided by the user. Mutational densities of commonly altered local regions of benign variations in control population and hotspot regions are plotted using gnomAD variants with allele frequencies (AF) below the cut off defined by CardioDB's maximum credible population AF and Clinvar pathogenic variants respectively. Read depth and allele frequencies of gnomAD variants are plotted to represent presence or absence of the variants and coverage of the region.

**Results:** With concise visualization and comprehensive feature annotations to variants and genes including GWAS summary statistics, MutEviz is a rich one-stop resource and valuable tool that allow clinicians to meet the research needs head-on. For example, the tool provides supportive evidence to show variant T449I in MYH7 gene to be predictive pathogenic for hypertrophic cardiomyopathy (prevalence 1 in 500; maximum allowed AF 2e-04) as several tools predict this variant adversely affect protein function, it falls in the local Clinvar hotspot region, as well as Constrained Coding Regions of the myosin head domain and it is not detected in GnomAD, 1KG, and ESP databases.

**Conclusion:** MutEviz is easy-to-use visualization tool that delivers information-dense, publication-quality mutation plots. Through the presentation of clear and crisp visualization of variants overlaid on perceptible protein diagrams with ample functional annotation, the tool provides a valuable addition to enable the next round of applications in precision medicine with quick turnaround times.

## PgmNr 2783: Identification of causal transcriptional components for genome-wide association studies by integrating epigenomic data and sequence-based variant prediction.

### Authors:

D. Lee; A. Chakravarti

[View Session](#) [Add to Schedule](#)

**Affiliation:** New York University School of Medicine, New York, NY.

---

Most genetic variants associated with diseases and traits from genome-wide association studies (GWAS) are thought to be transcriptional regulatory. Thus, identification of major transcriptional components (transcription factors, regulatory elements, causal variants, and target genes) for GWAS signals is a critical step to understand their underlying molecular mechanisms. To this end, we developed an integrative computational framework, combining genomic and epigenomic data with sequence-based predictive models for regulatory elements and variants. We first perform genome-wide identification of regulatory elements and their variants for each tissue of interest. This regulatory element/variant map is the basis of identifying potential causal variants at GWAS loci in a cell-type specific way. We next connect these trait-associated regulatory variants to their putative target genes using eQTL analyses and additionally predict their potential transcription factors (TFs) through TF binding site motif analysis. Our strategy is likely to identify causal and functional variants and their target genes since we restrict our attention to regulatory variants with plausible functions. As a result, we can establish a genetic link from a variant to a gene to a phenotype for each locus in a tissue-specific way. To validate our approach, we first performed partitioned heritability analysis for several phenotypes across multiple tissues to show that heritability is specifically enriched in phenotypically relevant tissues only. We found 16.5x heritability enrichment of QT-interval in cardiac regulatory variants, 15.2x heritability enrichment for hypothyroidism in thyroid regulatory variants, and 16.8x heritability enrichment for diastolic blood pressure in artery regulatory variants. Given the successful identification of the cardiac tissue from the heritability enrichment analysis for QT-interval, we further predicted target genes in the individual loci using GTEx eQTL data: 73 cardiac regulatory variants that are significantly associated with QT-interval are also heart left ventricle eQTL for 31 distinct genes in 18 loci. Notable genes include *NOS1AP*, *FADS2*, *NDRG4*, *SCN5A*, *KCNH2*, and *ANKRD9*. The identified regulatory variants and their target genes provide a testable hypothesis for further experimental validation. Our approach is generic and can be extended to any GWAS and any tissue of interest.

## PgmNr 2784: Decision support for cardiovascular genetic risk assessment.

### Authors:

M. Murugan <sup>1</sup>; E. Venner <sup>1</sup>; V. Yi <sup>1</sup>; J. Easton Marks <sup>3</sup>; T. Vi <sup>1</sup>; S. Lu <sup>1</sup>; L. Wang <sup>1</sup>; D. Murdock <sup>1</sup>; G. Metcalf <sup>1</sup>; C. Kovar <sup>1</sup>; V. Korchina <sup>1</sup>; D. Muzny <sup>1</sup>; E. Boerwinkle <sup>1,2</sup>; R. Gibbs <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) University of Texas Health Science Center, Houston, TX; 3) CIMTelligence Systems LLC, Houston, TX

---

Cardiovascular disease (CVD) is the leading cause of death for both men and women in the United States with estimates of genetic contributions to risk ranging from 5 to 30%. The Human Genome Sequencing Center Clinical Lab (HGSC-CL), along with sister departments at the Baylor College of Medicine and within the Common Spirit hospital network, have launched a CVD risk gene panel genetic screen (HeartCare™) to assess the merits of genetic testing and DNA screening, create a robust data set for genomic research and to provide a platform for development of data management tools.

Critical needs for the program included serving the clinical components of sample accessioning, consent management, DNA sequencing and EHR integration. Further, an interpretation and decision support pipeline were built, in order to speed the linking of DNA variation to clearly mandated physician action. This includes bioinformatics components for inferring clinical variant interpretation, pharmacogenomics and polygenic risk scores for clinical reporting. Components of this pipeline prompt clinical care directly by triggering best practice alerts in the EHR. For example, Lipoprotein(a) testing is automatically recommended via prompts for patients who are determined to carry LPA genetic risk variants. In aggregate these automated decision support tools greatly reduce the burden on non-genetic physician specialists.

The translational bioinformatics pipeline and infrastructure built for this CVD pilot program, is future proofed and engineered to support other clinical applications at the HGSC-CL. The use of a hybrid cloud and local infrastructure ensures effective capacity, security and compliance management. EHR integration implementations created for this program establish a baseline for future growth and standardization for integration and for building clinical decision support applications.



## PgmNr 2785: Annotation and functional characterization of lncRNAs using integrative GWAS, eQTL and network analysis in lncRNAKB.

### Authors:

M. Pirooznia; F. Seifuddin; A. Suresh; K. Singh; Y. Chen; V. Chaitankar; I. Tunc

[View Session](#) [Add to Schedule](#)

**Affiliation:** NHLBI, NIH, Bethesda, MD

---

Advent of high-throughput sequencing technologies and development of efficient tools to analyze big data has led to the surprising discovery that only ~2% of the human genome is protein coding and that a majority of the genome though transcribed, falls into the non-protein coding transcript category. Several groups of researchers have used manual and automated techniques to identify and annotate lncRNAs in the human genome. However, questions remain about the purpose and function of these lncRNAs and whether they play a critical role in normal cell function and/or in disease. Here we present a pipeline that uses publicly available RNA-seq (across 31 solid organ human tissues) and genome-wide association (GWAS) data to functionally characterize ~100,000 lncRNAs available in the lncRNAKB built by step-wise integration of six commonly used lncRNA human annotation databases. First, our analysis addresses the conservation, tissue-specificity, and coding potential of all the lncRNAs curated in the lncRNAKB. Since the expression patterns of lncRNAs are considered to be tissue-specific, tissue-specific metrics were calculated using the RNA-seq data. A tau statistic, which determines the overall tissue specificity of genes, and preferential expression measure (PEM), which scores genes by individual tissue, were computed. In order to determine coding potential of lncRNAs, the FEELnc package which implements a random forest machine learning algorithm (considering various sequence features) was used to determine the coding potential as well as annotate these with respect to protein-coding mRNAs. Next, we constructed tissue-specific protein coding-lncRNA co-expression networks to functionally characterize the cellular processes that the lncRNAs may be involved in using Weighted Gene Co-expression Network Analysis (WGCNA). Finally, we perform tissue-specific expression quantitative trait locus (eQTL) analysis and overlay these to GWAS summary data to identify subsets of single nucleotide polymorphisms (SNPs) in lncRNA that may have an association between gene expression and disease phenotype using Summary-data Mendelian Randomization (SMR) analysis. Put-together, these analyses will provide deep insight into the underlying biology of lncRNAs in cell function and disease across several human tissues. Creating a co-expression network sheds light on the function of lncRNAs in heart failure and how they might contribute to gene expression and pathology.

## **PgmNr 2786: Quality of whole genome sequencing from blood versus saliva derived DNA in cardiac patients.**

### **Authors:**

R. Yao <sup>1</sup>; O. Akinrinade <sup>1</sup>; M. Chaix <sup>1</sup>; S. Mital <sup>1,2</sup>

[View Session](#)[Add to Schedule](#)

### **Affiliations:**

1) Genetics and Genome Biology, Hospital for Sick Children, Toronto, Canada; 2) Department of Pediatrics, Hospital for Sick Children, University of Toronto, Toronto, Canada

---

Whole-genome sequencing (WGS) is becoming an increasingly important tool for detecting genomic variation. Blood derived DNA is the current standard for WGS for research or clinical purposes. We compared the level of microbial contamination, sequencing coverage, as well as yield and concordance of single-nucleotide polymorphism (SNP) and copy number variant (CNV) calls in WGS from paired blood and saliva samples from 5 pediatric heart disease patients. We found that although saliva samples contained a higher proportion of sequence reads that map to the human oral microbiome, these reads were readily excluded by mapping the reads to the human reference genome. Sequencing coverage was low only in 1 of 5 saliva samples. Over 95% SNPs (including rare SNPs) but <80% CNVs called in blood genomes were detected in paired saliva genomes. These findings suggest that most good quality saliva samples can serve as an alternative to blood samples for detection of sequence variants from WGS in cardiovascular disease patients.

## PgmNr 2787: Examining the lipidomic footprint of hypertension in African Americans.

**Authors:**

D. Nguyen; A. Gaye; G. Goodney; G. Gibbons

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Human Genome Research Institute, National Institutes of Health

---

**Background**

Hypertension (HTN) is prevalent in African Americans (AA), affecting approximately 40 percent of the AA population. As a multifactorial condition, HTN involves many physiological processes and metabolic changes. Recent advances in metabolomic profiling has given rise to the opportunity to examine metabolites involved with HTN. The link between lipids and HTN in African Americans has not been well characterized. This study aims to identify lipid metabolites related to HTN in a sample of AA.

**Methods**

The data consists of plasma lipidomic profiles of 165 AA (99 hypertensives and 66 normotensives) from the Minority Health Genomics and Translational Research Bio-Repository Database (MH-GRID) study.. LC/QTOF MS was used for lipidomics analyses with a total of 407 lipids identified. These lipids were then used as predictor variables in a Random Forests (RF) classification to identify sets of lipid metabolites that predict HTN.

**Findings**

RF analysis identified 7 metabolites that classify hypertensive versus control with an area under the curve (AUC) of 0.80. Five of the seven lipids identified were phosphocholines. The other two lipids were cholesterol and a triglyceride.

**Conclusion**

It has been shown that the levels phospholipids like phosphocholines play an important role in the development of various components of metabolic syndrome. Although the lipid metabolites identified in this analysis are not well characterized, the unbiased and data-driven approach of this analysis suggests that they are likely to play an important role in the in the pathophysiology of HTN. Further work is underway to understand the link between these metabolites and HTN and if it's specific to the genetic ancestry of AA.

## PgmNr 2788: Calling and imputation of the common $\alpha$ -globin copy number variant with whole genome sequencing data in TOPMed and association with hematologic and other clinical phenotypes.

### Authors:

M. Jiang <sup>1</sup>; Y. Su <sup>2</sup>; T.W. Blackwell <sup>3</sup>; A. Correa <sup>4</sup>; G. Abeçasis <sup>3</sup>; A.P. Reiner <sup>5</sup>; Y. Li <sup>2,6,7</sup>; L.M. Raffield <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Applied Physical Sciences, University of North Carolina, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC; 3) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 4) Department of Medicine, University of Mississippi Medical Center, Jackson, MS; 5) Department of Epidemiology, University of Washington, Seattle, WA; 6) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 7) Department of Computer Science, University of North Carolina, Chapel Hill, NC

---

Recent work has shown that inheriting a single copy of the sickle cell causing  $\beta$ -globin variant rs334 (i.e. sickle cell trait) can be associated not only with alterations of blood cell indices and hemoglobin A1c levels, but also with increased risk of certain medical conditions, such as chronic kidney disease. In individuals with sickle cell disease, co-inheritance of another globin gene variant, the 3.7 kb  $\alpha$  globin gene deletion (a common cause of  $\alpha$ -thalassemia in African populations), can modify the risk of stroke and other sickle cell disease complications. In this work, we first aim to better catalog this CNV in population-based cohorts via whole genome sequencing data from the NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, assessing the performance of different structural variant calling methods. We next evaluate the ability to impute this CNV using SNP genotypes from commercial arrays. Lastly, we conduct association analysis for main effects of the  $\alpha$  globin CNV and its interactions with other globin gene variants on a battery of hematological traits and other clinical phenotypes including stroke, chronic kidney disease, and hemoglobin A1c levels. Preliminary results from 2,916 African Americans in the Jackson Heart Study demonstrated highly reliable calling using either GenomeSTRiP or LUMPY (correlation >0.99). In addition, we can achieve reasonable imputation quality for structural variant calls ( $r^2 = 0.629$  using minimac3) for imputation from Affymetrix 6.0 SNP genotypes in JHS. Finally, association analysis confirmed the CNV's significant main effects as well as interaction effects with other  $\beta$  globin variants on multiple traits. For example, the CNV is associated with higher hemoglobin A1c ( $p = 0.0001$ ) and also attenuates the increased risk of anemia in carriers of sickle cell trait ( $p_{interaction} = 0.031$ ). We anticipate that accurate calling and imputation of this CNV in additional African American and Hispanic/Latino cohorts will allow powerful interrogation of its main and interaction effects on a wide range of blood cell and other clinical phenotypes.

## PgmNr 2789: Transcriptome signature of declined kidney function in hypertensives from the Minority Health-GRID study.

### Authors:

M. Pavia; A. Gaye; G. Goodney; G. Gibbons

[View Session](#) [Add to Schedule](#)

**Affiliation:** NHGRI, NIH, Bethesda, MD.

---

### Background

The disproportionate burden of hypertension and chronic kidney disease (CKD) amongst African Americans is a well-documented public health concern in the United States. However, the relationship between these two conditions and the genes that predict them, particularly in this understudied population, requires further investigation. The purpose of this study was to explore whole-blood genome-wide transcriptome to identify clusters of genes related to declined kidney function in hypertensive subjects.

### Methods

The analysis used whole-blood transcriptome of 180 individuals from the Medicine Minority Health Genomics and Translational Research Bio-Repository Database (MH-GRID) Study. Cases are hypertensive with signs of declined kidney function based on eGFR and albumin-creatinine ratio (ACR). Weighted Gene Co-expression Network Analysis (WGCNA) was used to identify clusters of genes associated with the phenotype of interest.

### Results

A total of 3 network modules of respectively 133, 704 and 110 were found positively associated with reduced kidney function in hypertensive subjects. One of the modules is also associated with systolic and diastolic blood pressure (SBP and DBP) whilst the other 2 are associated with inflammasome proteins including MMP-3, MMP-7 and MMP-9. The top pathways enriched in the modules of co-expressed genes are respectively *Receptor-Mediated Endocytosis(RME)*, *STAT3* and *B cell Development*.

### Conclusion

The network analysis approach revealed a cluster of co-expressed genes enriched in RME pathway. The endocytic pathway is central in the uptake of low molecular weight protein in the proximal tubule of the kidney; if impaired this process leads to proteinuria. Matrix metalloproteinase (MMP) genes are involved in the degradation of the extracellular matrix (ECM) which is crucial in the development of CKD.

# PgmNr 2790: Hybrid *de novo* genome assembly of individuals with cystic fibrosis using linked- and long-read sequencing.

## Authors:

S. Mastromatteo <sup>1,2</sup>; B. Thiruvahindrapuram <sup>1</sup>; J. Whitney <sup>1</sup>; W. Sung <sup>1</sup>; Z. Wang <sup>1</sup>; G. Kaur <sup>1</sup>; F. Lin <sup>2</sup>; B. Xiao <sup>2</sup>; J. Gong <sup>2</sup>; N. Panjwani <sup>2</sup>; A. Bao <sup>2</sup>; A. Chen <sup>2</sup>; J.M. Rommens <sup>2,4</sup>; L.J. Strug <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 3) Statistical Sciences and Biostatistics, University of Toronto; 4) Department of Molecular Genetics, University of Toronto

---

Short read sequencing with alignment to reference genome is the conventional approach to whole genome analysis but has difficulty resolving low-complexity repeats and structural variants (SVs), and performs poorly in variant phasing for haplotype construction. Exemplary challenges are genetic modifier loci of Cystic Fibrosis (CF) identified by GWAS, where several align to regions of the genome that are repetitive, GC-rich, or contain SVs (eg. *MUC5b*, *PRSS1*, *SLC9A3*). To understand the mechanisms by which these modifier loci contribute to CF disease severity we aimed to improve genome resolution with *de novo* assembly of individual genomes.

The 10x Genomics Chromium platform uses barcodes that link short reads from common DNA molecules. This linking information enables phasing and increased detection of SVs, but is still limited in low-complexity and repetitive regions. Pacific Biosciences (PacBio) technology allows sequencing of longer reads from single molecules but is highly error-prone and costly at high read-depth. We introduce a novel pipeline for building hybrid *de novo* assemblies using data from both platforms.

Our hybrid method uses separate 10x and PacBio assemblies as input. Scaffolds from 10x are split into 1kb chunks and aligned to the PacBio assembly. Contiguous chunks define blocks of homologous sequence between the assemblies and are used to fill gaps, correct misassemblies, scaffold contigs, and maximize nucleotide accuracy. A single hybrid assembly is produced and phased using both 10x and PacBio reads.

Our hybrid assembly pipeline will be applied to genomic DNA from participants of the Canadian CF Gene Modifier study (n=100 planned; 30x and 50x coverage for 10x and PacBio, respectively). Results for the first 8 hybrid assemblies displayed a mean of 2.81Gb in 837 scaffolds compared to 10x (2.87Gb, 19231) and PacBio (2.97Gb, 4055) assemblies. The average hybrid N50 (46Mb) was improved from the 10x (25Mb) and PacBio (15Mb) alone. An average of 44000 (98%) gaps within the 10x assemblies were resolved by our hybrid. Phasing produced large haplotype blocks, some on the scale of entire chromosomal arms.

The hybrid assemblies have resolved locus complexity at several CF modifier loci and informed specific regions of interest that are currently being investigated. Here we present our progress to

date, benchmark our method using data from the Genome in a Bottle Consortium, and demonstrate a high-quality hybrid even at reduced PacBio read depths.

## PgmNr 2791: Identification of novel genes associated with hypoxia response using Boolean implication relationships.

### Authors:

T. Stobdan<sup>1</sup>; D. Sahoo<sup>1,2</sup>; G.G. Haddad<sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, Univ California San Diego, La Jolla, California.; 2) Department of Computer Science and Engineering, Univ California San Diego, La Jolla, California.; 3) Department of Neurosciences, Univ California San Diego, La Jolla, California.; 4) Rady Children's Hospital, San Diego, California, US.

---

Hypoxia plays a major role in the etiology and pathogenesis of most of the leading causes of morbidity and mortality, whether cardiovascular diseases, cancer, respiratory diseases or stroke. Despite active research on hypoxia-signaling pathways, the understanding of regulatory mechanisms, especially in specific tissues, still remain elusive. With the accessibility of thousands of potentially diverse genomic datasets, computational methods are utilized to generate new hypotheses and expedite existing endeavors. Here we utilized Boolean relationships, a powerful method to probe correlated genes, of a well-known hypoxia-responsive gene, *VEGFA*, with very large human expression datasets ( $n = 25,955$ ) to identify novel hypoxia-responsive candidate gene/s. Subsequently, we utilized *in-vitro* analysis using human endothelial cells exposed to 1% O<sub>2</sub> environment for 2, 8, 24 and 48 hours to validate few of the top candidate genes. There were 160 genes with correlation coefficient value of  $>0.5$  with *VEGFA*. Our preliminary analysis of the top 15 genes which were also in a Boolean relationship with *VEGFA* reveals *FAM114A1* as a novel candidate gene significantly upregulated and *EPS8* downregulated in the endothelial cells at 24 and 48 hours of 1% O<sub>2</sub> environment. Furthermore, an additional 11 genes were previously reported as hypoxia responsive genes, validating our result. At this rate of  $>85\%$  (13 out of 15) positive hits our results confirm that the additional *in-vitro* validation of Boolean correlated genes will reveal novel hypoxia responsive candidate genes. Furthermore, since Boolean logic is an empirically observed invariant of gene expression levels of two genes, this method can simultaneously identify both HIF-dependent and HIF-independent hypoxia-responsive genes.



## PgmNr 2792: Establishing quality control practices for phenome-wide association studies (PheWAS).

### Authors:

K. Passero <sup>1</sup>; J. Zhou <sup>2</sup>; B. Mueller-Myhsok <sup>3,5,6</sup>; M.E. Kleber <sup>4</sup>; W. Maerz <sup>4,7,8</sup>; M.A. Hall <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Huck Institutes of the Life Science, Pennsylvania State University, University Park, PA; 2) Department of Veterinary and Biomedical Sciences, College of Agricultural Sciences, The Pennsylvania State University, University Park, PA, USA; 3) Statistical Genetics, Max Planck Institute of Psychiatry, Munich, 80804, Germany; 4) Vth Department of Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; 5) Munich Cluster of Systems Biology, SyNergy, 81377 Munich, Germany; 6) Institute of Translational Medicine, University of Liverpool, Liverpool, L69 3BX, United Kingdom; 7) Synlab Holding Deutschland GmbH, SYNLAB Academy, Mannheim and Augsburg, Germany; 8) Medical University of Graz, Clinical Institute of Medical and Chemical Laboratory Diagnostics, Graz, Austria

---

Phenome-wide association studies (PheWAS) allow agnostic investigation of common genetic variants in relation to a variety of phenotypes. Preserving the power of PheWAS requires careful phenotypic quality-control (QC) procedures. While the QC of genetic data is well-defined, there are no established practices for QC of multi-phenotypic data. Manually imposing sample size restrictions, identifying variable types/distributions, and locating problems such as missing data or outliers is arduous in large, multivariate datasets. To this end, we utilized the novel R package CLARITE (CLEaning to Analysis: Reproducibility-based Interface for Traits and Exposures), which allows efficient QC. Data from the Ludwigshafen Risk and Cardiovascular Health Study (LURIC) was used for a coronary-related PheWAS in 2,824 samples. Initially, genome-wide single nucleotide polymorphisms (SNPs) were tested for association with 15 disease phenotypes using logistic regression. No results reached genome-wide significance. One result, mortality after follow-up, was close to significance ( $p:9.57e-07$ ) and involved in a missense mutation in *PDZRN4* (rs285584). A second PheWAS assessed association between genome-wide SNPs and 14 polyunsaturated fatty acids (PUFAs), as omega-3 fatty acids are implicated in heart disease. Seven SNPs were genome-wide significant in association with dihomo- $\gamma$ -linolenic acid. Five of those SNPs were within fatty acid desaturases *FADS1* and *FADS2*, which are key catalyzers of PUFA synthesis and well referenced in GWAS literature. The remaining SNPs were in *FEN1* and *MYRF*. Mutations in *MYRF* have been identified in congenital cardiac abnormalities. PheWAS is a useful tool to elucidate the genetic architecture of complex disease phenotypes within a single experimental framework. Here, we utilized CLARITE across a diverse set of variables, demonstrating its utility in QC of various variable types, which would be arduous to perform by hand. Finally, we provide a generic QC pipeline for phenotype data, widely applicable to genome-wide datasets.

## PgmNr 2793: Temporal variation of regulatory organization based on longitudinal RNA-seq data in twins.

### Authors:

A. Ramisch <sup>1,2</sup>; D. Ribeiro <sup>3</sup>; J.S. El-Sayed Moustafa <sup>4</sup>; G. Leday <sup>5</sup>; Y. Jiao <sup>6,7,8</sup>; G. Nicholson <sup>8</sup>; M. Stevens <sup>1,2</sup>; M. Abdalla <sup>6,7,8</sup>; C. Menni <sup>4</sup>; C. Holmes <sup>8</sup>; T.D. Spector <sup>4</sup>; M.I. McCarthy <sup>6,7,9</sup>; S. Richardson <sup>5</sup>; O. Delaneau <sup>3</sup>; K.S. Small <sup>4</sup>; E.T. Dermitzakis <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetic Medicine and Development, Faculty of Medicine, University of Geneva, Geneva, Switzerland; 2) Institute of Genetics and Genomics in Geneva, Geneva, Switzerland; 3) Systems and Population Genetics Group, Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; 4) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 5) Medical Research Council Biostatistics Unit, Cambridge, United Kingdom; 6) Wellcome Centre Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 7) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 7LE, UK; 8) Department of Statistics, University of Oxford, Oxford, OX1 3TG; 9) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE, UK

---

A new study by our lab (Delaneau et al. Science 364 (6439), eaat8266, 2019) used population-based chromatin data to identify cis-regulatory domains (CRDs), which show cell type-specific coordinated chromatin activity and follow the local chromatin organization into topologically associated domains.

However, for most human genetic studies, chromatin data is not available and the identification of CRDs becomes difficult and expensive. Therefore, we extended our previous work to characterize regulatory organization of the genome using only gene expression data from RNA-seq studies and genotypes. Here, we define co-expression domains (CODs) as groups of significantly correlated proximal genes (Ribeiro et al. ASHG 2019). We use a gene-central approach, in which we identify for each gene individually all correlated genes located within a cis-window of 1Mb. From the resulting list of co-expressed pairs of genes we summarize connected pairs into CODs which therefore can span several Mbps and may overlap with each other.

In our current work we aim to understand how the regulatory organization changes over time using a longitudinal gene expression dataset. The MutliMuTHER study includes 335 individuals with whole-blood RNA-seq data collected at three time-points over seven years, as well as imputed genotypes from the TwinsUK cohort (El-Sayed Moustafa et al. ASHG 2019). We identify CODs in each of the three time-points as well as integrating across the three time points, and explore their association to genetic variation by performing QTL analysis, which identifies genetic variants associated to the level of expression coordination of the genes involved in each of the CODs. Furthermore, we investigate the stability of CODs over time, as well as age-associated changes in coordinated regulation. Using the unique properties of twins studies, we have performed a variance decomposition analysis to estimate heritability, i.e., the proportion of variance attributed to the genetic component in CODs as well as to the environmental contribution. Finally, we associate the CODs to clinically relevant phenotypes in an attempt to identify CODs involved in the regulation of cardiometabolic traits such as cardiovascular

diseases and diabetes.

## PgmNr 2794: Trans-ethnic GWAS with individual-level data identified potential risk for cleft lip.

### Authors:

Y. Jiang; C. Tian; S. Shringarpure; D. Hinds

[View Session](#) [Add to Schedule](#)

**Affiliation:** 23andme, Mountain View, California.

Cleft lip is one of the most common birth defects in the world. The prevalence for cleft lip varies across continents: it affects one in every 940 births in the US and 1 in every 500 births in East Asia. Although several studies have been conducted to identify the genetic etiology of the disease, the risk factors are not yet clear. The majority of these studies were conducted on case-parent trios of European- or East Asian-ancestry with follow up meta-analysis on the summary statistics. This poses two questions: 1) How well do these findings generalize to other under-represented populations? 2) Are meta-analyses of summary statistics under-powered compared to analysis utilizing individual-level data?

To tackle these issues, we conducted a trans-ethnic GWAS using de-identified and consented individual-level data in 23andMe's database. Our study included 1,410 cases and 146,0457 controls of European ancestry; 572 cases and 4,4611 controls of Non-European ancestry. Specifically, our non-European cohort including individuals from Latino, East Asian, African, South Asian, and Middle Eastern. We performed statistical tests using the generalized linear mixed model and estimated the distribution of the score-statistics using saddle-point approximation (implemented in SAIGE). We used 63,654 SNPs across autosomes to construct the genetic relationship matrix and did not observe genome-wide inflation caused by population structure. Our trans-ethnics mixed-model GWAS identified 8 loci (8q24, 10q25, 17p13, 20q12, 1p36, 1p22, 8q21, 2p21) reached genome-wide significance level ( $p\text{-value} < 5e-8$ ). These loci have been reported separately by various studies. We also found 10 more loci reached the suggestive significance level ( $p\text{-value} < 1e-6$ ). In parallel, we also ran GWAS on each population and conducted meta-analysis on the summary statistics. With nearly identical sample size, mixed model and meta-analysis have similar results, whereas mixed-model show stronger associations for certain loci. We will follow-up the hits with functional analysis.

In conclusion, with individual-level data across 6 populations, trans-ethnic GWAS on cleft-lip provides valid results without genome-wide inflation. With more related individuals in the database, e.g. UKbiobank, mixed-model will be a robust and powerful tool for carrying out trans-ethnic GWAS and help under-represented populations to benefit from genetic research.

## PgmNr 2795: Update from the NIH Oculocutaneous Albinism Natural History Protocol: Molecular characterization of the study cohort.

### Authors:

L. Lundh; D. Adams

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Human Genome Research Institute, National Institutes of Health, Bethesda, MARYLAND.

---

Oculocutaneous albinism (OCA) comprises of rare, inherited disorders of melanin production. Clinical manifestations include skin, hair and eye hypopigmentation along with disrupted visual development. Pigment reduction in OCA occurs along a spectrum from complete absence of melanin to near-expected levels for a given ancestry. All defined types of OCA are inherited in an autosomal recessive manner. The reported rate of successful molecular diagnosis varies depending on residual pigment production—greater than 90% if no pigment is made and as low as 50% with near-normal pigment levels. There is no known treatment for OCA. The NIH OCA Study, established in 2008, works with volunteer study participants with OCA. As part of the process of collating study data for publication, we are re-visiting the molecular diagnoses of our 132 study participants. Presented here is an update about our observations about molecular diagnosis rates. Study participants are evaluated at the NIH Clinical Center as part of an IRB approved protocol. Blood DNA is routinely extracted for analyses including the sequencing of known albinism genes. Sequencing methodology has changed over the course of the study and includes direct Sanger sequencing and short-read based capture panels. In early stages of the study, sequencing of less common OCA genes occurred only if more common gene sequencing was unrevealing. Approximately 95.5%, 65.2% and 28.8% of study participants have been sequenced for the two most common, the four most common, and all known-genes respectively. Successful molecular diagnosis includes 42.7% OCA type 1 (*TYR* gene), 14.5% OCA type 2 (*OCA2* gene), 0.8% OCA type 3 (*TYRP1* gene), and 4.0% OCA type 4 (*SLC45A2* gene). Overall, 37.9% of our cohort has an incomplete molecular diagnosis. Using visual phenotypic categories, the success of molecular diagnosis was 94% in individuals with no residual pigmentation, 72% with trace pigment production and 46% with greater-than-trace pigment production. Missing heterogeneity in OCA is frequently debated and suggested etiologies include missing genes, gene-gene interactions, background pigment variation threshold effects and non-coding variation. Our data set supports previous observations that the successful molecular diagnosis rate is reduced in the setting of residual pigment production. We present further detail about our cohort and describe ongoing initiatives designed to explore and resolve basic questions about OCA inheritance.

## PgmNr 2796: Type-2 diabetes and cognitive function: Unravelling a potential causal relationship using bidirectional Mendelian randomisation in the UK Biobank.

### Authors:

V. Garfield <sup>1</sup>; A.E. Farmaki <sup>1</sup>; S.V. Eastwood <sup>1</sup>; R.E. Farmer <sup>2</sup>; R. Mathur <sup>2</sup>; K. Bhaskaran <sup>2</sup>; L. Smeeth <sup>2</sup>; N. Chaturvedi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Population Science and Experimental Medicine, University College London, London, United Kingdom; 2) Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

---

Type 2 diabetes has previously been associated with poor cognition in the epidemiological literature, but the direction of effect and whether it is causal, remains unclear. We aimed to investigate, using a bidirectional Mendelian randomisation design, whether having diabetes is causally associated with poorer cognition, and/or whether poorer cognition might lead to diabetes. We used data from both men and women in the UK Biobank, a population-based study of 500,000 individuals, aged between 40-60 years. Diabetes was defined using a validated algorithm. Cognitive function tests at baseline included a reaction time test. Mendelian randomisation analyses used an inverse-variance weighted approach, and MR-Egger as a sensitivity analysis to detect unbalanced horizontal pleiotropy. Our type-2 diabetes genetic instrument consisted of 163 single nucleotide polymorphisms (SNPs) from the latest genome-wide association study, whilst our genetic instrument for reaction time comprised 43 SNPs. There were complete data for 349,326 white European participants who passed genetic quality control and had cognition data, with ~14,000 type-2 diabetes cases. Inverse-variance weighted Mendelian randomisation showed no evidence of a causal effect of diabetes on reaction time ( $\text{Exp}(\beta) = 1.00$ , 95% CI = 0.99; 1.00). In addition, there was no causal effect of reaction time on diabetes risk (OR = 0.96, 95% CI = 0.63; 1.46). MR-Egger indicated that there was unlikely to be unbalanced horizontal pleiotropy in either of the above,  $P_{\text{Egger intercept}} = 0.342$  for reaction time and  $P_{\text{Egger intercept}} = 0.457$  for diabetes. In UK Biobank, we found no evidence of a causal relationship between diabetes and reaction, or reaction time and diabetes, using bidirectional Mendelian randomisation. We conclude that in white Europeans, this relationship appears not to be causal, but this may not be the case in other ethnic groups, who are disproportionately burdened by diabetes.

## PgmNr 2797: Robust methods to combine *p*-values from independent experiments.

### Authors:

D. Nam <sup>1</sup>; S. Yoon <sup>1</sup>; T. Park <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biological Sciences, UNIST, Ulsan, 44919, South Korea; 2) Department of Statistics, Seoul National University, 08826, South Korea

---

Meta-analysis increases statistical power by combining statistic from multiple studies. Meta-analysis methods have mostly been evaluated under the assumption that all the data in each study are associated with the given phenotype. However, data in some studies can yield nearly random statistic due to different experimental conditions or genetic heterogeneity. Here, we propose two robust methods to combine *p*-values obtained from independent studies, a novel weighted Fisher's method and *p*-value of minimum marginal *p*-value in joint order distribution. *T*-test and GWAS simulation analysis demonstrated that proposed methods outperformed existing *p*-value combining methods and also compared favorably with state-of-the-art methods for GWAS, when only a small number of experiments (or cohorts) were associated with the phenotype. Finally, our methods were applied to exome sequencing data composed of eight cohorts and detected a novel locus (rs37411300; BUD13) associated with triglyceride level.

## PgmNr 2798: Resilience to polygenic risk of severe obesity: The Hispanic Community Health Study/ Study of Latinos (HCHS/SOL).

### Authors:

V.L. Buchanan <sup>1</sup>; M. Graff <sup>1</sup>; A.E. Justice <sup>2</sup>; L. Fernandez-Rhodes <sup>3</sup>; R.A.J. Smit <sup>4</sup>; Y. Wang <sup>1</sup>; R. Rohde <sup>1</sup>; Q. Qi <sup>5</sup>; C.R. Isasi <sup>5</sup>; R.J.F. Loos <sup>4</sup>; K.E. North <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 2) Biomedical and Translational Informatics, Geisinger, Danville, PA; 3) Biobehavioral Health, Pennsylvania State University, University Park, PA; 4) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY

**Introduction:** Severe obesity (SevO) is a global public health threat, especially for African and Native American ancestries, and is highly heritable with multiple polygenes, environmental, and behavioral determinants. Polygenic risk scores (PRS) are a novel approach to quantify inherited susceptibility that affords new opportunities for clinical prevention and mechanistic assessment. Yet even when PRS perform well, those at highest risk may be resilient to obesity. Thus, we applied PRS to Hispanic/Latino (H/L) participants of HCHS/SOL to find factors associated with leanness in the context of an obesogenic environment and high polygenic risk.

**Methods:** HCHS/SOL is a community-based prospective cohort of 16,415 H/L adults aged 18-74 years at screening from randomly selected households at 4 US field centers. We constructed PRS with PRSice using a BMI MEGA analysis of GIANT and UK Biobank and applied to 10,122 HCHS/SOL participants with BMI ( $29.8 \pm 6.0$ ) and genotype data (Illumina SOL Omni2.5M array imputed to 1000 Genomes Phase 1). We categorized resilience as normal weight (NW) persons in the top PRS decile ( $N=113$ , BMI= $22.9 \pm 1.8$ , age  $41.6 \pm 14.8$ , 42% male) and compared their lifestyle and cardiometabolic traits to those of SevO persons in the same PRS decile ( $N=121$ , BMI= $45.3 \pm 4.7$ , age  $46.6 \pm 12.6$ , 19.4% male) and to NW in the lowest decile ( $N=330$ , BMI= $22.5 \pm 1.9$ , age  $45.9 \pm 15.0$ , 43% male), adjusting for age, sex, and population substructure.

**Results:** The NW resilient group was significantly healthier than the SevO top PRS group with respect to hypertension (12.4 vs 40.5%), T2D (15.0 vs 41.3%), HDL-C (53.8 vs 45.0mg/dL), and triglycerides (102.1 vs 143.7mg/dL), potentially related to higher physical activity (2.8 vs 1.5hrs/day), shorter time living in the US (19.5 vs 28.3yrs), higher rates of foreign birth (13.1 vs 21.0%), and higher current smoking (34.5 vs 17.4%). Yet, the NW resilient group had more prevalent T2D (16.7 vs 8.5%) than the NW low PRS group, despite similar estimates of central fat. Sensitivity analyses showed consistent effects across Caribbean vs Mainland ancestry.

**Conclusion:** Those resilient to obesity are more physically active, more likely foreign born, and spent fewer years living in the US. Although resilient individuals appear to be metabolically healthy, they have higher T2D prevalence. Identification and characterization of modifiable factors influencing resiliency to genetic risk may lead to changes in how we target, prevent and treat SevO.



## PgmNr 2799: Genome-wide polygenic risk scores for 23andMe phenotypes based on UK Biobank GWAS summary statistics.

### Authors:

W. Wang; S. Shringarpure; A. Auton

[View Session](#) [Add to Schedule](#)

**Affiliation:** 23andMe, Mountain View, California.

---

Polygenic risk scores (PRS) from large scale genome-wide association studies (GWAS) have been commonly applied to the prediction of risk of developing a complex disease. It is expected that larger datasets and improved statistical methodology will result in more accurate predictions, but it remains unclear how PRS generated in one large-scale dataset can be applied to another. For this reason, 23andMe's large-scale and diverse database that combines genotype data with self-reported phenotype data offers a great opportunity to obtain a more comprehensive assessment of PRS approaches.

We applied LDpred on UKBB GWAS summary statistics to construct polygenic risk scores in the 23andMe database and evaluated the performance of the risk prediction for potential prevention.

We constructed PRS with around 4.5 million genetic variants for over 100 phenotypes. We tested the prediction performance of the PRS in the 23andMe database by assessing the prevalence stratified by PRS. We undertook prospective validation on selected phenotypes to check the incidence rate across age based on genetic risk levels, using 23andMe health update data from health history survey where customers update their health status annually.

We assessed the prediction performance over 100 phenotypes, such as type 2 diabetes, deep vein thrombosis, gout, pulmonary embolism, high blood pressure, type 1 diabetes and heart diseases. For each 23andMe phenotype, the prevalence in the high risk group stratified by PRS which developed from UK Biobank GWAS is significantly higher than the prevalence of the rest of the population. For example, the prevalence of type 2 diabetes in the top 5% high risk group is 1.9 times higher than the rest. By checking the 1-year incidence rate of type 2 diabetes across age in 23andMe health update data based on different PRS levels, we found the cumulative 1-year incidence is 0.46% at age 60 and 0.9% at age 90 in the top 5% high risk group, compared to 0.21% at age 60 and 0.46% at age 90 for the remaining population. For the bottom 5% low risk group, the cumulative 1-year incidence is 0.05% at age 60 and 0.24% at age 90, compared to 0.24% at age 60 and 0.5% at age 90 for the rest of the population.

The polygenic risk scores developed from UK Biobank GWAS summary statistics show promise for predicting performance on 23andMe phenotypes. The progression analysis on 23andMe health update data highlights the potential for enhanced screening and prevention based on genetic risk.

## PgmNr 2800: Detection of copy-number duplication at 5q31.3 and its association with obesity in Arab population.

### Authors:

P. Hebbar <sup>1,2</sup>; R. Nizam <sup>1</sup>; S. Elsa John <sup>1</sup>; M. Melhem <sup>1</sup>; J. Pitkaniemi <sup>3</sup>; J. Tuomilehto <sup>3</sup>; F. Al-Mulla <sup>1</sup>; T. Alphonse Thanaraj <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Bioinformatics, Dasman Diabetes Institute, Shark, Kuwait city, Kuwait; 2) Doctoral program in Population Health, Department of Public Health, Faculty of Medicine, University of Helsinki, Finland; 3) Department of Public Health, University of Helsinki, Helsinki, Finland

---

**Aim:** Obesity is a growing health concern for the Arabian Peninsula. Yet, there is a lack of convincingly determined genetic risk variants - single nucleotide polymorphisms (SNP) and copy number variants (CNV) - for adult obesity in Arabs. In this study, we aimed at delineating copy number variations associated with quantitative traits relating to obesity in Arab population.

**Method:** We performed a genome-wide association analysis of CNVs identified from genome-wide genotyping data from 950 native Arab adult individuals from Kuwait and performed validation with whole-exome sequencing data from a cohort of 191 native Arab adult individuals. We used PennCNV program to call CNVs from genome-wide genotype florescent intensity signals and PennCNV-ExomeSeq to call CNVs from exome sequence read depths. We used ParseCNV suite to create individual files for Deletion and Duplication and then to perform statistical association tests using linear regression on quantitative phenotypes such as BMI, Weight, Waist circumference, Height.

**Results:** CNV discovery using genome-wide genotyping data delineated a duplication at genomic region chr5:140201219-140210616 comprising *PCDHA1-PCDHA6* genes. Further validation using exome sequencing data also confirmed CN-duplication at this region. Both genome-wide genotyping and validation using exome data showed consistency in CNV calls with 6% frequency for CN-duplication. Statistical association of CN-duplication showed that the CN-duplication associated with elevated levels of anthropometric traits such as BMI (beta=2.659; p-value=0.0006) and Weight (beta= 6.562; p-value=0.003).

**Conclusion:** The CN-duplication comprising protocadherin genes has been previously associated with childhood obesity and extreme obesity in Korean population. Our study further supports the involvement of protocadherin genes in obesity traits in Arab population.

## PgmNr 2801: Alterations in serum biomarkers and bioavailable testosterone are associated with loss of the Y chromosome in UK Biobank men.

### Authors:

S. Lin <sup>1</sup>; W. Zhou <sup>1,2</sup>; M. Yeager <sup>1,2</sup>; S. Chanock <sup>1</sup>; M. Machiela <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; 2) Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Frederick, MD

---

Post-zygotic genetic alterations occur in hematopoietic stem cells and can lead to clonal mosaicism. Clonal mosaicism has been associated with increased risk of chronic diseases such as leukemia and cardiovascular disease; however, the biological mechanisms linking mosaicism to disease risk are largely unknown. Detected mosaicism could serve as a proxy of genomic maintenance or may reflect a cellular environment favorable for chronic disease onset. Mosaic loss of chromosome Y (mLOY) is the most frequently detected large mosaic copy number alteration with studies reporting frequencies of mLOY >20% in elderly men. We investigated potential associations between mLOY and serum biomarkers in UK Biobank men. After exclusion of men with cancer, the analytic cohort consisted of 208,775 men age 37 to 73 years old. We detected mLOY by scanning for deviations in the median log<sub>2</sub> R ratio (mLRR) of array probes in male-specific region of the Y chromosome. Blood biochemistry measurements were regressed against chromosome Y mLRR and related risk factors (age, ethnicity, smoking, alcohol consumption, body mass index, and disease diagnosis) to identify possible associations. Our analysis identified associations between mLOY and increased total testosterone, sex-hormone binding globulin (SHBG), high-density lipoprotein (HDL), apolipoprotein A (apoA) as well as decreased urate, triglyceride, low-density lipoprotein (LDL), cystatin C, cholesterol, and apolipoprotein B (apo B). We also performed analyses on the concentration and percentage of free and bioavailable testosterone and found lower levels of free and bioavailable testosterone in men with mLOY. To further investigate the association between mLOY and blood biochemistry measurements, we constructed a polygenic risk score (PRS) for mLOY based on 156 known SNPs associated with mLOY. Using this mLOY PRS, we found associations between mLOY and blood biochemistry panels persisted in apo B, cholesterol, LDL, SHBG, urate, free and bioavailable testosterone. Collectively, these results suggest modified sex hormone and serum biochemistry levels in men with mLOY. Further longitudinal studies are needed to disentangle whether these observations reflect a common cause between alterations in these serum biomarkers and mLOY or if alterations in these serum levels are potentially risk factors for or outcomes of mLOY.

## PgmNr 2802: Impact of reference panel choice for imputation on genome-wide association study results for type 2 diabetes in Arab population.

### Authors:

H. Almeer <sup>1</sup>; O. Alsmadi <sup>2</sup>; N. Elkum <sup>3</sup>; M. Saad <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar; 2) King Hussein Cancer Center, Amman, Jordan; 3) Research Department, Sidra Medicine, Doha, Qatar

---

The prevalence of Type 2 Diabetes Mellitus (T2DM) varies substantially among ethnicities because of dissimilarity in lifestyle and genetic makeup. Genome-Wide Association Studies (GWAS) with T2DM have mostly been conducted in populations of European and Asian origin, and it remains to be confirmed whether they will be similar in the Arab population.

In this study, we conducted a GWAS for T2DM in a cohort of Kuwait-resident Arab population (498 cases and 1141 controls), genotyped on the Illumina HumanCardio-MetaboChip array. We evaluated the accuracy of imputation with Minimac3 and its impact on GWAS results using four reference panels: (1) 1000 Genomes Project data (1KG), (2) a public dataset of 108 healthy Qataris with Whole Genome Sequence (Q108), (3) the two previous datasets merged (1KG+Q108), and (4) a cohort of ~1000 Arab subjects, containing both healthy and T2DM patients, with Whole Exome Sequence.

Our preliminary results showed that the 1KG imputation generally led to the best performance across different Allele Frequency intervals. The Q108 imputation yielded similar performance to 1KG imputation for common variants despite its small size. Merging the two datasets decreased imputation performance.

For all imputations we performed, we confirmed two T2DM susceptibility genes, *TCF7L2* (Lead SNP was rs34872471, *P* value=9.58E-08, OR=1.6, 95% CI = [1.34,1.9]) and *GLIS3* (Lead SNP was rs10814915, *P* value 2.57E-05, OR =1.44, 95% CI = [1.22,1.71]). We also identified three genes, using 1KG imputation only, that could play a role in T2DM. These genes were *KCNE4*, *PTPRT*, and *PTRF*.

## **PgmNr 2803: Analyses of variants in GCK, RAPGEF, G6PC2, IGF, related genes for type 2 diabetes and fasting plasma glucose risk in Korean repeated community-based cohort.**

### **Authors:**

J. Lee; K.Y. Jung; S. Moon; Y.J. Kim; B.J. Kim

[View Session](#) [Add to Schedule](#)

**Affiliation:** National institute of health, Korea, Cheongju-si, Chungcheongbuk-do, Republic of Korea

Type 2 diabetes mellitus (T2DM) is one of complex disease (T2DM). By association for various traits related to T2DM, important variations were reported consistently. In National Institution of Health in Korea, 10,038 participants were selected by the Korean Genome Epidemiology Study project (KoGES) and generated by the Korean Association Resource (KARE) through the Korean Genome Analysis Project (KoGAP). Type 2 diabetes was diagnosed based on the criteria of WHO.

We were analyzed genome-wide association for these traits like Type 2 diabetes mellitus, FPG(Fasting plasma glucose), BMI(body mass index), etc. and then also considering their relation. Moreover, Glycemic traits were used screening diabetes. We focused two traits (T2DM and FPG) in Korean repeated community based on cohort because FPG trait also diagnosed T2DM. We had analyzed clinical, epidemiological data that community based on cohort in two repeated (baseline and 12-year follow-up). We also genotyped and imputed using 1000 Genomes.

The association of each traits baseline and repeated were same tendency or effects in baseline and repeated (12-year follow-up).

The genes that had different aspects for FPG and T2DM were analyzed using linear regression adjust BMI (Body Mass Index). Especially, rs1799884 was significant for FPG in baseline and follow-up(P Value=5.37E-05, 2.97E-04) and also for T2DM(PValue=2.07E-08, 5.57E-05). Estimations of absolute beta had a tendency for increasing absolute beta estimates in follow-up data compare to baseline. Our results maintained even whenever variant is not significant in each trait (FPG or T2DM). Most studies that GCK gene is associated with T2DM or FPG. In Korean population had not reported the association for T2DM. However, we have got the results that association for FPG and T2DM but its results is significant. Our finding that rs1799884 is significant for two traits (FPG and T2DM) could help to identify high risk group of T2DM than any other locus.

**Acknowledgements:** This study was supported by intramural grants from the Korea National Institute of Health (2019-NI-097-00) and with bioresources from National Biobank of Korea, the Center for Disease Control and Prevention, Republic of Korea.

## PgmNr 2804: The functional impacts of rare coding variants in 46,000 individuals on 23 quantitative phenotypes.

### Authors:

P. Dornbos<sup>1,2,3</sup>; R. Koesterer<sup>1,2</sup>; M. Udler<sup>1,2,3,4</sup>; L. Caulkins<sup>1,2</sup>; J. Flannick<sup>1,2,3,5</sup>; Accelerating Medicines Partnership (AMP) T2D-GENES Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Metabolism Program, Broad Institute of MIT and Harvard, Cambridge, MA; 2) Program in Medical & Population Genetics, Broad Institute, Cambridge, MA; 3) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 4) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 5) Department of Pediatrics, Harvard Medical School, Boston, MA

---

While common variant genome-wide association studies (GWAS) have identified many complex trait-associated genetic loci, identifying causal genes and mechanisms remains a challenge. Through an analysis of 23 cardiometabolic quantitative traits (QTs) in 46,000 exome sequenced individuals from five ancestries, we ask to what extent can rare coding variants complement GWAS by providing insight into (a) effector genes that mediate GWAS associations and (b) specific molecular gene perturbations that impact disease risk. Exome-wide analysis indicated a total of 394 variant associations that were significant; nearly all are within 250 kb of known GWAS loci and collectively suggest >20 genes as novel effector candidates. Gene-level analyses of rare loss-of-function and damaging missense mutations produced 34 exome-wide significant associations; all are within 250 kb of known GWAS loci and one in particular suggests *ABCA6* as an effector gene which mediates an increase in LDL levels. We next evaluated whether rare coding variants could offer molecular insights even within genes that did not achieve exome-wide significance. First, for 18 QTs, we used the MAGMA gene-level analysis method to calculate gene-level GWAS association scores; the top 100 genes ranked by MAGMA had (collectively) a significant ( $p < 0.05$ ) excess of rare coding variant associations for six QTs and a near significant excess ( $p < 0.1$ ) for four QTs. Second, for six QTs, we curated gene-sets implicated in the trait by mouse knockouts and found that three had a significant excess of rare coding variant associations with another trait near significance. These results suggest that many disease-relevant genes harbor rare coding “allelic series” with varied molecular and phenotypic effects. In many cases, distinct allelic series exist for multiple traits; we find 56 genes in which distinct sets of variants associate with at least one QT below  $p < 1e-4$  and a second QT below  $p < 0.05$ . Such genes offer the potential to explore selective targeting of a gene as a means to affect one phenotype but not others; for example, within *SLC30A8*, a known type-2 diabetes (T2D) gene, we find three distinct allelic series for: T2D, fasting glucose, and fasting insulin/BMI/hip circumference. Collectively, these results confirm that rare coding variants offer a valuable resource for probing the molecular function of genes. Finally, we describe a public web-portal for viewing these data to assist with downstream functional analyses.

## PgmNr 2805: Building genetic risk scores with penalized regressions to infer whether an environmental factor attenuates or exacerbates the adverse influence of a candidate gene.

### Authors:

W. Lin <sup>1,2</sup>; Y. Lin <sup>1</sup>; C. Chan <sup>2,3</sup>; Y. Liu <sup>4</sup>; S. Tsai <sup>5,6,7</sup>; P. Kuo <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 2) Department of Public Health, College of Public Health, National Taiwan University, Taipei, Taiwan; 3) Institute of Occupational Medicine and Industrial Hygiene, College of Public Health, National Taiwan University, Taipei, Taiwan; 4) Center for Neuropsychiatric Research, National Health Research Institutes, Zhunan, Miaoli County, Taiwan; 5) Department of Psychiatry, Taipei Veterans General Hospital, Beitou District, Taipei, Taiwan; 6) Division of Psychiatry, National Yang-Ming University, Taipei, Taiwan; 7) Institute of Brain Science, National Yang-Ming University, Taipei, Taiwan

---

The detection of “gene-environment interactions” (GxE) is important and is even more challenging than the detection of genetic main effects. Although some gene-based GxE methods have been developed, very few GxE findings have reached the genome-wide significance level (i.e.,  $p < 2.5E-6$ , where  $2.5E-6$  is the commonly used significance level in genome-wide gene-based analyses). Most evidence of GxE was discovered from candidate gene analyses in which genome-wide association study (GWAS) hits were targeted. It is of interest to know whether an environmental factor (E) attenuates or exacerbates the adverse influence of a candidate gene. Different from most set-based gene-environment interaction tests, genetic risk score (GRS) approaches can not only provide a  $p$ -value for the significance of “gene-environment interaction” (GxE) but also guide inferences on how E modifies the effect of a gene. If E exacerbates the adverse influence of a gene, GRS constructed by the elastic net (ENET) or the least absolute shrinkage and selection operator (LASSO) is recommended. However, the performance of ENET or LASSO will be compromised if E attenuates the adverse influence of a gene, in which the sensitivity of selecting trait-associated SNPs is much lower than that in the exacerbation scenario. Instead, using the ridge regression (RIDGE) is more powerful in this situation. Applying RIDGE to 18,424 subjects in the Taiwan Biobank, we showed that performing regular exercise can attenuate the adverse influence of *the fat mass and obesity-associated (FTO)* gene on 4 obesity measures, including body mass index ( $p = 0.0033$ ), body fat percentage ( $p = 0.0097$ ), waist circumference ( $p = 0.0254$ ), and hip circumference ( $p = 0.0023$ ). Moreover, we replicated that *the fibroblast growth factor 5 (FGF5)* gene has a stronger effect on diastolic blood pressure with a higher BMI ( $p = 0.0106$ ). Our approaches provide an important inference on whether an environmental exposure attenuates or exacerbates the adverse influence of a candidate gene.



## PgmNr 2806: Imputing gene-expression in the Population Architecture of Genetic Epidemiology (PAGE) diverse samples.

### Authors:

H.M. Highland <sup>1</sup>; L.M. Huckins <sup>2</sup>; G.L. Wojcik <sup>3</sup>; C.J. Hodonksy <sup>1</sup>; A.R. Baldassari <sup>1</sup>; K.L. Young <sup>1</sup>; S. Bien <sup>4</sup>; PAGE Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Univ North Carolina at Chapel Hill, Chapel Hill, North Carolina.; 2) Icahn School of Medicine at Mount Sinai, New York NY; 3) Stanford University, Stanford CA; 4) Fred Hutchinson Cancer Research Center, Seattle WA

---

Complex traits and diseases are influenced by genetic factors, in part through altering gene expression. Leveraging existing reference panels, such as GTEx, the heritable component of gene expression can be imputed and tested for association with various traits. While these approaches have successfully identified new gene associations across a variety of traits, it remains unclear whether variant predictors derived from predominantly European descent reference panels can be used for discovery in more ancestrally diverse populations. The Population Architecture of Genetic Epidemiology (PAGE) study applied the PrediXcan approach to impute gene expression (GReX) based on models derived from the GTEx Consortium data using GWAS data from up to 49,781 ancestrally diverse PAGE participants (35% African American, 45% Hispanic/Latino, and 10% Asian) to identify 98 novel (>1MB from GWAS signals) genes associated (experiment-wide Benjamini-Hochberg FDR<0.05) with one or more of the 31 cardiometabolic traits evaluated in one or more relevant tissues. These findings include novel associations between QT-interval and *PCCB* GReX in whole blood, aortic, and atrial tissue (FDR= 0.006). *PCCB* encodes a subunit of PCC, for which deficiency causes long QT syndrome. Further, to assess the impact of ancestrally diverse samples, we compared the number of significant gene associations for height and BMI detected in PAGE to those detected in a subset of 50,000 participants from UK Biobank. In whole blood, we identified 82 height associated genes in UK biobank and 20 in PAGE with eight overlapping. For BMI we identified twelve genes in UK Biobank and three in PAGE, with two overlapping. Lastly, using principal component analysis in the PAGE dataset, we compared the extent of genetic differentiation of PrediXcan model variants to a set of control variants matched on various genomic features. We found that variants used for transcriptome imputation are less differentiated in terms of genetic ancestry than would be expected by chance. Collectively, these results suggest that while transcriptome imputation models derived from European-specific reference panels can be used for discovery, the potential gains from such approaches will be significantly diminished for populations with different ancestral backgrounds; therefore, further investment in multitissue transcriptomic data from ancestrally diverse samples is essential to equitable advance multi-omic research.



## PgmNr 2807: Distinct metabolomic signatures of central obesity in the Atherosclerosis Risk in Communities (ARIC) Study.

### Authors:

K.L. Young <sup>1</sup>; H.M. Highland <sup>1</sup>; B. Yu <sup>2</sup>; M. Grove-Gaona <sup>3</sup>; E. Boerwinkle <sup>2</sup>; K.E. North <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Univ North Carolina, Chapel Hill, North Carolina.; 2) University of Texas Health Science Center at Houston, Department of Epidemiology, Human Genetics and Environmental Sciences, Houston, TX, USA; 3) University of Texas Health Science Center at Houston, Human Genetics Center, Houston, TX, USA

While the prevalence of obesity has more than doubled in the US, not all obese individuals have the same risk for adverse health outcomes. As central obesity is more metabolically active, metabolites may provide insights into mechanisms leading to disease. We hypothesize that central obesity (measured as waist-to-hip ratio adjusted for BMI) has an identifiable metabolic signature distinct from metabolites associated with BMI. GWAS of obesity associated metabolites (mGWAS) can identify associations linked to metabolic transporters, regulators, or enzyme coding genes, allowing for discovery of the best biological candidates in known GWAS regions, as well as identifying novel central obesity genes. Our study leverages existing data in the Atherosclerosis Risk in Communities (ARIC) Study (N=4,032), a prospective, longitudinal cohort study of cardiovascular disease in European and African Americans (AA), to identify metabolomic signatures of central obesity.

We estimated associations between baseline natural log (ln) transformed WHRadjBMI and 245 baseline standardized serum metabolites, using ancestry and sex-stratified additive linear models, followed by meta-analysis in SAS 9.3. We identified 7 metabolites significantly associated with increased lnWHRadjBMI ( $p=2.04E-4$  (0.05/245\*2 traits) – including 1-stearoylglycerophosphoethanolamine – as well as 5 metabolites associated with decreased lnWHRadjBMI – including g-glutamylvaline and g-glutamylleucine. Notably, none of these metabolites were associated with BMI. All 7 metabolites positively associated with lnWHRadjBMI are lipids – supporting prior GWAS implicating lipid pathways, while 4 of the metabolites associated with decreased lnWHRadjBMI are amino acids or peptides.

Metabolites significantly associated with lnWHRadjBMI were carried forward for mGWAS ( $n=12$ ), followed by meta-analysis in METAL. An intronic variant in *ALDH1A2* is associated with increased 1-stearoylglycerophosphoethanolamine levels in all strata ( $\beta=0.24$ ,  $SE=0.03$ ,  $p=8.18E-22$ ,  $EAF=0.39$ ); this variant was previously associated with plasma lipid levels, including HDL and triglycerides. Two variants near *GGT1* are associated with  $\gamma$ -glutamylleucine ( $\beta=0.33$ ,  $SE=0.05$ ,  $p=6.23E-11$ ,  $EAF=0.82$ ) and  $\gamma$ -glutamylvaline ( $\beta=0.50$ ,  $SE=0.08$ ,  $p=6.55E-10$ ,  $EAF=0.93$ ) levels in AA only; *GGT1* codes for an enzyme involved in protein metabolism. These results highlight the potential for metabolomics to identify new central obesity signals and refine obesity phenotypes.

## PgmNr 2808: Investigating the use of machine learning methods to build risk prediction models for complex disease.

### Authors:

J.P. Cook <sup>1</sup>; J.Y. Goulermas <sup>2</sup>; A.P. Morris <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; 2) Department of Computer Science, University of Liverpool, Liverpool, United Kingdom

---

Large-scale population biobanks offer exciting opportunities to develop risk prediction models for complex diseases because of the availability of genetic data and extensive lifestyle and clinical information. Unlike traditional polygenic risk scores, machine learning methods can be utilised to build risk prediction models that include both genetic and non-genetic features, and interactions between them.

We have performed a simulation study to assess the utility of several machine learning methods (gradient boosting machines, deep learning neural networks, and random forests) to generate prediction models for type 2 diabetes (T2D), applied using the H2O package, using data from the UK Biobank. 20,000 participants were randomly selected according to their T2D status (10,000 cases and 10,000 controls). Five relevant clinical factors (age, sex, body mass index, diastolic blood pressure and systolic blood pressure) were selected for entry into the model alongside a set of 1-100 SNPs, simulated with varying minor allele frequency and relative risk of disease. Irrelevant clinical factors were also selected to assess whether the methods identify them as unimportant for disease prediction.

All methods successfully identified the most strongly associated genetic and non-genetic factors as the most important features for prediction, and assigned the least importance to the irrelevant factors. Results also indicated that the inclusion of strongly associated genetic variants increases the predictive accuracy of the model compared to using clinical factors alone, while the inclusion of more modestly associated variants does not appear to improve predictive power.

## PgmNr 2809: Type 2 diabetes and autonomic function: A bidirectional Mendelian randomisation study in the UK Biobank.

### Authors:

A. Farmaki <sup>1</sup>; V. Garfield <sup>1</sup>; S.V. Eastwood <sup>1</sup>; R.E. Farmer <sup>2</sup>; R. Mathur <sup>2</sup>; K. Bhaskaran <sup>2</sup>; L. Smeeth <sup>2</sup>; N. Chaturvedi <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Population Science and Experimental Medicine, University College London, London, United Kingdom; 2) Department of Non-communicable Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom

---

People with type 2 diabetes (T2DM) are at risk of vascular and neurological complications. The autonomic nervous system (ANS) controls metabolically relevant organs such as the liver, pancreas and adipose tissue. This study aims to explore the direction of causality between ANS function, using heart rate as a proxy, and T2DM.

Data from 400,000 Europeans in UK Biobank were used and T2DM status assigned using a validated algorithm. Mendelian randomisation analyses used an inverse-variance weighted (IVW) approach, and MR-Egger as a sensitivity analysis for horizontal pleiotropy. Independent variants ( $R^2 < 0.3$ ) from the latest genome-wide association studies were used as genetic instruments. Firstly, the causal effects of heart rate on T2DM status, overall obesity (body mass index-BMI) and central obesity (waist to hip ratio-WHR) were examined, using a genetic instrument with 14 variants. Secondly, to investigate the causal effect of T2DM, BMI and WHR on heart rate, 163, 899 and 366 genetic variants were used, respectively.

Using IVW, there was no evidence for a causal effect of heart rate on T2DM (OR [95%CI]: 0.998 [0.981, 1.016]), BMI ( $\beta$  [95%CI]: 0.003 [-0.015, 0.022]) or WHR ( $\beta$  [95%CI]: 0.017 [-0.014, 0.048]) and no evidence of horizontal pleiotropic effects ( $P_{\text{Egger intercept}} > 0.1$ ). However, there was a modest causal effect of T2DM and BMI on heart rate ( $\beta$  [95%CI]: 0.310 [0.171, 0.449] and 0.871 [0.601, 1.140], respectively) with no evidence of horizontal pleiotropy ( $P_{\text{Egger intercept}} > 0.1$ ). Although IVW suggested a causal effect of WHR on heart rate ( $\beta$  [95%CI]: 1.627 [1.208, 2.046]) there was likely to be unbalanced horizontal pleiotropy ( $P_{\text{Egger intercept}} = 0.026$ ).

Findings suggest that diabetes and obesity may cause autonomic dysfunction, with no evidence for the reverse direction. However, there is a need for better autonomic function genetic instruments to robustly study the effect of ANS dysfunction on diabetes and obesity.

## PgmNr 2810: Mendelian randomisation reveals that testosterone causally improves the metabolic profile of middle-aged men.

### Authors:

K. Kentistou<sup>1,2</sup>; P. Joshi<sup>1</sup>; P. Timmers<sup>1</sup>; J. Wilson<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, EH8 9AG, UK.; 2) Centre for Cardiovascular Sciences, Queen's Medical Research Institute, University of Edinburgh, Edinburgh, EH16 4TJ, UK.; 3) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK

---

**Motivation:** Androgens have been repeatedly shown to regulate fat mass and body composition in mammals. Several cardiometabolic diseases (CMD), such as heart disease, diabetes and dyslipidaemia are associated with a reduction in circulating levels of testosterone, while androgen supplementation has been shown to improve blood lipid levels and general metabolic health, in men. Observational studies report the opposite in women, whereby increased androgen levels increase abdominal adiposity and the risk for cardiovascular disease. This relationship between sex hormones and CMD is seen as cyclical, i.e. it is difficult to distinguish forward or reverse causality between them. Here, we use a mendelian randomisation (MR) approach to estimate the direct causal effect of testosterone on anthropometric and CMD traits.

**Methods:** A GWAS of testosterone was conducted in UK Biobank (n=278,405) and genetic instruments were created, by LD pruning of the summary statistics. Non-overlapping publicly available GWAS on cardiometabolic phenotypes were used as outcomes. These included sex-stratified measures of adiposity (BMI, WHR, etc.) and only sex-combined blood lipids and incidence of disease. Inverse variance weighted MR was used to determine the effects of the hormonal exposures on the outcomes.

**Results:** For anthropometric measures, where sex-stratified analyses were possible, we found an effect of testosterone on waist circumference of -0.08 (SE=0.03, P=5E-3) in men. This means that an increase in circulating testosterone of 4 nM, reduces waist circumference by 1 cm. This effect size is the same as the observational association, within UK Biobank. Due to moderate power, we were not able to preclude small deleterious effects in women, but any such effects appear to be smaller to the beneficial ones in men. In the sex combined analyses of blood lipid levels, testosterone causally decreased triglycerides (P=6E-9) and increased HDL cholesterol levels (P=3E-7), confirming a metabolically favourable effect.

**Conclusions:** Our results show a causal effect between testosterone and metabolic health in men, recapitulating the observational association between them and reinforcing intervention studies in the literature. Delineating the effect of testosterone and other sex-hormones on metabolic health and disease could lead to novel preventative insights.

## PgmNr 2811: Expression levels of *F13A1* transglutaminase in adipose tissue correlates with metabolic health in obesity and with genes linked to hypertrophic adipocyte phenotype.

### Authors:

M. Arora <sup>1</sup>; K.H. Pietiläinen <sup>2</sup>; M.T. Kaartinen <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Experimental Medicine, Faculty of Medicine, McGill University, Montreal, Quebec, Canada; 2) Obesity Research Unit, University of Helsinki, Finland; 3) Division of Biomedical Sciences, Faculty of Dentistry, McGill University, Montreal, Quebec, Canada

---

Obesity increases the risk of developing debilitating and lethal comorbidities, however, it is estimated that 10-25% of obese humans remain healthy. This *metabolically healthy obesity* (MHO) has been speculated to arise from protection from development of insulin resistance and altered adipose tissue properties of individuals. To investigate adipose tissue-genes causative to obesity and metabolic health, Pietiläinen group examined genome wide gene expression in adipose tissue and adipocytes of monozygotic Finnish, BMI-discordant lean-obese twin pairs. The analyses identified 28 potentially causative genes in these pairs; out of which, the gene with strongest causative relation to obesity was *F13A1* (Factor XIII-a transglutaminase) which we have linked to metabolic health in mouse studies. It was also reported that the BMI-discordant monozygotic twins could be divided into groups that demonstrated metabolic differences; MHO-Group 1: insulin sensitive, low liver fat% in obese twin, and metabolically sick obesity MSO-Group 2: less insulin sensitive, high liver fat% in obese twin. Here we have used the twin GWA/Affymetrix mRNA and metabolic data to examine *F13A1* mRNA levels and correlation with possible markers of adipogenesis, tissue expansion and metabolic disturbances. The aim of this study was to examine expression levels of *F13A1* and its link to metabolic health. We performed statistical analyses of the data set (52 individuals, 26 twins) by means of regression, correlations, and descriptive statistics, in the MHO-Group 1 and MSO-Group 2. We report that *F13A1* mRNA levels are significantly higher in obese twins and highest in the obese twins of the MSO group. Conversely, lowest averaged levels were observed in the lean twins of MHO group. Adipose tissue cellularity in the MHO group was significantly higher compared to MSO where adipocyte size was significantly higher, suggesting different mode of expansion (hyperplasia vs. hypertrophy, respectively) that may be genetically regulated. Linear regression and clustering analysis between *F13A1* and expression of genes known to be highly produced in large-adipocytes showed strongest linear correlation to IL8 ( $R^2=0.81$ ), CXCL2 ( $R^2=0.62$ ) (inflammatory cytokines) and DCN ( $R^2=0.59$ ), (an extracellular matrix component). In summary, the data suggests that *F13A1* levels in adipose tissue link to metabolic health outcome of obesity and may be linked to adipose tissue capacity to expand and to respond to energy surplus.

## PgmNr 2812: Cross-phenotype association analysis identifies 62 adiposity-increasing loci that protect against cardiometabolic comorbidities.

### Authors:

L.O. Huang <sup>1</sup>; A. Rauch <sup>6,7</sup>; U. Schick <sup>2</sup>; N. Yang <sup>2</sup>; S. Mandrup <sup>6</sup>; T.O. Kilpeläinen <sup>1,5</sup>; R.J.F. Loos <sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genomics and Metagenomics in Metabolism, Novo Nordisk Foundation Center for Basic Metabolic Research, Copenhagen, Denmark, DK-2200; 2) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai Hospital, New York, NY 10029 USA; 3) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai Hospital, New York, NY 10029 USA; 4) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai Hospital, New York, NY 10029 USA; 5) Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai Hospital, New York, NY 10029 USA; 6) Functional Genomics & Metabolism Research Unit, Department of Biochemistry and Molecular Biology, University of Southern Denmark, Odense M, Denmark, DK-5230; 7) Molecular Endocrinology & Stem Cell Research Unit (KMEB), Department of Endocrinology and Metabolism, Odense University Hospital and Department of Clinical Research, University of Southern Denmark, Odense C, Denmark, DK-5000

---

**Background:** Some individuals remain cardiometabolically healthy despite being obese, the so-called metabolically healthy obese (MHO). The biological mechanisms that underlie MHO are poorly understood in current literature. In this study, we aim to 1) identify novel adiposity-increasing variants with protective cardiometabolic effects, 2) identify the tissues where these loci are enriched in, 3) find experimental evidence that links adipocyte function to these loci.

**Methods:** 1) We performed pairwise meta-analysis of adiposity and cardiometabolic traits from summary statistics of 11 GWAS using CPASSOC, including three adiposity traits (body fat percentage, BMI, and waist-to-hip ratio) and eight cardiometabolic traits (HDL-C, LDL-C, triglycerides, fasting insulin and glucose, and blood pressure levels, and incident coronary artery disease and type 2 diabetes). We searched for genome-wide significant associations ( $P < 5 \times 10^{-8}$ ) with any of the 24 trait pairs from one adiposity and one cardiometabolic trait. 2) We used DEPICT to perform tissue enrichment analysis on 62 loci and 3 gene clusters identified through K-means clustering. 3) We conducted *in vitro* experiments on human mesenchymal stem cells to ascertain functional role of the identified loci in adipocytes.

**Results:** 1) The meta-analysis identified 62 independent loci of which 32 are novel, several others reside in gene regions such as *IRS1*, *COBLL1*, *PPARG*, *TOMM40*, and *VEGFA*, which are previously reported MHO loci. 2) Tissue enrichment analyses implicate strongest enrichment in adipose tissue, adrenal cortex, adrenal gland, liver, and arteries. In cluster-specific analysis, we identified a cluster of 7 SNPs for which the BMI/BFP increasing alleles associated with protective effect on cardiometabolic outcomes, mediated through favorable fat distribution with lower WHR. Tissue enrichment analysis of this cluster shows a broad impact on adipose tissues, liver, digestive and cardiovascular systems. 3) We found evidence that the 62 SNPs are enriched for variants linked to differential gene expression in adipose tissue as well as in regulatory elements that gain activity during adipogenesis.

**Conclusion:** Our integrated approach using GWAS meta-analysis, computational tools, and *in vitro* experiments identified novel MHO loci and provided insight into biological pathways that may underlie MHO, and could shed light on future drug development targeting improvement in cardiometabolic profile of unhealthy obese individuals.

## **PgmNr 2813: Association of genetic risk scores for adiposity-related traits with body fat percentage and related phenotypes in the Jackson Heart Study.**

### **Authors:**

M. Yaser <sup>1</sup>; L.M. Raffield <sup>2</sup>; L.A. Lange <sup>3</sup>; A. Correa <sup>4</sup>; K.C. Taylor <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Epidemiology, School of Public Health & Information Sciences, University of Louisville, Louisville, KY, USA; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 3) Department of Medicine, University of Colorado Denver, Aurora, CO, USA; 4) Departments of Medicine, Pediatrics and Population Health Science, University of Mississippi Medical Center, Jackson, MS, USA

---

### **Background:**

Compared to European ancestry individuals, African ancestry individuals with comparable anthropometric measures have significantly lower abdominal adiposity deposits. Here we investigate the extent to which genetic risk scores for well-studied adiposity phenotypes, such as body mass index (BMI), also predict other adiposity measures, such as % body fat or visceral:subcutaneous (VAT:SAT) ratio, in order to better understand the genetic underpinnings of different patterns of adiposity.

### **Methods:**

Genetic risk score analyses for anthropometric measures were undertaken among 2420 African Americans aged 28-97 years from Jackson Heart Study (JHS). Anthropometric polygenic risk scores (PRS) for BMI, waist-hip ratio (adjusted for BMI), waist circumference (WC) (adjusted for BMI), and % body fat mass were calculated for each individual as the weighted sum of phenotype-increasing alleles. Genome-wide significant variants from prior studies using European or African ancestry individuals in the GWAS catalog were included. Associations between each PRS and inverse normally transformed adiposity measures were examined using multivariable linear regression models adjusted for age, gender, and up to 10 ancestry principal components.

### **Results:**

The BMI PRS was found to be a significant positive predictor of % body fat mass ( $\beta=0.005$  per allele,  $P=2.19\times10^{-3}$ ) and subcutaneous adiposity ( $\beta=0.004$ ,  $P=2.38\times10^{-2}$ ). The % fat mass PRS was significantly associated with subcutaneous ( $\beta=0.022$ ,  $P=1.70\times10^{-10}$ ) but not visceral adiposity ( $\beta=0.011$ ,  $P=5.64\times10^{-2}$ ). No significant associations were found between % fat mass PRS or BMI PRS with waist circumference, waist-hip ratio or VAT:SAT ratio. Likewise, the WC-PRS and WHR-PRS were not significantly associated with % fat mass.

### **Conclusion:**



These analyses suggest; (a) that genetically driven increases in percentage body fat mass strongly associate with subcutaneous adiposity but not with the more deleterious visceral adiposity; and (b) that percentage body fat mass is strongly associated with overall BMI associated genetic variants but not with variants associated with measures of central adiposity (WC and WHR). More study is needed to understand how genetic variants associated with these anthropometric traits may contribute to observed differences in adiposity patterns between African and European ancestry individuals.

## PgmNr 2814: A PheWAS and LabWAS approach in studying rs560887 SNP in the *G6PC2* gene using BioVU.

### Authors:

K. Singh <sup>1,2</sup>; S. Goleva <sup>1,2,3</sup>; K.J. Bosma <sup>3</sup>; R. O'Brien <sup>3</sup>; L.K. Davis <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN

---

The glucose-6-phosphatase (G6Pase) enzyme system is located in the endoplasmic reticulum and catalyzes the hydrolysis of glucose-6-phosphate (G6P) to glucose and inorganic phosphate. G6Pase system is composed of a glucose transporter and a G6P/Pi transporter, encoded by the *SLC37A4* gene along with a catalytic subunit, which can be one of three isoforms, G6PC1, G6PC2 or G6PC3. *G6PC2* is expressed predominantly in pancreatic islet beta cells. G6Pase activity is abolished in *G6pc2*KO mice, which results in a leftward shift in the dose response curve for glucose-stimulated insulin secretion (GSIS) leading to a reduction in fasting blood glucose (FBG). Consistent with these mouse studies, genome wide association studies have linked the rs560887 SNP in the *G6PC2* gene to variations in FBG. Molecular studies have shown that the rs560887-G allele represents a gain of function that is associated with increased *G6PC2* RNA splicing which is predicted to lead to increased full length G6PC2 protein expression and elevated glucose cycling. Our aim here was to find other disease associations with rs560887-G allele. We performed a phenome-wide association study to examine the effect of this allele on 2000 different medical diagnoses present in the electronic health records (EHR) of 62,400 patients treated at Vanderbilt University Medical Center. The rs560887-G allele was significantly associated with acute pancreatitis (OR = 1.37,  $p = 1.75 \times 10^{-5}$ ) in patients without Type 2 diabetes, after adjusting for sex, genetic ancestry, age, and genotyping batch. We also conducted a Lab-wide association study including 272 heritable lab traits commonly measured in a clinical setting. We found rs560887-G allele to be significantly associated with decreased glucose ( $\beta = -0.049$ , SE = 0.005,  $p = 2.4 \times 10^{-22}$ ), which confirmed molecular studies in mice. The direction and magnitude of the associations were the same for both diabetic and non-diabetic patients but failed to reach statistical significance after correction for multiple testing in the Type 2 diabetic population. The rs560887-G allele was also significantly associated with decreased HbA1C ( $\beta = -0.045$ , SE = 0.011,  $p = 2.9 \times 10^{-5}$ ) in patients without Type 2 diabetes, which matches the GWAS data. Our future analysis will focus on studying how the SNP associates in populations with diverse ancestral backgrounds including African, Asian, and European ancestry.

## PgmNr 2815: A multivariate Bayesian genetic association analysis of a *CREBRF* variant and adiposity-related phenotypes.

### Authors:

J.Z. Zhang <sup>1</sup>; J.C. Carlson <sup>1</sup>; N.L. Hawley <sup>3</sup>; G. Sun <sup>4</sup>; H. Cheng <sup>4</sup>; T. Naseri <sup>5</sup>; M.S. Reupena <sup>6</sup>; R. Deka <sup>4</sup>; S.T. McGarvey <sup>7,8</sup>; R.L. Minster <sup>2</sup>; D.E. Weeks <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 3) Department of Epidemiology (Chronic Disease), School of Public Health, Yale University, New Haven, CT; 4) Department of Environmental Health, College of Medicine, University of Cincinnati, Cincinnati, OH; 5) Ministry of Health, Government of Samoa, Apia, Samoa; 6) Lutia i Puava ae Mapu i Fagalele, Apia, Samoa; 7) International Health Institute, Department of Epidemiology, School of Public Health, Brown University, Providence, RI; 8) Department of Anthropology, Brown University, Providence, RI

While obesity has become a global epidemic, it is especially prevalent within the Samoan population. Over 80% of the Samoan adults are classified as either overweight or obese. It is therefore important to understand the interlinkages between genotypes and adiposity-related phenotypes. Previous studies have shown that a minor allele of *CREBRF* (encoding *CREB3* regulatory factor) rs373863828 (p.Arg457Gln) is strongly associated with BMI and reduced risk of type 2 diabetes in Polynesians. Although this variant is exceedingly rare in non-Polynesian populations, it is relatively common in the Samoan and Polynesian populations (MAF = 0.259); rs373863828 may have been subject to positive selection during historic periods of food scarcity on Polynesian islands.

The study population consisted of 1,810 unrelated Samoans from the Independent State of Samoa. We performed a Bayesian association test between rs373863828 and a panel of 10 anthropometric and lipid profile measurements – body mass index (BMI), fat mass index (FMI), fat-free mass index (FFMI), height, waist circumference, hip circumference, HDL cholesterol (HDL-C), LDL cholesterol (LDL-C), net triglycerides, and total cholesterol. The phenotype panel was quantile normalized to multivariate normality to reduce the risk of false positive associations due to influence of outlier observations. Analysis was carried out with mvBIMBAM, a Bayesian multivariate framework that has been shown to be more powerful than univariate association analysis. In addition to BMI, this variant has univariate associations with hip circumference, abdominal circumference, and height.

Under a  $\log_{10}$  Bayes factor threshold of 5.0, significant associations were found between rs373863828 and the overall phenotype panel ( $\log_{10}$  Bayes factor = 6.5), BMI (5.9), FMI (6.2), and abdominal circumference (5.2). There was evidence for a suggestive association with hip circumference (4.9) and a lack of evidence of association with FFMI (3.5), total cholesterol (1.4), net triglycerides (-0.2), LDL-C (0.8), HDL-C (-0.4), and height (0.1). The analysis, which accounts for multivariate correlation structures between phenotypes, provides further evidence for the association of rs373863828 with measures of total and regional adiposity while not supporting the previously observed univariate association with height. Furthermore, these results suggest that adipose tissue deposits as opposed

to fat-free mass increases potentially contribute to increases in BMI.

## PgmNr 2816: Homozygosity mapping of pulmonary and asthma-related phenotypes in minority populations in the US.

### Authors:

Z.A. Szpiech<sup>1</sup>; K.L. Keys<sup>2</sup>; M.J. White<sup>2</sup>; A.M. Zeiger<sup>2</sup>; A. Li<sup>2</sup>; O.S. Risse-Adams<sup>2</sup>; P. Huang<sup>2</sup>; M.G. Contreras<sup>2</sup>; J. Elhawary<sup>2</sup>; E.Y. Lee<sup>2</sup>; L.-A. Samedy<sup>2</sup>; A.C.Y. Mak<sup>2</sup>; C. Eng<sup>2</sup>; D. Hu<sup>2</sup>; S. Huntsman<sup>2</sup>; S.S. Oh<sup>2</sup>; S. Salazar<sup>2</sup>; E.G. Burchard<sup>2</sup>; R.D. Hernandez<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Auburn University, Auburn, AL; 2) Department of Medicine, University of California San Francisco, San Francisco, CA; 3) Department of Human Genetics, McGill University, Montreal, QC; 4) Genome Quebec Innovation Center, McGill University, Montreal, QC

---

Homozygosity mapping has long been useful for identifying the genetic basis of recessive Mendelian diseases in inbred pedigrees. However, with increased density of genetic markers and improved methodology for detecting short autozygous segments, homozygosity mapping has become a viable approach for studying complex phenotypes in outbred populations. We use homozygosity mapping to investigate the genetic basis of asthma-related and lung function phenotypes in three minority populations in the US: African Americans, Puerto Ricans and Mexican Americans. Asthma is a chronic inflammatory disorder characterized by recurrent respiratory symptoms and reversible airway obstruction. It is the most common chronic childhood disease with variable prevalence across our study populations, with Puerto Ricans having the highest rates (36.5%), African Americans having intermediate rates (13.0%), and Mexican Americans having the lowest rates (7.5%).

We perform homozygosity mapping of both asthma status and an important measure of lung function, the forced vital capacity (FVC) of a patient's lungs. We find several population-specific homozygous regions significantly associated with these phenotypes, including several on or near genes previously associated with lung disease or lung function phenotypes. Many of these associations were found in the direction of a protective effect or increased lung function, suggesting that the underlying haplotypes in these associated regions may contain protective/beneficial variants related to asthma and lung function. Furthermore, these mapped regions did not appear significant when performing standard GWAS on the genotypes of these same samples, suggesting that our homozygosity mapping approach provides additional value over standard GWAS analyses.

## PgmNr 2817: Large scale GWAS of hypothyroidism and health risks in treated vs. untreated hypothyroidism.

### Authors:

M.P. Reeve<sup>1</sup>; W. Zhou<sup>2</sup>; J. Karjalainen<sup>1,2</sup>; M. Kurki<sup>1,2</sup>; P. Della Briotta Parolo<sup>1</sup>; C. Willer<sup>4,5,6,7</sup>; B.O. Åsvold<sup>8,9</sup>; M. Daly<sup>1,2,3</sup>; A. Palotie<sup>1,2,3</sup>; S. Ripatti<sup>1,2,3</sup>; FinnGen

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 4) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA; 5) Department of Internal Medicine, Division of Cardiology, University of Michigan Medical School, Ann Arbor, Michigan, USA; 6) Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA; 7) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 8) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 9) Department of Endocrinology, St. Olavs hospital, Trondheim University Hospital, Trondheim, Norway

---

The prevalence of overt hypothyroidism in Europe is between 0.2% and 5.3% and occurs more frequently in women, patients with autoimmune diseases, individuals >65, and whites. In iodine-sufficient areas, autoimmune thyroiditis (Hashimoto's disease) is the most common cause - though because of the lack of specific symptoms, there is up to a 5% rate of untreated hypothyroid disease. We undertook a combined meta-analysis of data from FinnGen (n=140,000) and UK Biobank (n=330,000) and found >50 independently associated loci. In addition to confirming previously documented hits (including PTPN22, SH2B3, FOXE1, IFIH1, TYK2) we identify about 40 novel hits including several corresponding to alleles specifically enriched through the founding bottleneck of the Finnish population. These include a missense variant in ZAP70, a gene in which recessive mutations cause immunodeficiency.

Using the results of this scan, as well as from a recent study of quantitative thyroid stimulating hormone (TSH) levels in the HUNT study and Michigan Genomics Initiative, we explore the impact of treated and untreated hypothyroidism on risk to all late-onset common diseases by utilizing polygenic risk scores (PRS) as a proxy for the severity of hypothyroidism in conjunction with the comprehensive national medication reimbursement data available nationwide since 1995 in Finland in the FinnGen cohort. With this approach, we can highlight specific instances where treatment for hypothyroid disease ameliorates downstream health consequences, reinforcing the importance of increasing the detection of untreated hypothyroid disease in the population

## PgmNr 2818: A new biomarker for idiopathic pulmonary fibrosis by combining Mendelian randomization and a metabolomics-based case-control study.

### Authors:

A. Cerani <sup>1,2</sup>; S. Ross <sup>1,2</sup>; D. Assayag <sup>2</sup>; D.A. Schwartz <sup>3</sup>; M. Lathrop <sup>4</sup>; P. Wolters <sup>5</sup>; J.B. Richards <sup>1,2,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada; 2) Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada; 3) Department of Medicine, University of California, San Francisco, CA, USA; 4) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 5) Department of Medicine, University of Colorado, Denver, CO, USA; 6) Department of Human Genetics / Department of Medicine, McGill University, Montreal, Canada

---

**Introduction:** Idiopathic pulmonary fibrosis (IPF) is a lethal disease with limited biomarkers and effective treatments. There is, therefore, an urgent need to find and validate biomarkers that could serve for diagnosis, prognosis, and/or as potential drug targets. We undertook combined genetic epidemiology and metabolomics studies to identify a metabolic biomarker for IPF.

**Methods and Results:** To identify potential metabolites associated with IPF, we looked for genetic markers jointly associated with blood metabolites levels and IPF risk using GWAS data for blood metabolites and IPF, respectively. Via this exploratory approach, we identified a shared genetic determinant near isovaleryl dehydrogenase (*IVD*) (SNP-metabolite association:  $-0.034$ ,  $p=1.87 \times 10^{-35}$ ; SNP-IPF risk association:  $OR=1.20$ , 95% CI: 1.12 to 1.28,  $p=7.4 \times 10^{-8}$ ). The known product of this gene is an enzyme whose inhibition is known to increase its substrate, the metabolite isovalerylcarnitine (IVC). Through two-sample Mendelian randomization (MR), we observed that genetically decreased IVC levels were strongly associated with increased risk of IPF ( $OR=12.13$ , 95% CI: 2.21 to 66.5,  $p=4 \times 10^{-3}$ ). We confirmed this result by using a different blood metabolomics GWAS in a second MR study ( $OR=2.91$ , 95% CI: 2.05 to 4.11,  $p=1.7 \times 10^{-9}$ ). We next found that the non-coding alleles from the two metabolite GWAS that were associated to decreased IVC blood levels increased *IVD* expression. Next we tested whether low IVC levels were associated with IPF risk in a study of 382 cases and 217 controls. We found that decreased blood IVC blood levels was associated with a 26% increase in IPF risk ( $OR: 1.26$  per 50 nM decrease, equivalent to an IQR drop in IVC; 95% CI: 1.00 to 1.60,  $p=0.048$ ). We are currently undertaking a replication case-control study.

**Conclusions:** IPF presents important unmet clinical needs. Building on strong complementary evidence from genetics and metabolomics data, we found that low IVC levels were associated with higher risk of IPF. IVC represents a novel biomarker and its enzyme, *IVD*, could represent an entirely novel IPF drug target.

## PgmNr 2819: Environment-dependent effect of *ALDH2* polymorphism on pediatric respiratory infections in Vietnam.

### Authors:

M. Yasunami <sup>1,2</sup>; R. Miyahara <sup>2</sup>; L.M. Yoshida <sup>2</sup>; H. Nakamura <sup>2</sup>; K. Takahashi <sup>2</sup>; D.T. Vu <sup>3</sup>; D.A. Dang <sup>3</sup>; H.T. Le <sup>4</sup>; H. Moriuchi <sup>5</sup>; K. Ariyoshi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Life Sciences Institute, Saga-ken Medical Centre Koseikan, Saga, Japan; 2) Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan; 3) National Institute of Hygiene and Epidemiology (NIHE), Hanoi, Vietnam; 4) Khanh Hoa Provincial Public Health Service, Nha Trang, Vietnam; 5) Graduate School of Medical Sciences, Nagasaki University, Nagasaki, Japan

---

**Background:** Aldehyde dehydrogenase 2 (ALDH2) plays a key role in the degradation of ingested ethanol as well as the detoxication of environmental aldehydes. A variant allele of *ALDH2* locus, *ALDH2\*2*, encodes an inactive subunit of the enzyme with dominant-negative properties which is prevalently observed in Asian populations. Accordingly, the carriers of *ALDH2\*2* show ethanol intolerance and susceptibility to various diseases including esophageal cancer which are believed to be caused by toxic substance otherwise eliminated by active ALDH2. Acetaldehyde in cigarette smoke can induce tissue damages even in passive smokers, which in turn may enhance the susceptibility to respiratory infectious agents. In the present study, the effects of environmental tobacco smoke (ETS) exposure from parental habitual smoking and *ALDH2\*2* carriage on the risk of pediatric respiratory infections were assessed by a prospective cohort study. **Methods:** 1999 newborn babies, who were born at a single institution in Nha Trang City, Khanh Hoa Province, Central Vietnam in one-year period from May 2009 to May 2010, were enrolled to a longitudinal observational study after obtaining the parental informed consent. The history of hospital admission and that of exposure to ETS at home were collected at the age of 24 months. *ALDH2* genotype was determined using umbilical cord blood DNA. The incidence of respiratory infections requiring hospitalization was compared between subpopulations with/without risk factors. **Results:** Among 1494 children who were followed up for 24 months (74.7% of enrolled), 231 hospital admissions because of acute lower respiratory tract infections (LRTIs) were detected (71.6 events/1000PYO). Exposure to ETS increase the chance to be hospitalized (un-adjusted RR = 1.49, P = 0.017) as reported previously (Miyahara R, *et al.* Sci Rep 7:45481(2017)). The genotype frequencies of *ALDH2* in the children followed-up were \*1/\*1 : \*1/\*2 : \*2/\*2 = 0.69 : 0.28 : 0.03, which were not deviated from Hardy-Weinberg equilibrium. Carriers of *ALDH2\*2* exhibited a tendency of increased LRTI hospitalization (p = 0.07). Stratified population analysis revealed the effect of *ALDH* genotype was augmented in the children with exposure to ETS (RR = 1.80, p = 0.004) but not in those without. **Conclusion:** *ALDH2\*2* enhance the risk of acute LRTIs conferred by the exposure to ETS at home in infants under this epidemiological setting.



## PgmNr 2820: Target identification in bronchiectasis disease using whole exome sequencing.

### Authors:

E.M. Schmidt<sup>1,2</sup>; W. Flowers<sup>2,3,4</sup>; K. Brown<sup>2,3,4</sup>; L. He<sup>1,2</sup>; K.H. Regan<sup>2,5</sup>; S.L. Spain<sup>1,2</sup>; R. McIntyre<sup>1,2</sup>; I. Dunham<sup>2,6</sup>; J.C. Barrett<sup>2,7</sup>; M. Begg<sup>2,8</sup>; A. Amour<sup>2,8</sup>; S. Nejentsev<sup>2,9</sup>; M. Ghoussaini<sup>1,2</sup>; A.M. Condliffe<sup>2,10</sup>; A.T. Hill<sup>2,5</sup>; E.M. Hessel<sup>2,8</sup>; D. Michalovich<sup>2,8</sup>; R.A. Floto<sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SA, UK; 2) Open Targets, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK; 3) Molecular Immunity Unit, Department of Medicine, University of Cambridge, Cambridge, CB2 0QH, UK; 4) Cambridge Centre for Lung Infection, Royal Papworth Hospital, Cambridge, CB2 0AY, UK; 5) Department of Respiratory Medicine, Royal Infirmary of Edinburgh, Edinburgh, EH16 4SA, UK; 6) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, UK; 7) Genomics plc, Cambridge, CB1 1BH, UK; 8) Refractory Respiratory Inflammation DPU, GlaxoSmithKline plc, Stevenage, Hertfordshire, SG1 2NY, UK; 9) Department of Medicine, University of Cambridge, Cambridge, CB2 0QQ, UK; 10) Department of Infection, Immunity and Cardiovascular Disease, School of Medicine, University of Sheffield, Sheffield, S10 2RX, UK

Bronchiectasis (BE) is a respiratory condition characterized by dilated bronchi, mucus hypersecretion, persistent lung inflammation, and chronic respiratory infection. Although BE can be caused by genetic disorders such as cystic fibrosis, primary ciliary dyskinesias, and immunodeficiencies (e.g. Activated PI3Kinase Delta Syndrome (APDS)), up to 50% of occurrences have no easily definable disease origin (idiopathic or post-infective). There are no licensed medicines specifically targeted for BE patients, presenting a need for deeper exploration into the genetic etiology of this disease and new therapeutic opportunities.

We performed whole exome sequencing of blood samples from 800 cases with idiopathic BE recruited from two respiratory clinics in the UK, and obtained exomes for 4,301 healthy donors from the UK Interval Cohort. We carried out single-variant and gene-based exome-wide association scans in a subset of unrelated individuals of European ancestry (743 cases and 3,992 controls) to estimate the aggregate burden of low frequency protein-altering variation in disease risk.

Of the 820,102 quality protein-coding sites identified, 2.6% are predicted high-confidence loss-of-function variants (stop-gained, splice site disrupting, frameshift), and 36.6% are predicted damaging missense variants (CADD>21 or MCAP>0.025). Single-variant association scans revealed exome-wide significant common variants (Firth logistic regression, MAF>1%) in two biologically plausible genes involved in epithelial cell function ( $P=7E-7$  and  $P=8E-7$ ), as well as rare variants (Fisher's Exact, MAF<1%) in two genes involved in adaptive immunity ( $P=5E-42$  and  $P=2E-23$ ). Notably, the combined burden of rare coding variants (SKAT-O, MAF<1%) resulted in exome-wide significant associations for three genes involved in epithelial barrier function ( $P=3E-32$ ), vesicular trafficking ( $P=4E-10$ ), and oxidative stress ( $P=3E-6$ ). The BE cohort is also composed of heterozygous carriers for 10 novel, potentially damaging missense variants in the *PIK3CD* and *PIK3R1* genes of the PI3Kdelta signalling pathway, as well as one carrier of the known APDS-causing *PIK3CD* E81K substitution (c.241G>A; p.Glu81Lys), subsequently confirmed by Sanger sequencing. Finally, we observe 30 heterozygous

carriers of at least 2 likely pathogenic coding variants in *CFTR*. These pathogenic variants in genes of immune and airway-related pathways identify novel mechanisms for disease susceptibility and guidance toward new drug targets.

## PgmNr 2821: Genome-wide gene-by-smoking interaction study of chronic obstructive pulmonary disease.

### Authors:

W. Kim<sup>1</sup>; D. Prokopenko<sup>2</sup>; P. Sakornsakolpat<sup>3</sup>; B.D. Hobbs<sup>4,5</sup>; M.H. Cho<sup>4,5</sup>; E.K. Silverman<sup>4,5</sup>; T.H. Beaty<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, MD; 2) Genetics and Aging Research Unit, Department of Neurology, Massachusetts General Hospital, Boston, MA; 3) Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand; 4) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 5) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA

---

**Introduction:** Risk for Chronic Obstructive Pulmonary Disease (COPD) is determined by both cigarette smoking and genetic susceptibility, but little is known about gene-by-smoking (GxSmoking) interaction. Our study aims to identify novel genetic variants by accounting for potential interaction with smoking on COPD risk. **Method:** We analyzed 200,766 subjects with European ancestry including 179,689 controls and 21,077 COPD cases based on pre-bronchodilator spirometry for moderate airflow limitation from the UK Biobank study. We sought to replicate the significant results in 2,533 control and 2,809 case non-Hispanic white former and current smokers from the COPD Gene cohort, examining the marginal genetic effects. We considered 3 binary smoking variables (exposed/non-exposed): ever/never, current/non-current, and heavy ( $\geq 10$  pack-years)/non-heavy smokers in a genome-wide analysis. We used logistic regression with a 2-degree-of-freedom (2df) test considering both genetic main effects and GxSmoking interaction effects. We examined variants achieving genome-wide significance ( $P < 5 \times 10^{-8}$ ) in our analysis but conditionally independent ( $P < 1 \times 10^{-5}$ ) from previously described GWAS variants. **Results:** We identified 3 variants - rs11735046 near *GYP A*, rs56326741 in *TN X B* and rs2604894 in *EGL N 2* that are distinct from previously reported COPD or lung function variants. At known loci of COPD, 3 variants at 2 loci - *CH R N B 4* and *M E C O M* showed a statistically significant interaction with smoking. For these 6 variants, 1 showed a stronger association among the non-exposed group, while 5 among the exposed group. Five human leukocyte antigen (HLA) imputed markers were statistically significant for 1df interaction with GxEver-smoking. In the COPD Gene cohort, one of the variants at the previously described 19q13 loci - rs2604894 showed nominal significance ( $P = 0.02$ ) in a marginal association with COPD. Two known loci - *CH R N B 4* and *M E C O M*, showed a stronger association with longer smoking duration. **Conclusion:** Despite the strong relationship between genetic variants and smoking in COPD, a genome-wide GxSmoking interaction study did not identify new genome-wide significant loci. Interaction effects and secondary signals at previously described loci, however, may help elucidate biologic mechanisms.

## PgmNr 2822: Genome-wide gene-smoking interaction analysis of lung function in UK Biobank.

### Authors:

C.A. Melbourne<sup>1</sup>; N. Shrine<sup>1</sup>; C. Batini<sup>1</sup>; I.P. Hall<sup>3,4</sup>; The SpiroMeta Consortium<sup>1</sup>; M.D. Tobin<sup>1,2</sup>; L.V. Wain<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Health Sciences, University of Leicester; 2) NIHR, Leicester Respiratory Biomedical Research Centre; 3) NIHR, Nottingham University Hospitals; 4) Division of Respiratory Medicine, University of Nottingham

---

Chronic obstructive pulmonary disease (COPD) encompassing chronic bronchitis and emphysema, is characterised by severe airflow obstruction and is a top contributor of worldwide mortality. Smoking is the biggest risk factor, whilst other environmental exposures also play a role such as air pollution, exposure to dust and previous chest infections. However, not all smokers develop COPD and around 20% of individuals with COPD have never smoked. Lung function measures used in COPD diagnosis have a strong genetic component, with 279 genetic signals identified to date. These signals are also associated with COPD risk. However, they only account for modest proportions of lung function heritability and have the same effects in ever and never-smokers. We hypothesised that there are additional genetic determinants of lung function and COPD that interact with cigarette smoking.

Presented here is the largest genome-wide gene-smoking interaction analysis in lung function to date, specifically for traits FEV<sub>1</sub> (forced expiratory volume in 1 second), FVC (forced vital capacity), the ratio of FEV<sub>1</sub>/FVC and PEF (peak expiratory flow). The analysis included 303,612 unrelated European individuals from UK Biobank. Phenotypes were inverse normalised adjusting for age, sex and height with 10 ancestry principal components and genotype array adjusted for during analysis. Interaction effect was determined using a gene-smoking interaction term in a linear regression model and a threshold of was used to identify discovery signals. Replication was sought using the SpiroMeta consortium consisting of 22 studies and up to 71,067 individuals.

Analysis of 8,647,748 variants identified 53 independent genetic signals at  $P < 5 \times 10^{-6}$ , none of which have previously been implicated for lung function or COPD. Twenty six signals were imputed with high quality in the SpiroMeta consortium (effective sample size  $> 50,000$ ), of which 11 had consistent direction of effect in ever and never-smokers. None reached a Bonferroni corrected threshold of for replication.

These 53 signals may influence estimates of relative and absolute genetic risk for poor lung function and COPD, and aid in the development of personalised medicine based on smoking behaviour. Larger sample sizes with denser imputation are however required to establish these signals as true positives.

## PgmNr 2823: Utilising human leukocyte antigen imputation to understand the role of genetic variation in immune system genes in a meta-analysis of idiopathic pulmonary fibrosis susceptibility.

### Authors:

M.L. Paynton <sup>1</sup>; R.J. Allen <sup>1</sup>; J. Oldham <sup>2</sup>; S. Ma <sup>3</sup>; T. Fingerlin <sup>4,5</sup>; R. Braybrooke <sup>6,7</sup>; I. Sayers <sup>7,8</sup>; I.P. Hall <sup>7,8</sup>; M.D. Tobin <sup>1,9</sup>; T. Maher <sup>10,11</sup>; P. Molyneaux <sup>10,11</sup>; D. Schwartz <sup>4,12,13</sup>; C. Flores <sup>14,15,16,17</sup>; I. Noth <sup>4</sup>; G. Jenkins <sup>6</sup>; E.J. Hollox <sup>18</sup>; L.V. Wain <sup>1,9</sup>; UK ILD Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Health Sciences, University of Leicester, Leicester, Leicestershire, United Kingdom; 2) Department of Internal Medicine, University of California Davis, Davis, United States of America; 3) Division of Pulmonary & Critical Care Medicine, University of Virginia, Charlottesville, United States of America; 4) Center for Genes, Environment and Health, National Jewish Health, Denver, United States of America; 5) Department of Biostatistics and Informatics, University of Colorado, Denver, United States of America; 6) Division of Epidemiology and Public Health, University of Nottingham, Nottingham, United Kingdom; 7) National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham University Hospitals, Nottingham, United Kingdom; 8) Division of Respiratory Medicine, University of Nottingham, Nottingham, United Kingdom; 9) National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Glenfield Hospital, Leicester, United Kingdom; 10) NIHR Respiratory Clinical Research Facility, Royal Brompton Hospital, London, United Kingdom; 11) National Heart and Lung Institute, Imperial College, London, United Kingdom; 12) Department of Medicine, University of Colorado Denver, Denver, United States of America; 13) Department of Immunology, University of Colorado Denver, Denver, United States of America; 14) Research Unit, Hospital Universitario N.S. de Candelaria, Universidad de La Laguna, Santa Cruz de Tenerife, Spain; 15) CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Spain; 16) Instituto Tecnológico y de Energías Renovables (ITER, S.A.), Santa Cruz de Tenerife, Spain; 17) Instituto de Tecnologías Biomédicas (ITB), Universidad de La Laguna, Santa Cruz de Tenerife, Spain; 18) Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom.

---

Idiopathic pulmonary fibrosis (IPF), a rare interstitial lung disease of an uncertain cause with poor prognosis, is believed to be the result of an abnormal wound healing response to injury leading to large areas of scarring and inflammation of the alveolar wall. Genome wide association studies (GWAS) have reported 17 signals of association with IPF susceptibility from across the genome.

The human leukocyte antigen (HLA) region is one of the most complex and variable regions in the genome and is known to influence risk for multiple diseases (including respiratory disorders such as asthma and chronic obstructive pulmonary disease). The extreme genetic diversity of the HLA, with each HLA gene potentially encoding thousands of different proteins, challenges the interpretation of association studies in the region. Specific imputation strategies have been designed to capture classical HLA gene alleles and amino acid changes. Association of the classic *HLA-DQB1\*06:02* allele has been previously reported in a cohort of fibrotic idiopathic interstitial pneumonias (including IPF).

In this study, we combined the latest single variant imputation panel (Haplotype Reference Consortium) and a bespoke HLA classical allele and amino acid imputation panel (T1DGC). We tested the association of 37,212 SNPs, 424 HLA alleles and 1,276 amino acid changes across the HLA region (including *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DRB1*) with IPF susceptibility in 2,769 IPF cases and 8,591 controls from 3 independent studies (UK [612 cases, 3,366 controls], Chicago [500 cases, 510 controls] and Colorado [1,515 cases, 4,683 controls]). Sex and 10 Principal Components were included as covariates.

Three independent signals passed a Bonferroni-corrected significance threshold of  $P < 2.8 \times 10^{-6}$  in a meta-analysis of the 3 studies. However, these three associations were not observed (at  $P < 0.05$ ) across all three contributing studies. The previously reported association of *HLA-DQB1\*06:02* was not replicated in this study.

To the best of our knowledge, we have undertaken the largest study of the role of HLA gene variation in IPF susceptibility to date, using a combination of dense SNP imputation and bespoke HLA allele imputation. No variants were consistently associated with IPF susceptibility in our data.

## PgmNr 2824: Comparisons between whole exome sequencing and whole genome sequencing in 25 individuals with autosomal recessive non-syndromic deafness.

### Authors:

S. Guo <sup>1</sup>; B. Guney <sup>1</sup>; J. Huang <sup>1</sup>; S. Blanton <sup>1,2</sup>; M. Tekin <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA

We have compared output of whole-exome sequencing (WES) and whole-genome sequencing (WGS) in 25 unrelated probands with autosomal recessive non-syndromic deafness. Agilent SureSelect Human All Exon 50 Mb kits and an Illumina HiSeq2000 instrument were utilized for WES; we used Illumina HiSeq X Ten or BGISEQ-500 platforms for WGS. Sequencing data were analyzed at our in house bioinformatics pipeline. We then compared the coverage in the regions targeted by WES capture

We subsequently compared the coverage of 16 autosomal recessive non-syndromic deafness genes *CDH23*, *CLDN14*, *GJB2*, *GRAP*, *ILDR1*, *LOXHD1*, *MPZL2*, *MYO15A*, *MYO7A*, *OTOA*, *OTOF*, *PCDH15*, *SLC26A4*, *TMC1*, *TMPRSS3*, and *USH1C*. These genes were selected as mutations in them were found to be causative in multiple families in our previous studies. Out of the 25 samples, we identified 7 causative variants via WGS as opposed to 4 causative variants via WES. All 7 variants were confirmed with Sanger sequencing. While WES and WGS appear to have similar coverage for deafness genes, WGS appears to be superior as it provides more uniform coverage to detect deafness mutations scattered throughout the coding sequence. In addition, WGS provides data for non-protein coding regions, structural variants, and potentially better detection of copy number variants.

	mean	1X	5X	10X	20X	SNV	Indels
WES	69	99%	95%	88%	77%	134119	9790
WGS	45	99.90%	99.70%	99.20%	98.80%	119911	21641
WES_DeafGenes	97	99.70%	98.40%	96.80%	92.40%	160	8
WGS_DeafGenes	47.5	100%	99.90%	99.60%	97.60%	147	12

## PgmNr 2825: Genome-wide analyses using UK Biobank data provide insights into the architecture of chronic kidney disease and kidney function.

### Authors:

Y. Lin <sup>1,2</sup>; J. Huang <sup>2</sup>; G. Liao <sup>3</sup>; L. Du <sup>4</sup>; Y. Guo <sup>1</sup>; W. Chung <sup>1</sup>; J. Li <sup>1</sup>; J. Si <sup>1</sup>; L. Liang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA.; 2) Department of Urology, West China Hospital, Sichuan University, Chengdu, P.R.China; 3) State Key Laboratory of Oral Diseases, National Clinical Research Center for Oral Diseases, Chinese Academy of Medical Sciences Research Unit of Oral Carcinogenesis and Management, West China Hospital of Stomatology, Sichuan University, Chengdu, P.R.; 4) Chinese Evidence-Based Medicine Center, West China Hospital, Sichuan University, Chengdu, P.R.China

Chronic kidney disease (CKD), affects ~11% of the adult population in Europe and North America. It is associated with high morbidity and in the advanced stage requires life supporting treatment by renal dialysis or transplantation. CKD is defined as an irreversible reduction in glomerular filtration rate (GFR), which can be estimated by both serum creatinine (eGFR<sub>crea</sub>) and serum cystatin C (eGFR<sub>cys</sub>). Previous largest genome-wide association studies for these serum biomarkers were limited to ~ 110 thousands individuals. Thus, we aimed to perform genome-wide association studies for CKD-related quantitative biomarker traits both from serum and urine, using the UK Biobank data collected on more than 400,000 individuals.

We used the linear mixed model approach as implemented in the software BOLT-LMM to account for cryptic relatedness and potential population stratification. We inverse normalized all the biomarker measures, then took the residuals using covariates, including sex, age, age squared, assessment center locations, genome arrays and 10 principal components. We analyzed on serum levels of creatinine (N=437,653) and cystatin C (N=437,839) and urinary levels of sodium (N=445,081) in white individuals from the UK Biobank. Then we used FUMA, a web-based platform, to perform functional mapping and annotation for the genetic components associated with each genetic trait.

Our genome-wide association study demonstrated strong associations with CKD-related biomarkers. For example, we identified 93,501 and 87,815 significant SNP associations ( $P < 5 \times 10^{-8}$ ) with eGFR<sub>crea</sub> and eGFR<sub>cys</sub> at serum level, respectively. The top locus for eGFR<sub>crea</sub> included *SPATA5L1* locus (rs2433601,  $P=7.4 \times 10^{-246}$ ), which encoded a protein suspected to be related to creatinine production. The top SNP associated to eGFR<sub>cys</sub> was identified at *SHROOM3* locus (rs4859682,  $P=1.95 \times 10^{-296}$ ) which encoded a protein may be involved in cell development in kidney. For urinary sodium, we identified 4,659 significant SNP associations ( $P < 5 \times 10^{-8}$ ). The top locus was found at *FGF21* locus (rs838133,  $P=1.2 \times 10^{-28}$ ), which encoded fibroblast growth factor 21 and was noted to correlate with the severity of albuminuria and loss of GFR.

In summary, we identified common variants associated with kidney function in large population-based studies in both serum and urinary level. We further characterized the shared genetic architecture of these serum and urine biomarkers with CKD using 11,000 cases in UK Biobank cohort.



## PgmNr 2826: Genome-wide association study of autoimmune thyroid diseases detects nine novel risk loci.

### Authors:

T.J. Meyers<sup>1</sup>; S.R. Rashkin<sup>1</sup>; J.S. Witte<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology and Biostatistics, UCSF School of Medicine, San Francisco, California.; 2) Institute for Human Genetics, UCSF School of Medicine, San Francisco, California.

**Introduction:** Autoimmune thyroid diseases (AITD) include Hashimoto's thyroiditis (HT) and Graves' disease (GD). The prevalence of HT and GD in the United States is about 5% and 0.5%, respectively. Both of these phenotypes are associated with papillary thyroid cancer in epidemiological studies. Genome-wide association studies (GWAS) have reported 32 genetic risk variants for HT and/or GD, with only one variant in HT alone.

**Methods:** A GWAS was performed in the UK Biobank (UKB) on 257 cases of HT and 541 cases of GD, with population-based controls matched 8 to 1. After imputation and quality control steps, 8.9 million autosomal single nucleotide polymorphisms (SNPs) with minor allele frequency greater than 1% were tested in logistic regression models for linear additive effects for each phenotype. SNP associations with p-value less than  $5 \times 10^{-8}$  were pruned by linkage disequilibrium (LD) to retain independent loci. SNPs were considered novel if they were not in LD with previously reported genome-wide significant SNPs ( $r^2 < 0.1$  within a 500-kb window).

**Results:** A locus on *PDXK* (chromosome 21) was associated with HT (lead SNP rs181726789; p-value =  $4.3 \times 10^{-8}$ ; odds ratio = 2.99, 95% confidence limits=2.02,4.42). This is a missense variant on a gene involved in vitamin B metabolism and which has not been associated with HT before. For GD, eight independent and novel loci were detected, seven located on chromosome 6. The lead SNPs for these seven loci were rs2523432, rs3132390, rs35744819, rs3130288, rs3130612, rs3094084, and rs2596474, of which the latter four are within the major histocompatibility complex (MHC). Another novel risk locus for GD was detected on chromosome 18 (lead SNP rs4239469 in an intronic region of *LINC01929*).

**Discussion:** This study contributes a known risk locus for HT and eight for GD. The analysis will be repeated in the full UK Biobank cohort (n=413,870 people).

## PgmNr 2827: High-sensitivity C-reactive protein and estimated glomerular filtration rate: A two-sample Mendelian randomization.

### Authors:

R. Fujii <sup>1,2</sup>; A. Hishida <sup>3</sup>; T. Nishiyama <sup>4</sup>; M. Nakatochi <sup>5</sup>; T. Kondo <sup>2</sup>; K. Wakai <sup>3</sup>; J-MICC Study Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Preventive Medical Sciences, Fujita Health University, Toyoake, Aichi, Japan; 2) Department of Pathophysiological Laboratory Sciences, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan; 3) Department of Preventive Medicine, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan; 4) Department of Public Health, Nagoya City University Graduate School of Medical Sciences, Nagoya, Aichi, Japan; 5) Department of Nursing, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan

Chronic inflammation is thought to be a risk factor for kidney disease. However, the plentiful discussion of this topic was controversial whether inflammatory status is either a trigger or results of chronic kidney disease. We aimed to investigate the causal relationship between high-sensitivity C-reactive protein (hs-CRP) and kidney function using two-sample mendelian randomization (MR). We conducted a MR study to estimate the effect of inflammation on kidney function as a part of the Japan multi-institutional collaborative cohort (J-MICC) Study, one of the largest genome cohort studies in Japan. hs-CRP concentrations were measured using a latex-enhanced immunonephelometric assay. Estimated glomerular filtration rate (eGFR) was calculated based on serum creatinine, age, and sex using the Japanese equation. Due to the limited sample for measured hs-CRP ( $n = 2,524$ ), we created a non-overlapping dataset from the J-MICC Study ( $n = 7,186$ ), and applied two-sample MR method, an application of MR to the summary statistics in non-overlapping datasets. Four independent single nucleotide polymorphisms (SNPs) known to be associated with CRP (rs3093077, rs1205, rs1130864, and rs1800947) were used to construct the instrumental variables (IVs) based on previous GWAS in European populations. F-statistics of IVs was 10.5, which indicates the causal estimates were unlikely to be biased. In addition to the inverse-variance weighted (IVW) method, we applied the weighted median (WM) method and the MR-Egger method to estimate causal relationship even when some genetic variants violate the important assumptions.

The mean age (SD) in CRP and eGFR datasets were 54.9 (9.3) and 53.7 (9.5), respectively. In linear regression adjusted for age, sex, cohort and top five principal components, natural log-transformed eGFR was not significantly associated with increase in hs-CRP in all three approaches (IVW: beta per SD = 0.004;  $p = 0.74$ , WM: beta per SD = 0.003;  $p = 0.83$ , and MR-Egger: beta per SD = 0.005;  $p = 0.89$ ). In MR-Egger, the estimate of intercept was zero (intercept = 0.000;  $p = 0.97$ ), which indicates that all IVs and the estimates from IVW were valid and not biased. In conclusion, causal inference approaches using MR does not support a causal effect of inflammatory status on eGFR level in this population.

## PgmNr 2828: Circulating MiRNAs were differentially expressed in T2D patients with fatty liver disease.

### Authors:

T. Qiuxiang; Y. Jianli; W. Jianping; H. Chang; P. Xingchen; C. Peng

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics, College of Basic Medical Sciences, Jilin University, Changchun, China 130021

**Background:** Fatty liver disease (FLD) is a diabetes complication which can develop into liver fibrosis or liver cancer. The incidence of FLD in type 2 diabetes (T2D) patients was much higher than the general population. However, the effective biomarker for early prediction of fatty liver disease risk in T2D patients with acceptable accuracy is not available.

**Aims:** In this study, we intended to investigate the dysregulated circulating miRNAs in relation to FLD of T2D patients, in the hope of constructing an effective prediction model for FLD in these vulnerable people.

**Methods:** We conducted a two-stage case-control study, using peripheral blood serum samples from 33 FLD and 9 non-FLD T2D patients with retinopathy as discovery stage, and 42 FLD and 28 non-FLD T2D patients without retinopathy as replication stage. Differential abundance of circulating miRNAs was measured using next-generation sequencing. The participants in this study were of Chinese Han ethnicity. Moreover, the fragments per million (FPM) values were used to construct a FLD risk prediction model. Target genes of the dysregulated circulating miRNAs were curated and subjected to a tissue-specific expression analysis using FUMA.

**Results:** The differential expression in of 21 mature miRNAs were identified in our discovery cohort ( $p \leq 0.01$ ,  $LFC \geq 2$  or  $LFC \leq -2$ ), however, none of them remained significant after multiple testing correction ( $FDR\ p > 0.05$ ). Among these, miRNAs miR-4646-5p was also significant in our replication stage ( $\log_2$  fold change = -1.71,  $p = 5.7 \times 10^{-4}$ ). We employed miR-4646-5p as the only biomarker to construct a FLD risk prediction model and assessed its predictive power in our replication cohort. The result showed that our risk prediction model was of marginal clinical relevance ( $AUC = 0.788$ , specificity = 1.000, sensitivity = 0.485). In addition, the target genes of miR-4646-5p were significantly abundant liver ( $p = 1 \times 10^{-7}$ ), heart ( $p = 1 \times 10^{-5}$ ) and pancreas ( $p = 1 \times 10^{-4}$ ).

**Conclusion:** Circulating miRNA miR-4646-5p was differentially expressed in serum samples of T2D patients with FLD as compared to non-FLD. This miRNA may be indicating novel molecular mechanisms of FLD onset in T2D patients, hence could potentially serve as biomarkers for FLD risk prediction when T2D diagnosis is made. Our prediction model was not clinically actionable, but could be refined in future studies with a larger sample size.

## PgmNr 2829: The association between genetic risk score for age at menarche or menopause and women's reproductive health-related traits in the Study of Women's Health Across the Nation.

### Authors:

W. Zhao <sup>1</sup>; J.A. Smith <sup>1</sup>; L.F. Bielak <sup>1</sup>; E. Ruiz-Narvaez <sup>2</sup>; M. Yu <sup>1</sup>; M. Hood <sup>1</sup>; P.A. Peyser <sup>1</sup>; S.L.R. Kardia <sup>1</sup>; S.D. Harlow <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, 48109; 2) Department of Nutritional Sciences, School of Public Health, University of Michigan, Ann Arbor, Michigan, 48109

---

Genome wide association studies (GWAS) of age at menarche and age at menopause have identified hundreds of genetic risk loci associated with reproductive timing. In order to better understand the roles of genetic loci contributing to women's reproductive health, we constructed standardized genetic risk scores (GRS) for age at menarche and age at menopause using SNPs at least suggestively associated ( $p < 1e-04$ ) with each respective trait in the largest GWASs to date. We then evaluated their association with eight traits related to women's reproductive health in an independent multi-ethnic cohort, the Study of Women's health Across the Nation (SWAN). White (N=726), Black (N=348), Chinese (N=139) and Japanese (N=140) participants were genotyped using the Illumina Multi-Ethnic Global Array, and genotypes were imputed to the 1000G Phase 3 v5 reference panel. The traits we evaluated included age at menopause, age at menarche, reproductive lifespan, duration of perimenopause and reproductive hormone levels (estradiol (E2) and follicle stimulating hormone (FSH)) before and after menopause. We observed a significant association between the menarche GRS and age at menarche, and the menopause GRS and age at menopause in almost all ethnicities with the strongest associations observed in Whites (beta=0.41 year increase with each SD of GRS,  $p=6.9e-15$  for menarche; beta=0.56,  $p=4.95e-05$  for menopause). In Whites alone, we observed a significant association between reproductive lifespan and both the menarche (beta=-0.54,  $p=2.0e-04$ ) and menopause GRSs (beta=0.76,  $p=2.9e-07$ ), the duration of perimenopause and the menopause GRS (beta=0.58,  $p=1.9e-4$ ), and FSH level and the menopause GRS (beta=-0.15 IU/L increase with each SD of GRS,  $p=0.01$ ). In conclusion, this study validates the established roles of these genetic risk loci in age at menarche and menopause in populations of multiple ethnicities and suggests additional roles in other reproductive health related traits. However, there was heterogeneity among ethnicities regarding the effect sizes and the observed significant effects were most prominent in Whites. This may be due to lack of power (small sample size) in other ethnicities or because the effect sizes of the genetic loci were estimated through GWAS conducted in European ancestry samples. This suggests an urgent need for large scale GWAS on reproductive health-related traits in non-European populations to better understand these genetic effects in other ethnicities.

## PgmNr 2830: Increased burden of common risk alleles in children with a significant fracture history.

### Authors:

T. Nakanishi<sup>1,2</sup>; D. Manousaki<sup>1,2</sup>; A. Kampe<sup>3</sup>; R. Makitie<sup>3,4</sup>; G. Bardai<sup>6</sup>; A. Belisle<sup>5</sup>; R. Li<sup>5</sup>; V. Forgetta<sup>1</sup>; O. Makitie<sup>3,4</sup>; F. Rauch<sup>6</sup>; B. Richards<sup>1,2,7,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Canada; 2) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 3) Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 4) Division of Endocrinology, Children's Hospital, Helsinki University Hospital and Research Program for Clinical and Molecular Metabolism, University of Helsinki, Helsinki, Finland; 5) McGill Genome Center, McGill University, Montreal, Canada; 6) McGill University, Ingram School of Nursing, and Shriners Hospitals for Children, Montreal, Canada; 7) Department of Medicine, Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Canada; 8) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK

---

**Background:** Extreme presentations of common disease in children are often presumed to be of Mendelian etiology. However, the contribution of common risk alleles to presumed Mendelian presentations has not been fully explored. We therefore tested whether children with a significant fracture history but no evidence of osteogenesis imperfecta (OI) are at increased polygenic risk for fracture as measured through a polygenic risk score.

**Methods:** A childhood significant fracture history was defined as the presence of low-trauma vertebral fractures or multiple long bone fractures. We generated a polygenic risk score for quantitative ultrasound speed of sound, termed "gSOS", derived from common risk alleles in 341,449 UK Biobank participants. Low gSOS predicts adult osteoporotic fracture. We tested if individuals from three cohorts with a significant childhood fracture history and no OI had lower predicted gSOS. The Canadian cohort included 94 children with suspected Mendelian osteoporosis. Of these, 26 had a positive and 68 a negative OI gene panel. Two Finnish cohorts included 59 children recruited due to a significant fracture history and 25 children referred for suspected Mendelian osteoporosis, among which 5 had positive OI gene panel. 11 individuals of non-European ancestry and individuals with OI, or who failed genotyping were excluded. We computed gSOS estimates in the remaining individuals, standardized them, and compared their mean to that of a UK Biobank subset. We computed the number of individuals with significant fracture history and identified OI and compared it to that of children having gSOS below the mean in excess of what would be expected by chance alone.

**Results:** The average gSOS across all three cohorts (n= 131) was -0.47 standard deviations lower than that in UK Biobank (n=80,027,  $P= 1.1 \times 10^{-5}$ ). The gSOS of 78 individuals suspected to have Mendelian osteoporosis was even lower (-0.76 standard deviations,  $P$  for comparison with UK Biobank  $=5.3 \times 10^{-10}$ ). Among 108 individuals with fractures tested for OI, while 30 had a mutation in an OI-related gene, there were 23 more subjects with a gSOS below the mean than would be expected by chance alone.

**Conclusions:** Individuals with a significant childhood fracture history but not OI have an increased burden of common fracture risk alleles, compared to the general population. This suggests that patients with presumed Mendelian osteoporosis should be investigated for a polygenic etiology of fracture.

## PgmNr 2831: Interactions of genetic variants with physical activity associated with bone mineral density in Hispanic children: The Viva La Familia Study.

### Authors:

R. Hou <sup>1</sup>; S.A. Cole <sup>2</sup>; M. Graff <sup>3</sup>; Y. Wang <sup>3</sup>; K. Haack <sup>2</sup>; S. Laston <sup>4</sup>; N.R. Mehta <sup>5</sup>; R.J. Shypailo <sup>5</sup>; M.L. Gourlay <sup>6</sup>; A.G. Comuzzie <sup>7</sup>; K.E. North <sup>3</sup>; N.F. Butte <sup>5</sup>; V.S. Voruganti <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Nutrition, University of North Carolina at Chapel Hill, Charlotte, North Carolina.; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) South Texas Diabetes and Obesity Institute and Department of Human Genetics, University of Texas of the Rio Grande Valley, Brownsville, TX; 5) USDA/ARS Children's Nutrition Research Center, Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Department of Family Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC; 7) The Obesity Society, Silver Spring, MD

**Background:** Osteoporosis is a complex disease that is affected by both genetic and environmental factors. Physical activity is an important environmental factor affecting accumulation and maintenance of bone mass, particularly in children. However, it is unknown if the benefits of moderate to vigorous physical activity (MVPA) on bone mineral density (BMD) vary based on the genetic risk in Hispanic children.

**Objective:** Our aim was to investigate if MVPA interacts with BMD-related single nucleotide polymorphisms (SNPs) to influence BMD in Hispanic children (n=750, 4-19y) of the cross-sectional Viva La Familia Study.

**Methods:** Physical activity was measured by accelerometers for 3 consecutive days and MVPA was defined by physical activity ratio greater than 3.0 or energy expenditure greater than 0.04 kcal/kg per minute. Total body BMD and lumbar spine BMD were measured by dual energy X-ray absorptiometry. A genetic risk score (GRS) was computed based on significant SNPs associated with BMD in our study and from the published literature. The reference alleles were coded as 2 for homozygous alleles associated with greater risk of low BMD, 1 for heterozygous alleles and 0 for homozygous for alternate alleles.

**Results:** The children's mean age and BMI were  $11.0 \pm 4.1$  y and  $25.1 \pm 7.6$  kg/m<sup>2</sup>, respectively. We found evidence of statistically significant effects of interaction between the GRS and MVPA on total body BMD and lumbar spine BMD ( $p < 0.05$ ). Higher GRS was associated with a greater decrease in total body BMD in low MVPA group ( $\beta = -0.040$ ,  $p = 1.1 \times 10^{-5}$ ), as compared to high MVPA group ( $\beta = -0.015$ ,  $p = 0.02$ ). Higher GRS was also associated with a greater decrease in lumbar spine BMD in low MVPA group ( $\beta = -0.042$ ,  $p = 0.0016$ ), as compared to high MVPA group ( $\beta = 0.008$ ,  $p = 0.4$ ). Further analysis with individual SNPs included in the GRS showed evidence of statistically significant interactions between rs370055571 (*MAP4K3*), rs452369 (*SLC8A1*) and rs10953178 (*SEM1*), and MVPA in relation to total body BMD and lumbar spine BMD ( $p < 0.05$ ).

**Conclusions:** MVPA seems to benefit bone health in Hispanic children who are genetically susceptible to low BMD compared to children who are less prone to low BMD. Future independent replication studies of our findings and functional studies for biological mechanisms are warranted.



## PgmNr 2832: Effect of genetically raised serum calcium on bone mineral density and risk of fracture in the general population: A Mendelian randomization study.

### Authors:

B. Richards<sup>1,2,3,4</sup>; A. Cerani<sup>2,3</sup>; S. Zhou<sup>2,3</sup>; V. Forgetta<sup>2</sup>; J.A. Morris<sup>3</sup>; K. Trajanoska<sup>5,6</sup>; F. Rivadeneira<sup>5,6</sup>; S.C. Larsson<sup>7</sup>; K. Michaelsson<sup>8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine, McGill University, Montreal, Quebec, Canada; 2) Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 3) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec, Canada; 4) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 5) Department of Internal Medicine, Erasmus MC, university Medical Center, Rotterdam, Netherlands; 6) Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, Netherlands; 7) Unit of Nutritional Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 8) Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

---

**Importance.** Calcium supplementation in the general population is common and often intended to reduce risk of fracture. Yet, calcium supplementation has been associated with increased risk of coronary artery disease and protective effects on bone health are still unclear.

**Objective.** We therefore tested whether genetically increased serum calcium was associated with improved bone mineral density and a reduction in osteoporotic fractures.

**Methods.** We used a two-sample Mendelian randomization design. A genome-wide association meta-analysis of serum calcium levels in up to 61,079 individuals was used to identify genetic determinants of serum calcium levels. The UK Biobank study was used to assess the association of genetic predisposition to increased serum calcium with estimated bone mineral density (eBMD) derived from heel ultrasound in 426,824 individuals who had on average calcium levels in the normal range. We next performed a fracture genome-wide association meta-analysis comprising 24 cohorts and UK Biobank including a total of 76,549 cases and 470,164 controls, who on average also had calcium levels in the normal range.

**Results.** A standard deviation increase in genetically-derived serum calcium (0.13 mmol/L or 0.51 mg/dL) was not associated with increased eBMD (0.003 g/cm<sup>2</sup>, 95% CI: -0.059 to 0.066; P = 0.92) or reduced risk of fractures (OR = 1.01, 95% CI: 0.89 to 1.15, P = 0.85) in inverse-variance weighted Mendelian randomization analyses. Sensitivity analyses did not provide evidence of pleiotropic effects.

**Conclusions.** Genetic predisposition to increased serum calcium levels in individuals with normal calcium levels is not associated with an increase in eBMD and does not provide clinically-relevant protection against fracture. Whether such predisposition mimics the effect of short-term calcium supplementation is not known. Given that the same genetically-derived increase in serum calcium is



associated with increased risk of coronary artery disease, widespread calcium supplementation in the general population may provide more risk than benefit.

## PgmNr 2833: Multi-omics integration of the human gut microbiome and metabolome with applications to osteoporosis.

### Authors:

J. Greenbaum <sup>1</sup>; K.J. Su <sup>1</sup>; X. Lin <sup>2</sup>; R. Gong <sup>2</sup>; H. Shen <sup>1</sup>; J. Shen <sup>2</sup>; H.M. Xiao <sup>2</sup>; H.W. Deng <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Bioinformatics and Gemoics, Department of Global Biostatistics and Data Science, Tulane University, New Orleans, LA, USA; 2) Department of Endocrinology and Metabolism, The Third Affiliated Hospital of Southern Medical University, Guangzhou, China; 3) School of Basic Medical Sciences, Central South University, Changsha, 410000, China

---

While a number of genetic factors have been implicated in the pathogenesis of osteoporosis, the influences of the gut microbiome and human metabolome remain largely unexplored. In this study we performed multi-omics integration analysis using metagenomic whole genome shotgun sequencing of stool samples, LC-MS untargeted metabolomics profiling, and DXA-derived bone mineral density (BMD) collected from 499 Chinese postmenopausal women. First, we developed a sparse canonical correlation analysis (sCCA) model using a lasso penalty to select the most highly correlated features between the multi-omics data that are also important for BMD. The performance of the model was evaluated through simulation, which revealed that the sCCA method is able to detect >95% of the correlated features while also maintaining a reasonable false positive rate under a variety of different scenarios. Next, the sCCA model was applied to the real data and the selected features were used as input to construct the Gaussian graphical model, where the significance of network edges was determined by the graph lasso penalty. We interrogated the partial correlation network to determine the hub microbes and metabolites with the most inter- and intra-omic connections. There were 20 features (16 microbes and 4 metabolites) with significant connections to BMD after adjustment for all other nodes in the graph. Finally, using structure learning algorithms we estimated the Bayesian network to identify the potential driver microbes and metabolites that may play key roles in the regulation of BMD. The findings provide novel insight into the biological mechanisms that may be involved in the development of osteoporosis and related phenotypes.

## PgmNr 2834: GWAS of copy number variations (CNVs) for cleft lip with or without cleft palate (CL/P) in a multi-ethnic study sample.

### Authors:

N. Mukhopadhyay <sup>1</sup>; M. Lee <sup>1</sup>; E. Feingold <sup>2,3</sup>; M.L. Marazita <sup>1,2,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oral Biology, University of Pittsburgh/School of Dental Medicine, Pittsburgh, Pennsylvania.; 2) Human Genetics, University of Pittsburgh/Graduate School of Public Health, Pittsburgh, Pennsylvania; 3) Biostatistics, University of Pittsburgh/Graduate School of Public Health, Pittsburgh, Pennsylvania; 4) Clinical and Translational Science Institute, University of Pittsburgh/School of Medicine, Pittsburgh, Pennsylvania; 5) Psychiatry, University of Pittsburgh/School of Medicine, Pittsburgh, Pennsylvania

---

Orofacial clefts (OFCs) are the most common craniofacial birth defect in humans with a prevalence of 1 in 700 live births worldwide. Approximately 70% of cases with cleft lip with or without cleft palate (CL/P) are considered to be non-syndromic. Non-syndromic CL/P has been shown to have a strong genetic component and a number of risk variants across the genome have been identified by previous studies. The CL/P risk loci identified by GWASs thus far focus mainly on common single nucleotide variants. In our study, we investigate the contribution of copy number variations (CNVs) to the risk of developing OFCs in a large multiethnic study consisting of 2,863 European subjects, 4,513 Latino subjects and 2,000 Asian subjects. CNVs were identified on chromosomes 1-22 using the PennCNV software on a ~600K SNP genotyping array. Thereafter, we performed whole-genome association of CL/P using the CNV status (present/absent) of individuals at SNP positions that overlapped with one or more CNVs. In the Latino subgroup, significant associations were observed in 7p14.1, 16q21 and 19p13.3 in the genes *TARP*, *POLR2C* and *POLR2E* respectively. In the Asian subgroup, genome-wide significant associations were observed in 4p16.3, 11p15 and chromosome 21q22.3 spanning genes *TACC3*, *MRGPRG*, and *COL6A1* respectively. Further, the European subgroup showed a genome-wide suggestive association at 19q13.3 near the *ITPKC* gene. None of our significant loci replicate known OFC genes. However, two independent studies of OFCs have previously reported the involvement of CNVs near 7p14.1 (PMID 24528994) and 4p16.3 (PMID 26561393) respectively. Further, the region corresponding to the suggestive association in the European sample has been reported by linkage and association studies of CL/P (PMIDs 8751880, 9003487, 9722939). These previous studies were conducted on samples different from ours. The significantly associated SNPs are near or within active transcription sites and/or super-enhancer regions involved in craniofacial development. Our study shows promising evidence that structural variation within regulatory regions of craniofacial development may contribute to the formation of OFCs. This research is supported by NIH grants R01-DE016148, R03-DE026469 and X01-HG007485.

## PgmNr 2835: Using external controls to account for mating asymmetry in maternal genetic association: Application to orofacial cleft.

### Authors:

J.E. Ewusie <sup>1</sup>; K. Burkett <sup>2</sup>; M.-H. Roy-Gagnon <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada; 2) Department of Mathematics and Statistics University of Ottawa, Ottawa, ON, Canada

---

Mating symmetry is a frequent assumption made when using case-parent triad data to study the effects of maternal and child genes on the risk of a disease with early onset. Mating symmetry indicates that for any possible parental genotype pair, the frequency of a given mother-father genotype assignment in the population is equal to the reverse father-mother genotype assignment. The violation of this assumption (known as mating asymmetry) may result in confounding, leading to spurious maternal associations in studies utilizing case-parent triads. The hybrid design proposed by Weinberg et al., relies on the use of control-parent dyads to test and account for mating asymmetry (MA). However, parents of unrelated controls are not genotyped in most ongoing studies. In this study, we modified the hybrid design (HD) approach by using control parents from an external dataset. We first investigated the performance of the modified hybrid design through simulations, by considering different levels of MA and sample sizes ( $n$ ) of external control-parent dyads, to assess the effect on the type 1 error and power to identify maternal effects. Results from our simulation study showed that the type 1 error rate was around 5% when the MA in the external control-parent dyads was no more than  $\pm 0.15$  points from the MA in the case-parent triads. Power of the hybrid design using external controls was above 80% for most scenarios considered. Using this modified approach, we investigated how adjusting for MA with external control-parent dyads from the HapMap project would change previous maternal genetic associations observed for orofacial clefts using GWAS data on over 2000 case-parent triads from an international consortium accessed through dbGAP. Our findings show that, in some situations, the HD with external controls may account for MA and provide valid tests for maternally contributed genotype effects when the mating symmetry assumption fails.

## PgmNr 2836: A new summary-data-based statistical test for discovering genetic variants influencing two traits with an application to GWAS of clefts.

### Authors:

D. Ray <sup>1,2</sup>; S. Venkataraghavan <sup>1</sup>; W. Zhang <sup>1</sup>; J.A. Bidinger <sup>1</sup>; E.J. Leslie <sup>3</sup>; M.A. Taub <sup>2</sup>; M.L. Marazita <sup>4,5,6</sup>; T.H. Beaty <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Johns Hopkins University, Baltimore, MD; 2) Department of Biostatistics, Johns Hopkins University, Baltimore, MD; 3) Department of Human Genetics, Emory University, Atlanta, GA; 4) Department of Oral Biology, Center for Craniofacial and Dental Genetics, University of Pittsburgh, Pittsburgh, PA; 5) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 6) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA

---

With a growing number of disease-/trait-associated genetic variants detected and replicated across genome-wide association studies (GWAS), scientists are increasingly noting the influence of individual genetic variants on multiple seemingly unrelated traits. Cross-phenotype association tests, applied on two or more traits, usually test the null hypothesis of no association of a genetic variant with any trait. Rejection of the null can be due to association between the genetic variant and a single trait, with no indication if the variant influences >1 trait.

We propose a new statistical approach, *metapleio2*, to discover genetic variants influencing two traits or phenotypic subgroups using GWAS summary statistics. The traits can be from two studies with/without overlapping samples. *metapleio2* gives an approximate asymptotic p-value for association with both traits and is computationally efficient to be implemented genome-wide. Our simulation studies show that *metapleio2* maintains appropriate type I error at stringent error levels when low-to-moderate sample size or effect size differences in the traits exist.

For an application, we chose non-syndromic orofacial clefts (OFCs)– a common human birth defect, which exhibits strong familial aggregation and has a complex etiology. OFCs are typically categorized as cleft lip with or without cleft palate (CL/P) and cleft palate alone (CP) based on epidemiologic and embryologic patterns. Historically, most studies of CL/P and CP have been analyzed separately. Evidence of shared genetic risk in *IRF6*, *GRHL3* and *ARHGAP29* regions exists, but no variant outside *FOXE1* has approached genome-wide significance in GWAS of CL/P and CP together. When applied to a multi-ethnic GWAS of 2847 CL/P and 611 CP case-parent trios, *metapleio2* identified 6 loci in or near recognized candidate genes: *PAX7* ( $p=7.3 \times 10^{-8}$ ), *IRF6* ( $p=4.6 \times 10^{-12}$ ), *DLG1* ( $p=5.6 \times 10^{-7}$ ), *LIMCH1* ( $p=5.2 \times 10^{-7}$ ), *SHROOM3* ( $p=8.5 \times 10^{-7}$ ) and *NOG* ( $p=6.3 \times 10^{-8}$ ), notably each significant variant appeared to increase risk for one cleft subgroup but decrease risk for the other. Additionally, we replicated known genetic variants in *FOXE1* ( $p=1.8 \times 10^{-7}$ ) and identified a locus in *RAB8A* ( $p=7.1 \times 10^{-7}$ ) that influence risk for both CL/P and CP.

In summary, *metapleio2* confirmed recognized genes and found evidence for new genetic regions exerting shared risk or with opposite effects on risk of CL/P and CP. It shows promise in providing

novel insights into the shared genetic architecture of traits.

## PgmNr 2837: Discovery of pleiotropic variants associated with multiple sclerosis and migraine.

### Authors:

M.K. Horton <sup>1,2</sup>; S.C. Robinson <sup>1</sup>; X. Shao <sup>1</sup>; H. Quach <sup>1</sup>; D. Quach <sup>1</sup>; V. Choudhary <sup>3</sup>; K. Bellesis <sup>3</sup>; T. Chinn <sup>3</sup>; L.F. Barcellos <sup>1,2</sup>; C. Schaefer <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetic Epidemiology and Genomics Laboratory, Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) Computational Biology Graduate Group, University of California, Berkeley, Berkeley, CA; 3) Kaiser Permanente Division of Research, Oakland, CA

---

Risk factors and symptoms of multiple sclerosis (MS) and migraine often overlap, and up to 69% of MS patients experience migraine. However, it remains unknown whether genetic risk variants are shared between the two disorders. We investigated whether genetic risk variants previously identified from GWAS of either condition could be contributing to both conditions (i.e., whether any variant exerted a pleiotropic effect). Data from 1,073 MS cases and 12,000 controls (white, non-Hispanic) from Kaiser Permanente Northern California were utilized. Migraine status was obtained through self-report and a validated electronic health record algorithm, and whole genome data were available. MS or migraine SNPs were identified from prior GWAS, and after quality control, 902 SNPs with MAF>1% were available for analysis. A method by Lutz et al. (2017) was used to identify pleiotropic SNPs in which observed *P* values were compared to genotype permuted *P* values for both phenotypes. SNPs significantly associated with MS and migraine were subsequently used in logistic regression models to estimate the association between each variant and phenotype, adjusting for ancestry. To account for ascertainment bias from obtaining a secondary phenotype (migraine) from a case-control study design that selected participants based on their MS status, the migraine model adjusted for a propensity score representing the probability of MS case-control status given covariates. Preliminary results showed five SNPs were significantly associated with MS and migraine: rs6677309, rs10801908, rs1335532, rs62420820, and rs17066096. Three were protective for MS (rs6677309, rs10801908, and rs1335532) and all increased odds of migraine. Implicated genes include *CD58*, which modulates regulatory T-cells, and several immune genes (*IL20RA*, *IL22RA2*, *IFNGR1* and *TNFAIP3*) within the 6q23 chromosomal region. Results illuminate the shared genetics of MS and migraine, and because several variants increase risk of migraine but decrease risk of MS, there are implications for targeted therapies.

## PgmNr 2838: Genetic variants from immune regulatory genes at MHC region predict *Clostridioides difficile* infection: A genome-wide association study using comprehensive electronic health records.

### Authors:

J. Li <sup>1</sup>; Y. Zhang <sup>2</sup>; A. Jilg <sup>3</sup>; H. Khara <sup>4</sup>; D. Wolk <sup>5</sup>; M. Komar <sup>4</sup>; A. Kolinovsky <sup>6</sup>; D. Rolston <sup>3</sup>; V. Abedi <sup>1</sup>; M. Lee <sup>2</sup>; Regeneron Genetic Center

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical and Translational Informatics, Geisinger Medical Center, Danville, PA 17822, USA; 2) Genomic Medicine Institute, Geisinger Medical Center, Danville, PA 17822, USA.; 3) Division of Internal Medicine, Medicine Institute, Geisinger Medical Center, Danville, PA, 17822, USA; 4) Department of Gastroenterology and Hepatology, Geisinger Medical Center, Danville, PA 17822, USA; 5) Molecular and Microbial Diagnostics and Development, Geisinger Medical Center, Danville, PA 17822, USA; 6) Phenotype Core, Geisinger Medical Center, Danville, PA 17822, USA

---

*Clostridioides difficile* (*C. difficile*), is a major cause of diarrhea following antibiotic treatment or after community contact with hypervirulent strains. Host genetic susceptibility to *C. difficile* infection (CDI) has not been studied on a large-scale with a non-hypothesis driven strategy. The purpose of this study is to identify common variants associated with CDI in a community population as well as antibiotic-treated subset. CDI cases and controls were identified by leveraging the comprehensive Geisinger Electronic Health Record (EHR) database. Geisinger PHASE I and II cohorts were genotyped separately. To account for the potential population structure and relatedness, a linear mixed model was selected for the genome-wide association study (GWAS). After meta-analysis of two GWAS data sets, rs115062572 ( $p=3.51E-06$ ), within the MHC region, was among the top tier signals associated with CDI. Robust evidence of functional impact on neighborhood genes was identified. Subset analysis of patients with notable antibiotic use prior to the index date was performed to characterize a high-risk cohort and the GWAS was reconducted. Subset analysis revealed that rs114751021 at the MHC region was the top signal with a genome-wide significant association with CDI (OR = 2.42; 95%CI = 1.84-3.11;  $p=4.50E-08$ ); while rs115062572 still demonstrated significant association with CDI in antibiotic-treated samples (OR = 2.29; 95%CI = 1.70-2.99;  $p=1.40E-06$ ). Since this locus can be enriched in this CDI high-risk cohort, fine mapping was implemented to identify the causal SNPs or genes. The subsequent eQTL analysis, transcriptomic imputation and association, colocalization testing, and PheWAS annotation, suggested that these SNPs serve as eQTL for NOTCH4 and pQTL for C4A/C4B, and also prioritized the causal SNPs/genes. Minor allele carriers with increased risk for CDI were predicted to have higher levels of NOTCH4 and C4A/C4B, suggesting increased inflammatory response to *C. difficile*. Known associations between IL-8 SNPs (rs4073 and rs2227306) and CDI were replicated ( $p_{\text{lowest}} = 0.036$  and  $0.017$ , respectively). No significant interaction between the top SNPs at the MHC region and other risk factors was identified in association with CDI. This study provides the first evidence that genetic variants at MHC region are associated with CDI, and that those variants significantly alter gene expression of NOTCH4, C4A/C4B, and some MHC molecules in disease-relevant tissues.



## PgmNr 2839: Epistatic combinations of multiple SNPs commonly confirmed in two different case-control studies of rheumatoid arthritis in Japanese populations.

### Authors:

M. Shibata <sup>1</sup>; A. Terada <sup>2</sup>; D. Okada <sup>1</sup>; T. Kawaguchi <sup>3</sup>; I. Yamaguchi <sup>3</sup>; Y. Kamatani <sup>4</sup>; F. Matsuda <sup>3</sup>; J. Sese <sup>2,5</sup>; R. Yamada <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Unit of Statistical Genetics, Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 2) Humanome Lab., Inc., Tokyo, Japan; 3) Unit of Human Disease Genomics, Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan; 4) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan

Genome-Wide Association Study (GWAS) has played a great role as indicating the interaction between genetic variants, especially Single Nucleotide Polymorphism (SNP), and disease, phenotype and behavior for decades. Most of these studies have been conducted to detect association between one SNP and a disease; however, one SNP cannot cause complex diseases. This shortcoming leads to missing heritability problem in GWAS [1]. To overcome this problem, we have developed LAMP algorithm [2], a multiple testing method to detect combinatorial effect. This algorithm analyzes criteria whose frequency of appearance is small and does not change the probability of a false positive [3]. Using this approach in GWAS application, LAMPLINK [4] has become one of the tools that can detect combinations of SNPs that are statistically and significantly associated with the target phenotype.

In this study, LAMPLINK was applied to a case-control study of Rheumatoid Arthritis (RA) from Japanese population, involving 497,190 SNPs and 11,755 individuals, and 61 statistically significant combinations were detected in total. The reproducibility of the combinations was checked in another study of RA from Japanese population, involving 1998 individuals. We confirmed that some combinations consisting of 3 or more SNPs have adjusted p value less than 0.05 in the dataset, showing the common existence of the significant associations of the SNP combinations to RA at least in the independent Japanese populations. Interestingly, the combinations indicate high odds ratio while the independent SNPs in the combinations doesn't have high odds, suggesting the importance to check the combinatorial epistatic effects. Overall our results may provide new aspects for understanding combinations of SNPs.

[1] Blanco-Gómez et al. *Bioessays*. 2016

[2] Terada et al. *PNAS* 2013

[3] Tarone et al. *Biometrics* 1990

[4] Terada et al. *Bioinformatics* 2016

## PgmNr 2840: GWASCellTyper: A fast wrapper for GWAS cell-type enrichment packages.

### Authors:

A. Voda <sup>1,2</sup>; L. Jostins-Dean <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Kennedy Institute of Rheumatology, Oxford, United Kingdom; 2) Saint Edmund Hall, University of Oxford, UK; 3) Christ Church, University of Oxford, UK

---

Genome wide association studies (GWAS) have revealed numerous genetic loci associated with many heritable diseases. A key challenge in translating these findings into biological understanding is finding the tissues and cell types where these risk variants act. A number of gene expression datasets, including both bulk and single-cell RNA-seq (scRNA-seq) experiments, have been used to show interesting cell type enrichments in schizophrenia, Crohn's disease (CD) and other pathologies. There are also multiple statistical packages available for testing enrichment, as well as multiple techniques for detecting and scoring cell-type specificity of genes. Running these tests at scale can be computationally and logistically challenging, which makes finding the right test, and assessing how robust results are to test and parameter choice, difficult.

Here we present a new R package, GWASCellTyper, to efficiently test for enrichment of signal from GWAS summary statistics in cell types using a range of established tests, including LDSC, Magma and SNPsea. GWASCellTyper is designed in a modular fashion to allow flexibility of use. For example, users can enter clustered or unclustered scRNA-seq data, or use pre-computed gene sets to compute enrichment. The package is multithreaded to allow scalable computation, and can compare different techniques for selecting cell-type specific genes (e.g. differential expression vs specificity scoring).

We apply this package to data from a Crohn's disease GWAS using a range of bulk and scRNA-seq datasets. We demonstrate a consistent enrichment of CD risk loci in T cells across tests and datasets, particularly CD4+ memory cells, along with various myeloid cell types, particularly stimulated macrophages. We also find cell types specific to particular datasets or tests, including hematopoietic stem cells and gamma-delta T-cells. We find that enrichment tests that use quantitative gene expression levels generally identify more enriched cell types than tests that only use a discrete list of cell-specific genes. Finally, we show that single-cell datasets often lack power compared to cell sorted bulk datasets, likely due to greater measurement error, but that computational approaches to reduce measurement noise, such as metacell or index cell techniques, can partially restore this power.

We aim to introduce additional enrichment tests, other functional enrichment tests and increased portability into future versions of GWASCellTyper.

## PgmNr 2841: Polymorphisms in the *CLEC4E* human immune gene associate with the transition between latent and pulmonary tuberculosis in Guinea-Bissau (West Africa).

### Authors:

J. Olvany <sup>1</sup>; M. White <sup>10</sup>; L. Sausville <sup>2</sup>; G. Tavera <sup>4</sup>; C. Wejse <sup>7,8</sup>; A. Tacconelli <sup>9</sup>; G. Sirugo <sup>5,6</sup>; S. Williams <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; 2) Departments of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 3) Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 4) Clinical Translational Science Collaborative, Case Western Reserve University, Cleveland, OH; 5) Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 6) Division of Translational Medicine and Human Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 7) Center for Global Health, School of Public Health, Aarhus University, Skejby, Denmark; 8) Bandim Health Project, Danish Epidemiology Science Centre and Statens Serum Institute, Bissau, Guinea-Bissau; 9) Centro Ricerca, Ospedale San Pietro Fatebenefratelli, Rome, Italy; 10) Department of Medicine, Lung Biology Center, University of California San Francisco, San Francisco, California

---

Tuberculosis (TB) is the leading cause of death from a single infectious agent. Although the bacillus responsible for disease, *Mycobacterium tuberculosis* (MTB), is estimated to infect 1.7 billion people worldwide, only a small portion of these people (5-10%) will transition from the latent disease state into active TB. Because only a small fraction of infected people develop active disease, we hypothesized that underlying host genetic variation associates with developing active disease. Variation in *CLEC4E* has been of interest in previous association studies with the same variant showing either no effect or protection from TB. One of the studies using a South African cohort theorized that *CLEC4E* contributes to susceptibility to TB but could not statistically support it, while a Northern Chinese study found it confers protection. For this study we assessed 60 SNPs in 12 immune genes, including *CLEC4E*, using a case-control study from Guinea-Bissau. The 289 cases and 322 controls differed in age, sex, and ethnicity all of which were included in adjusted models. Initial association analysis with unadjusted logistic regression revealed putative association with nine SNPs ( $p < 0.1$ ). Any SNP that was significant below a  $p$ -value of 0.05 was then assessed in an adjusted model. Of the six SNPs that remained significant, three of them were assigned to the *CLEC4E* gene (rs12302046, rs10841847, and rs11046143). Of these, only rs10841847 passed FDR adjustment for multiple testing. Based on these initial association tests, *CLEC4E* seemed to be the best single indicator of PTB risk in this population, thus a deeper analysis of the gene was completed. Haplotype analysis (2-SNP and 3-SNP windows) showed that minor alleles in segments including rs10841847 were the only ones to pass the threshold of global significance as compared to other haplotypes ( $p$ -value  $< 0.05$ ). Linkage disequilibrium patterns showed that rs12302046 exists in high LD with rs10841847 ( $r^2 = 0.63$ ), and all other SNPs lost significance when adjusted for rs10841847 effects. These findings indicate that rs10841847 in the *CLEC4E* is the single best predictor of pulmonary tuberculosis risk in our study population. These results provide experimental evidence for the hypothesis set forth by the study from South Africa.



## PgmNr 2842: Detecting and quantifying clonal expansions in immune repertoire sequencing data.

**Authors:**

B. Howie; P. Fields; H. Al-Asadi; T. Ching; R. Gitterman

[View Session](#) [Add to Schedule](#)

**Affiliation:** Computational Biology, Adaptive Biotechnologies, Seattle, WA

---

A key feature of the adaptive immune system is the ability of lymphocytes to bind non-self antigens, clonally replicate, and carry out effector functions. Immune repertoire sequencing of somatically rearranged T-cell and B-cell receptors (TCRs and BCRs) is a powerful and sensitive tool for measuring clonal population sizes. By tracking changes in clonal populations over time, repertoire sequencing can identify specific TCRs or BCRs that have expanded or contracted in response to perturbations of the immune system, such as the introduction of a pathogen. In this work, we propose and evaluate statistical methods for two related problems: identifying which clones have responded to a stimulus and quantifying the strength of a polyclonal immune response. A core challenge in developing these methods was to make them robust to the number of sampled lymphocytes, which can vary substantially across time points and individuals due to technical and biological processes. To demonstrate these methods, we analyzed peripheral blood from nine volunteers who received the YF-17D vaccine for yellow fever virus, with samples collected immediately before vaccination as well as 14 and 90 days after vaccination. We show that the number of expanded clones can be compared across individuals by correcting for sampling depth, and we describe a clonal expansion score that captures both the breadth and depth of an immune response.

## PgmNr 2843: *HLA B\*57:03* and *C\*03:02* are associated with slow pediatric HIV progression in Botswana and Uganda.

### Authors:

S. Kyobe <sup>1</sup>; S. Mwesigwa <sup>1</sup>; M. Wayengera <sup>1</sup>; G. Anabwani <sup>2</sup>; G.P. Kisitu <sup>3</sup>; K. Maplanka <sup>4</sup>; I. Kasvosve <sup>4</sup>; B. Nsangi <sup>3</sup>; M. Marape <sup>2</sup>; T.J. Masego <sup>4</sup>; C.W. Brown <sup>5</sup>; A. Kekitiinwa <sup>3</sup>; M.L. Joloba <sup>1</sup>; W.S. Mpoloka <sup>4</sup>; G. Mardon <sup>6</sup>; J. Kyosiimire-Lugemwa <sup>7</sup>; N.A. Hanchard <sup>6</sup>; M. Matshaba <sup>2</sup>; Collaborative African Genomics Network (CAfGEN) of the H3Africa Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Microbiology, Makerere University, Kampala, Central Region, Uganda; 2) Baylor-Botswana Children's Clinical Centre of Excellence, Gaborone, Botswana; 3) Baylor-Uganda Children's Clinical Centre of Excellence, Kampala, Uganda; 4) University of Botswana, Gaborone, Botswana; 5) University of Tennessee Health Science Center, Memphis, Tennessee, USA; 6) Baylor College of Medicine, Houston, Texas, USA; 7) MRC/UVRI and London School of Hygiene and Tropical Medicine, Entebbe, Uganda

### Background

HIV remains at epidemic proportions in sub-Saharan Africa especially in pediatrics, despite effective antiretroviral therapy (ART). HIV infected individuals characteristically progress to AIDS leading to mortality, however, 1-5% do not progress to AIDS for at least 10 years in the absence of ART. This restriction of HIV progression is determined by both host genetic and viral factors. The *human leucocyte antigen (HLA) class I* gene is strongly associated with long-term non-progression (LTNP) of HIV in adult Caucasians and Africans. However, *HLA* is highly variable and shows high geographical diversity. In this study we determine the class I *HLA* alleles associated with HIV disease progression in Botswana and Uganda.

### Methods

From a cohort of about 12,000 children enrolled into pediatric HIV care centers in Uganda and Botswana since 2003, 1,000 participants were retrospectively recruited using the electronic medical records. They were stratified equally into two groups of rapid progression and LTNP according to the WHO classification of HIV disease progression. Class I *HLA* typing was done using HLAreporter that derives high-resolution allelotypes from whole exome sequence (WES) data.

### Results

Of the 1,000, 850 participants had WES successfully done and class I *HLA* allele data was correctly typed for 800 participants. After adjusting for gender and country of origin, we identified *HLA-B\*57:03* (OR 3.21 95%CI 1.50-6.85, p=0.030) and *C\*03:02* (OR 4.49 95%CI 1.64-12.23, p=0.049) alleles that are associated with LTNP in children in Uganda and Botswana. In addition, we confirm earlier observations that *HLA-B\*15:10* (OR 0.46 95%CI 0.31-0.79, p=0.030) is associated with rapid HIV progression as seen in Caucasian populations.

### Conclusion

This study demonstrates the utility of whole exome sequencing to study the association of class I *HLA*

and HIV disease progression in African populations. Consistent with previous studies, we identified a major role of *HLA-B* in HIV disease progression, as well as an unreported, uncommon *HLA-C* allele associated with HIV restriction in African pediatric populations. We recommend studies to understand the functional mechanisms of these alleles that are potential targets for HIV immunotherapeutics and vaccines.

## PgmNr 2844: CNV calling from exome data using machine learning detects at least 90% of pathogenic CNVs detectable by microarrays.

### Authors:

P. Danecek <sup>1</sup>; T. Fitzgerald <sup>2</sup>; G. Gallone <sup>3</sup>; D. de Vries <sup>4</sup>; M.E. Hurles <sup>1</sup>; on behalf of the DDD study

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 2) European Molecular Biology Laboratory, Hinxton, Cambridgeshire, United Kingdom; 3) Max Planck Institute for Molecular Genetics, Berlin, Germany; 4) University of Groningen, The Netherlands

---

The diagnosis of genetic diseases in clinical settings relies on performing several different assays, from karyotyping and array-based Copy Number Variant (CNV) detection approaches to whole genome or whole exome sequencing (WES). With the continuing decrease of sequencing costs it would be advantageous to determine all types of genetic variation from a single assay.

CNVs are an important class of disease associated variation. CNVs remain more difficult to call with high accuracy than single-nucleotide variants or short indels. Multiple methods and algorithms have been developed, but no large-scale assessment comparing the accuracy and sensitivity of array-based vs WES-based approaches has been performed. The Deciphering Developmental Disorders (DDD) study provides an ideal data set to test this because of its scale and large overlap of samples with both exon-resolution array-CGH and exome CNV calls.

We sequenced more than 32 thousand exomes from families of 13,462 patients with severe developmental disorders. For 6,832 probands both aCGH and WES data are available which makes it a unique data set to compare the characteristics of callsets obtained from the two platforms. To determine CNVs from the aCGH data we developed and applied an approach, CNsolidate, that integrates CNV calls from six different calling algorithms to give higher accuracy than is achievable with any one caller.

To determine CNVs from WES data, we evaluated four callers (XHMM, CONVEX, CLAMMS and CANOES). We show that sensitivity and specificity of all callers is low and that a Random Forest machine learning approach can be used to integrate the calls. The final callset performs better than any of the individual callers (as observed independently also by other groups).

We find that the Random Forest approach applied on WES data identifies at least 90% of pathogenic CNVs identified from the aCGH data. We note that aCGH arrays in DDD have considerably higher resolution than the arrays typically used in a clinical setting, and that the DDD study is impoverished for large pathogenic CNVs relative to a prospective clinical cohort, therefore the true sensitivity of the WES callset compared to standard clinical aCGH would be even higher.

We also analysed the parental origin of de novo CNVs and show that the proportion that are maternal in origin is significantly higher than for de novo SNVs.



# PgmNr 2845: Dynamic incorporation of multiple in-silico functional annotations empowers rare variant association analysis in large-scale whole genome sequencing studies.

## Authors:

Li <sup>1</sup>; Z. Li <sup>1</sup>; H. Zhou <sup>1</sup>; S. Gaynor <sup>1</sup>; J. Rotter <sup>2</sup>; C. Willer <sup>3,4,5</sup>; G. Peloso <sup>6</sup>; P. Natarajan <sup>7,8,9</sup>; X. Lin <sup>1,7,10</sup>; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Lipids Working Group

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA.; 2) The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA 90502, USA.; 3) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA; 4) Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA.; 5) Department of Human Genetics, University of Michigan, Ann Arbor, MI 48109, USA.; 6) Department of Biostatistics, Boston University School of Public Health, Boston, MA 02118, USA.; 7) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA.; 8) Center for Genomic Medicine and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA 02114, USA.; 9) Department of Medicine, Harvard Medical School, Boston, MA 02115, USA.; 10) Department of Statistics, Harvard University, Cambridge, MA 02138, USA.

## Introduction

Large-scale whole genome sequencing studies have enabled the analysis of rare variants (RVs) associated with complex human traits. Commonly used RV association tests (RVATs) have limited scope to leverage the functions of variants.

## Methods

We propose STAAR (variant-Set Test for Association using Annotation infoRmation), a scalable and powerful method to increase the power of RVATs by effectively incorporating both variant functional categories and multiple complementary functional annotations using a dynamic weighting scheme. For the latter, we introduce “annotation Principal Components” (aPCs), multi-dimensional summaries of *in-silico* coding and noncoding variant functional annotations. STAAR accounts for population structure and relatedness, and can be used for analyses of continuous and dichotomous traits. We highlighted STAAR-O, the omnibus test which aggregates multiple annotation-weighted tests including the burden test, SKAT, and ACAT-V in the STAAR framework. We focused on two types of WGS RV association analysis using STAAR-O: gene-centric functional element-based analysis by grouping variants into functional categories for each protein-coding gene and agnostic genetic region analysis using sliding windows. We applied STAAR-O to identify RV-sets associated with four quantitative lipid traits (LDL-C, HDL-C, TG and TC) in 12,316 discovery samples and 17,822 replication samples from the NHLBI Trans-Omics for Precision Medicine (TOPMed) program.

## Results

In RV gene-centric analysis, STAAR-O identified 25 significant associations with lipids traits. After conditioning on known lipids-associated variants, 13 out of the 25 associations remained significant and could be validated in replication phase, including the association between *NPC1L1* missense RVs and LDL-C, which was not detected by the conventional variant-set tests and has not been observed

in previous studies. In RV sliding window analysis, STAAR-O detected 59 significant 2kb sliding windows associated with lipid traits. 12 sliding windows remained significant after conditioning on known variants and could be validated in replication phase, including associations between an intergenic region near *APOC1P1* and LDL-C that were not detected using conventional tests.

### **Conclusion**

By incorporating multiple variant functional annotations via aPCs, STAAR-O empowers rare variant association analysis and detected novel rare variants association with lipid traits using the TOPMed WGS Freeze 5 data.

## PgmNr 2846: MetaPhat: Detecting and decomposing multivariate associations from univariate genome-wide association statistics.

### Authors:

J. Lin <sup>1</sup>; R. Tabassum <sup>1</sup>; S. Ripatti <sup>1,2,3</sup>; M. Pirinen <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computational Genomics, CdG COE, FIMM, University of Helsinki, Helsinki, Helsinki, Finland; 2) Public Health, University of Helsinki, Helsinki, Finland; 3) Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA, USA; 4) Department of Mathematics and Statistics, University of Helsinki, Helsinki, Finland

---

Multivariate statistical tools that analyze jointly multiple genome-wide association studies have become widely-used due to the increased number of phenotypes gathered in biobanks. While these tools have been shown to considerably boost statistical power over univariate tests, a remaining challenge is to interpret which traits are driving the multivariate association and which traits are just passengers with minor contributions to the genotype-phenotype association.

MetaPhat is a novel computational tool to conduct genome-wide association studies of related multivariate traits based on univariate summary statistics and to decompose the significant multivariate associations into sets of statistically central traits. The decomposition is achieved by tracing the multivariate association p-value from testing on the canonical correlation value, beginning with the full model while traits are iteratively dropped. We use two new concepts to define the central traits. The *driver* traits are a small subset without which the association p-value does not reach the significance level. The *optimal* traits are a small subset of traits that together make the association statistic optimal.

Using univariate GWAS summaries for 21 heritable and correlated polyunsaturated lipid species from 2,045 Finnish samples, MetaPhat detected 7 independent significant SNPs ( $p\text{-value} < 5 \times 10^{-8}$ ). As an example, the missense variant rs7412 in the *APOE* gene, known to associate with low-density lipoprotein cholesterol, was detected in a multivariate analysis by MetaPhat, although the smallest p-value of rs7412 was found to be  $1.1 \times 10^{-4}$  from any univariate GWAS of the 21 traits. Our decomposition revealed that this association was driven by two polyunsaturated lipids with C20:4 acyl chain fatty acid properties. Learning this information without the systematic iterative procedure implemented in MetaPhat would have been infeasible, as it had required considering millions of possible subsets of the 21 traits. MetaPhat, instead, considers only a *few hundred subsets* of the 21 traits and completes decomposition of a variant association in just a few minutes.

MetaPhat is an open source application written in Python with built-in support for multi-processing, quality control, clumping and intuitive visualizations using R. The software is easy to use and suitable for cloud computing. The implementation is freely available under MIT license at <https://sourceforge.net/projects/meta-pheno-association-tracer>

## PgmNr 2847: Common genetic variation as a modifier of familial hypercholesterolemia in the Old Order Amish.

### Authors:

G. Tzoneva<sup>1</sup>; N. Lin<sup>1</sup>; B. Ye<sup>1</sup>; A. Damask<sup>1</sup>; S. Khalid<sup>1</sup>; S. Balasubramanian<sup>1</sup>; A. Baras<sup>1</sup>; M. Zhang<sup>2</sup>; K.A. Ryan<sup>3</sup>; B.D. Mitchell<sup>4,5,6</sup>; C. Paulding<sup>1</sup>; C. Van Hout<sup>1</sup>; N. Gosalia<sup>1</sup>; A.R. Shuldiner<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Regeneron Genetics Center, Tarrytown, NY; 2) Office of Research Affairs, University of Maryland School of Medicine, Baltimore, MD; 3) Department of Medicine, University of Maryland, Baltimore, MD; 4) Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD; 5) Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 6) Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, MD

Founder populations like the Old Order Amish (OOA) have a unique genetic architecture with a high representation of drifted alleles such as *APOB* p.R3527Q, a known familial hypercholesterolemia (FH) allele, present at a frequency of 6.5% in the OOA compared to only 0.06% in non-Finnish Europeans in gnomAD.

We assessed whether common genetic variation in the OOA modifies the effect of *APOB* p.R3527Q on LDL cholesterol (LDL-C) by performing genome-wide gene-by-gene interaction screens using merged exome and array genotyping data on 5,924 OOA individuals (768 of which carry at least one copy of *APOB* p.R3527Q). We utilized linear mixed models to account for relatedness and adjusted for sex, age, age<sup>2</sup> and study. *APOB* p.R3527Q was associated with an increase of 75.7 ( $\pm$  1.7) mg/dL in LDL-C per allele ( $p < 1 \times 10^{-303}$ ), but there were no genome-wide significant interaction effects with *APOB* p.R3527Q. We identified a nominal interaction effect of -0.31 SD in LDL-C ( $p = 1.02 \times 10^{-5}$ ) with rs6531176, located in a regulatory region 1.39 Mb upstream and in moderate linkage disequilibrium (LD) with *APOB* p.R3527Q.

Next, we evaluated the impact of a polygenic risk score (PRS) on LDL-C levels in the OOA. We generated a PRS for LDL-C using previously reported GWAS summary statistics from 188,577 European-ancestry individuals in the Global Lipids Genetics Consortium (GLGC). We pruned the PRS of any variants in LD with *APOB* p.R3527Q ( $r^2 \geq 0.1$ ) by removing a region of 22.8 Mb around *APOB* p.R3527Q from the GLGC source data and assessed the PRS effect on LDL-C in *APOB* p.R3527Q reference and heterozygous carriers. *APOB* reference individuals in the top PRS decile had LDL-C levels of 139.8 ( $\pm$  1.6) mg/dL, while *APOB* reference individuals in the lowest PRS decile had LDL-C levels of 103.5 ( $\pm$  1.2) mg/dL (a difference of 36.3 mg/dL). In comparison, *APOB* p.R3527Q heterozygous carriers in the top PRS decile had LDL-C levels of 211.2 ( $\pm$  4.8) mg/dL, while *APOB* p.R3527Q heterozygous carriers in the lowest PRS decile had LDL-C levels of 181.5 ( $\pm$  3.9) mg/dL (a difference of 29.7 mg/dL). These differences were associated with a nominal interaction effect between the LDL-C PRS and the *APOB* p.R3527Q allele with an effect size of -0.11 SD ( $p = 3.07 \times 10^{-3}$ ).

These results expand our understanding of the allelic architecture of LDL-C in the OOA, indicating that common genetic variation can modify LDL-C in individuals with monogenic FH, findings that are likely

generalizable to FH in other populations.

## PgmNr 2848: Evidence for a causal role of triglycerides in increasing blood pressure: A multivariable Mendelian randomization study of over 820,000 subjects.

### Authors:

H. Rasheed<sup>1,2</sup>; J. Zheng<sup>2</sup>; B. Brumpton<sup>1,2,3</sup>; J. Rees<sup>4</sup>; E.B. Stovner<sup>1</sup>; C. Willer<sup>5</sup>; S. Burgess<sup>4,6</sup>; G.D. Smith<sup>2</sup>; K. Hveem<sup>1</sup>; B.O. Åsvold<sup>1,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norway; 2) MRC Integrative Epidemiology Unit, University of Bristol, UK; 3) Clinic of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim, Norway; 4) Cardiovascular Epidemiology Unit, University of Cambridge, Cambridge, UK; 5) Department of Internal Medicine, University of Michigan, Ann Arbor, MI, US; 6) MRC Biostatistics Unit, University of Cambridge, Cambridge, UK; 7) Department of Endocrinology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway

**Background:** Higher low-density lipoprotein cholesterol (LDL-C), lower high-density lipoprotein cholesterol (HDL-C) and higher triglyceride (TG) levels have been associated with high blood pressure (BP) in observational studies.

**Aim:** To investigate the causal associations of serum lipid levels (LDL-C, HDL-C, and TG) with systolic (SBP) and diastolic blood pressure (DBP) using univariable and multivariable Mendelian Randomization (MR).

**Methods:** Using a two-sample MR, an instrument of 392 lipid-associated independent variants, selected from the Global Lipids Genetics Consortium (n=331,368) was examined for a causal association with SBP and DBP in 69,248 participants from the HUNT cohort and 757,601 participants from the combined UK biobank and International Consortium of Blood Pressure (UKB-ICBP) cohort. Estimates were meta-analyzed across cohorts. Besides inverse variance weighted (IVW) for univariable MR, multivariable MR was used to disentangle the independent effect of individual lipid traits on SBP and DBP. Evidence of pleiotropy was evaluated using MR-Egger. The estimates ( $\beta$ [95% CI]) represent the age- and sex-adjusted change in blood pressure (mmHg) per 1 standard deviation (SD) genetically attributed lipid level.

**Results:** For SBP, IVW meta-estimates in combined cohorts supported a meaningful change in SBP with HDL-C (-0.39[-0.64,-0.14],  $P=0.002$ ) and TG (0.63[0.11,1.16],  $P=0.02$ ), with comparable estimates in UKB-ICBP only. Multivariable meta supported substantial effect for TG (0.61[0.11,1.10],  $P=0.01$ ) only.

In both HUNT and UKB-ICBP cohorts, IVW estimates showed evidence for casual effect of HDL-C (-0.29[-0.55,-0.03],  $P=0.03$  and -0.26[-0.47,-0.04],  $P=0.02$ , respectively) and TG (0.53[0.22,0.84],  $P=7.7 \times 10^{-4}$  and 0.59[0.35,0.82],  $P=1.2 \times 10^{-6}$ , respectively) with DBP with consistent estimates from meta-analysis (HDL-C: -0.27[-0.43,-0.11],  $P=0.001$  and TG: 0.57[0.38,0.76],  $P=3.2 \times 10^{-9}$ ). Multivariable MR along with its meta estimate supported only the independent casual association of TG

(0.53[0.18,0.88],  $P=0.003$  and 0.64[0.36,0.91],  $P=5.7 \times 10^{-6}$ , 0.60[0.38,0.81],  $P=6.3 \times 10^{-8}$ , HUNT, UKB-ICP, Meta respectively) with DBP. MR-Egger regression provided little evidence of pleiotropy.

*Conclusion:* In our cohorts, univariable MR provided evidence for the causal association of multiple lipid traits with blood pressure traits. However, multivariable MR only supported the independent causal role of TG in increasing blood pressure.

## PgmNr 2849: Multi-trait genome-wide multivariate polygenic risk score improves disease prediction of complex traits.

### Authors:

Y. Tanigawa <sup>1</sup>; J. Qian <sup>2</sup>; T.J. Hastie <sup>1,2</sup>; M.A. Rivas <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical Data Science, Stanford Univ, Stanford, California.; 2) Statistics, Stanford Univ, Stanford, California.

---

Polygenic prediction has been extensively used to identify individuals with elevated risk in a given population. Albeit advancements in modern high-dimensional statistics, most of the genetic risk score models are built on a stepwise procedure for each trait, independently: clumping and thresholding based on the significant associations from univariate marginal GWAS summary statistics, limiting the incorporation of informative alleles with modest effect for single or multiple traits and the overall predictive performance. To address this, we built genome-wide multivariate penalized regression models and their corresponding polygenic risk scores (PRSs) using batch screening iterative lasso (BASIL) algorithm implemented in R snpnet package across 94 phenotypes in UK Biobank, including 38 biomarkers and 46 cancer phenotypes. Among the improved predictive performance, we found our model explains 56% of the variance in Lipoprotein (a) levels and predicts prostate, skin, and breast cancers with ROC-AUC of .64, .61, and .59, respectively, using the features in the array-genotyped data alone. To investigate and summarize correlated predicted disease risks across traits, we applied an unsupervised statistical learning approach and characterized latent components of PRSs (multi-PRS LCs). By quantifying phenotype pairwise similarities within multi-PRS LCs, we found the 22 distinct phenotype pairs with more than .5 of absolute value of similarity score within the biomarker PRSs: creatinine and eGFR ( $< -.999$  cosine similarity), LDL and Apolipoprotein B ( $> .990$ ), and HDL cholesterol and Apolipoprotein A ( $> .0.985$ ), all of which are consistent with coherent variant selection in the penalized multivariate Lasso regressions. Using multi-PRS LCs, we further identified outlier individuals and characterized the directions (i.e. the composition of multi-PRS LCs) and magnitude of the elevated genetic risks. We found a set of outlier individuals who have elevated risks for creatinine, cystatin-C, and Urea. Together, our results highlight the benefit of genome-wide characterization and integrative analysis of PRSs across correlated traits. The resource made available with the study, which includes the weights of the individual trait-specific polygenic risk scores from the penalized Lasso regression models and combined multi-PRS LC scores, will accelerate further investigation of PRSs and pleiotropic effects of complex traits.



## PgmNr 2850: Integrative transcriptome-wide association framework identifies many novel replicating genes for lipid traits.

### Authors:

Y. Veturi<sup>1</sup>; T. Hoffmann<sup>2</sup>; E. Theusch<sup>3</sup>; Y. Bradford<sup>1</sup>; H. Hakonarson<sup>4</sup>; I. Kullo<sup>5</sup>; J.E. Miller<sup>1</sup>; P. Sleiman<sup>4</sup>; D. Schaid<sup>6</sup>; W. Wei-Qi<sup>7</sup>; C.M. Stein<sup>7</sup>; D.R. Velez Edwards<sup>8,9,10</sup>; Q. Feng<sup>8</sup>; M. Medina<sup>3</sup>; R. Krauss<sup>3</sup>; N. Risch<sup>2</sup>; M. Ritchie<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Epidemiology & Biostatistics, University of California and San Francisco, San Francisco, CA, 94117, USA; 3) Children's Hospital Oakland Research Institute, Oakland, CA, 94609, USA; 4) Center for Applied Genomics, Children's Hospital of Philadelphia, PA 19104, USA; 5) Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN 55905, USA; 6) Department of Health Sciences Research, Mayo Clinic, Rochester, MN 55905, USA; 7) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37230, USA; 8) Department of Biomedical Informatics in School of Medicine, Vanderbilt University, Nashville, TN 37230, USA; 9) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN 37230, USA; 10) Division of Quantitative Science, Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN 37230, USA

Transcriptome-wide association studies (TWAS) have identified thousands of genes whose *cis*-regulated expression levels associate with a host of complex traits and diseases. In this study, we devised an integrative framework that combines TWAS with probabilistic fine mapping across multiple large-sized cohorts. We performed TWAS using S-PrediXcan on summary-statistics of lipid traits - high-density lipoprotein (HDL), low-density lipoprotein, total cholesterol and triglycerides (TG). Analyses were performed using 31,838 European American (EA) adults from electronic Medical Records and Genomics (eMERGE) network Phase III, 76,627 Non-Hispanic EA adults from Genetic Epidemiology Resource on Adult Health and Aging (GERA), and 188,578 predominantly EA individuals from Global Lipids Genetics Consortium (GLGC). We used tissue-specific weights (~8300 genes) derived from Genotype Tissue Expression Project v7 for adipose subcutaneous (AS), adipose visceral omentum, liver, small-intestine terminal ileum (SI), and whole blood. We identified 375 Bonferroni-significant genes (137 novel and 238 previously known). 42 of the novel genes “replicated” (same direction of effect for the same trait-tissue pair) in at least two cohorts and had an FDR-adjusted gene expression prediction accuracy p-value < 0.05. After conditioning on independently associated SNPs at a locus (GCTA-COJO), we conducted statistical colocalization analyses in each of five tissues (coloc) and filtered out LD-contaminated genes (probability of independent GWAS and eQTL signals > 0.5). We subsequently performed probabilistic fine mapping of causal gene sets (FOCUS) and identified 29 genes that had posterior inclusion probabilities > 0.5 in at least two cohorts for the same trait-tissue pair. We also estimated local SNP heritability for all four lipid traits (HESS). Some of our novel genes replicated in all three cohorts for the same trait-tissue pair (*TRIM74*: AS-HDL, *ZSWIM1*: AS-TG, and *MT3*: SI-HDL) and one replicated in two cohorts for all five tissues (*RP11-109L13.1*: TG).

*RP11-109L13.1* was in a 1.3 MB region that includes *APOC3/A1/A4/A5*; this region explained 2.2%

(GERA) and 1.3% (GLGC) of variation in TG. We are using data from Million Veteran Program, Penn Medicine Biobank, and the UK Biobank as additional replication cohorts and are also running S-PrediXcan “phenome-wide”. We demonstrate the power of this framework over standard GWAS to identify novel genes that have strong evidence for causality.

## **PgmNr 2851: Uncommon, non-coding variants in the *APOE* region are associated with Alzheimer's disease risk, independent of the *APOE* $\epsilon 2/\epsilon 3/\epsilon 4$ alleles.**

### **Authors:**

E.E. Blue <sup>1</sup>; A.Q. Cheng <sup>2</sup>; C.E. Yu <sup>3,4</sup>; Alzheimer's Disease Genetics Consortium

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Division of Medical Genetics, Univ of Washington, Seattle, Washington.; 2) Department of Biostatistics, Univ of Washington, Seattle, WA; 3) Geriatric Research, Education, and Clinical Center, VA Puget Sound Health Care System, Seattle, WA; 4) Department of Medicine, Division of Gerontology and Geriatric Medicine, University of Washington, Seattle, WA

---

The two single nucleotide polymorphisms (SNPs, rs429358 and rs7412) defining the *APOE*  $\epsilon 2/\epsilon 3/\epsilon 4$  alleles are the most well-established genetic loci for the risk of Alzheimer's disease (AD). Other SNPs in the *APOE* region have also been associated with AD risk, including rs2075650 in the intron 2 of *TOMM40* and rs4420638 located 3' of *APOC1*. It has been difficult to determine whether these associations are independent of the strong *APOE* effect, as both rs2075650 and rs4420638 are in linkage disequilibrium (LD) with  $\epsilon 4$  and each other ( $0.30 \leq r^2 \leq 0.65$  in the 1000 Genomes Europeans). Using the Alzheimer's Disease Genomics Consortium (ADGC) genotype data near *APOE* ( $\pm 500$ kb; imputed with 1000 Genomes reference) among subjects with European ancestry, we tested for association between SNPs (minor allele frequency, MAF,  $\geq 0.01$ ) and AD disease status with four mixed models adjusting for sex, population structure, and relatedness: model 1 (N=18,770) without *APOE* adjustment, model 2 (N=18,770) adjusted for the count of *APOE*  $\epsilon 2$  and  $\epsilon 4$  alleles, model 3 (N=8,878) restricted to *APOE*  $\epsilon 3$  homozygotes, and model 4 (N=1,503) restricted to *APOE*  $\epsilon 4$  homozygotes. We find stronger LD between rs429358, rs2075650, and rs4420638 in ADGC data relative to the 1000 Genomes Europeans ( $0.50 \leq r^2 \leq 0.83$ ). As expected, rs2075650 and rs4420638 are strongly associated with AD without *APOE* adjustment (model 1; p-value  $\leq 3.19 \times 10^{-228}$ , odds ratio, OR,  $\geq 2.59$ ). While rs2075650 is significantly associated with AD risk among the *APOE*  $\epsilon 4$  homozygotes (model 4; p-value = 0.0468, OR = 1.33) and trends towards significance under models 2 and 3 (p-value < 0.10), rs4420638 is no longer associated with AD after *APOE* adjustment or stratification. Among the other SNPs near *APOE*, we find significant association between rs192879175 and AD among the *APOE*  $\epsilon 3$  homozygotes (model 3; p-value =  $8.30 \times 10^{-6}$ , OR = 0.50) and suggestive association between 12 additional SNPs and AD after correcting for the estimated number of independent tests (p-value <  $4.74 \times 10^{-5}$ ). These 13 SNPs are uncommon among the *APOE*  $\epsilon 3$  homozygotes (MAF < 0.10), and most are non-coding SNP estimated to have a protective effect for AD. Over half of these 13 SNPs were nominally associated with AD risk in the model with *APOE* adjustment (model 2; p-value < 0.05), with the same direction of effect across models. Together, these results suggest that uncommon, non-coding variants in the *APOE* region also contribute to AD risk independent of the *APOE*.

## PgmNr 2852: Detectable chromosome X mosaicism is rare in peripheral leukocytes of UK Biobank men.

### Authors:

M. Machiela<sup>1</sup>; W. Zhou<sup>1,2</sup>; S. Lin<sup>1</sup>; M. Yeager<sup>1,2</sup>; S. Chanock<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland, USA;  
2) Cancer Genomics Research Laboratory, Frederick National Laboratory for Cancer Research, Frederick, Maryland, USA

---

Genetic mosaicism is the presence of clonal populations of cells harboring post-zygotic mutations that are absent in inherited germline DNA. Age-related mosaicism of the sex chromosomes occurs more frequently than autosomal mosaicism, with detectable mosaic Y loss affecting up to 20% of elderly men and base pair adjusted rates of mosaicism on the female X chromosome observed at 4 times the rate of the autosomes. To our knowledge, the frequency of large structural (>2 Mb) mosaicism on the X chromosome has not been investigated in male populations. We characterize the frequency of genetic mosaicism of the male X chromosome in 196,219 men from the UK Biobank. We systematically scanned for deviations in median log R ratio (mLRR) data of chromosome X probes extracted from Affymetrix UK BiLEVE and UK Biobank Axiom genotyping arrays of blood-derived DNA. We detected 12 (0.01%) men in the UK Biobank with deviations in chromosome X LRR suggesting potential chromosome X mosaicism ( $|\text{mLRR}| \geq 0.15$ ), including 1 man with a potential mosaic gain on Xq and 11 men with gains spanning the whole X chromosome. Only 2 men with mosaic whole chromosome X gains surpassed the median  $\text{mLRR} \geq 0.25$  threshold, suggesting a high proportion of cells in these men are affected by mosaic X events. We did not detect evidence for mosaic losses in UK Biobank men. There was no evidence in the UK Biobank suggesting detected X mosaicism is age-related. We also detected 24 (0.01%) males in the UK Biobank with constitutional XXY (Klinefelter syndrome). The rarity of chromosome X mosaicism in males relative to females likely reflects the importance of the genetic content of the X chromosome for cellular growth and survival. Whereas in females the loss or gain of an inactivated copy of the X chromosome may be tolerated, clonal loss or gain of the X chromosome in males is rare.

## **PgmNr 2853: Direct-to-consumer genetic testing data identified the possible risk markers associated with the oral cavity health among Japanese subjects.**

### **Authors:**

S. Nogawa <sup>1</sup>; S. Yusuke <sup>2</sup>; H. Nakamura <sup>2</sup>; S. Morishita <sup>2</sup>; H. Jia <sup>3</sup>; K. Saito <sup>1,3</sup>; H. Kato <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Genequest Inc. , 5-29-11. Siba, Minato-ku, Tokyo 108-0014, Japan; 2) Research and Development Headquarters, Lion Corporation, Tajima100, Odawara, Kanagawa 256-0811, Japan; 3) Laboratory of Health Nutrition, Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1, Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

---

[Background and Objective] The importance of oral cavity, an organ required for ingesting food, has been increasingly appreciated in parallel with the growing concerns on healthy longevity. However, only a few studies have been conducted to understand the genetic background of oral cavity health, particularly among the Japanese. Thus, to gain new insights about the genetic background of oral cavity health, we conducted the present study to identify the risk markers associated with oral diseases and problems.

[Method] We performed a genome-wide association analysis of more than 10,000 Japanese participants recruited via the direct-to-consumer genetic testing service in Japan. Their genomic DNA was extracted from saliva and was genotyped for about 300,000 single nucleotide polymorphisms (SNPs) using SNP array. We conducted web-based questionnaires about the oral cavity of the subjects. Some examples of the questions asked in the questionnaire are as follows; "Have you (or your relative) ever been diagnosed with periodontal disease by your doctor? (yes/no)", "Have you (or your relative) ever been diagnosed with dental decay by your doctor? (yes/no)", and "Are you worried about bad breath? (yes/no)". The genome-wide association analysis was performed on the acquired data by logistic regression model adjusted for age and sex using PLINK software, and P-value  $< 5 \times 10^{-8}$  was regarded as being genome-wide significant and P-value  $< 1 \times 10^{-5}$  was regarded as being genome-wide suggestive. To evaluate the associated SNPs, we performed mediation and stratified analyses.

[Results and Discussion] We identified some genome-wide suggestive SNPs ( $p < 1 \times 10^{-5}$ ) that were associated with periodontal disease and bad breath. These SNP loci were previously reported in non-Japanese population studies focused on the oral cavity. In conclusion, we identified the candidate risk makers associated with oral cavity health. However, further studies are required to examine the reliability of these markers owing to the low sample size and moderate p-values in our current study. Our findings may provide new clues to understand the underlying mechanisms that influence oral cavity health, and are expected to be exploited as biomarkers for personalized oral cavity treatments.

## PgmNr 2854: Association analysis of handgrip strength in targeted loci from GWAS using longitudinal measures and whole-genome sequence data from the Trans-Omics for Precision Medicine (TOPMed) Program.

### Authors:

C. Sarnowski <sup>1</sup>; H. Chen <sup>2,3</sup>; M.L. Biggs <sup>4</sup>; S. Wassertheil-Smoller <sup>5</sup>; J. Bressler <sup>2</sup>; M.R. Irvin <sup>6</sup>; K.A. Ryan <sup>7</sup>; J.R. O'Connell <sup>7</sup>; D.W. Fardo <sup>8</sup>; D. Karasik <sup>9</sup>; D.P. Kiel <sup>9,10</sup>; J. Murabito <sup>11,12</sup>; K.L. Lunetta <sup>1</sup>; the TOPMed Longevity and Healthy Aging Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 2) Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 3) Center for Precision Health, School of Public Health and School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX; 4) Cardiovascular Health Unit, Department of Medicine and Department of Biostatistics, University of Washington, Seattle, WA; 5) Department of Epidemiology and Population Health, Albert Einstein College of medicine, Bronx, NY; 6) Department of Epidemiology, University of Alabama at Birmingham School of Public Health, Birmingham, AL; 7) Program for Personalized and Genomic Medicine, Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 8) Department of Biostatistics and Sanders-Brown Center on Aging, University of Kentucky, Lexington, KY; 9) Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA; 10) Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA; 11) The Framingham Heart Study, Framingham, MA; 12) Section of General Internal Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA

**Background:** Low or declining handgrip strength, a widely used proxy of muscular fitness and a marker of frailty, predicts a range of morbidities and all-cause mortality. We performed association analyses of handgrip strength in 16 targeted GWAS loci (Willems et al, 2017) using longitudinal handgrip measurements and whole-genome sequence data from the Trans-Omics for Precision Medicine (TOPMed) program.

**Methods:** We analyzed 12,342 ethnically diverse participants from six cohort studies (Amish, ARIC, CHS, FHS, HyperGEN, and WHI) who had between two and six handgrip measures per exam, from one to nine separate exams over time. We selected the maximum observation per participant at each exam and used all exams per participant, totaling 32,266 observations. We conducted association analyses with GMMAT using linear mixed models adjusted for age, sex, height, BMI, study, age×sex, BMI×sex, study×sex, and 11 ancestry principal components. We included random effects for study, kinship and participant to account for correlation of measures across exams. We also performed stratified analyses in European (EA) and African (AA) ancestry ( $N_{EA}=9,538$ ;  $N_{AA}=2,663$  participants). We defined loci using a 500kb window around the lead 2017 GWAS variants.

**Results:** Participants had a mean age of 66.6 years (SD: 13.8) and a mean handgrip of 29.4 kg (SD: 10.8). We detected a significant ( $P \leq 6.9 \times 10^{-6}$ ) 17q21 association (lead TOPMed variant rs4793937,

MAF=0.18,  $P=1.9\times 10^{-7}$ ,  $P_{EA}=4.2\times 10^{-6}$ ,  $P_{AA}=0.003$ ) in the homeobox B3 (*HOXB3*) gene. The association of the lead 2017 GWAS variant (rs2288278, MAF=0.34) was restricted to EA and modest in our sample ( $P_{EA}=0.005$ ), which may be due to the younger age of the 2017 GWAS participants (mean age (SD): 56.8 years (8.0)). The association of rs4793937 remained significant after conditioning on rs2288278 ( $P=2.1\times 10^{-6}$ ), suggesting two distinct signals. The correlation between rs4793937 and rs2288278 was higher in EA ( $r^2=0.41$ ) than in AA ( $r^2=0.11$ ), which may explain why the 2017 GWAS identified only one signal in *HOXB3*. Using GTEx and FUSION quantitative trait loci (QTLs) results of expression (eQTLs) and DNA methylation (mQTLs) in human skeletal muscle tissue, we found that both variants are eQTLs of *HOXB3* and mQTLs of *HOXB3* CpG sites.

**Conclusion:** Leveraging multi-ethnic populations along with longitudinal data can help identify additional signals in GWAS loci associated with complex traits. Further investigation of the function of *HOXB3* in muscle cell is needed.

## PgmNr 2855: Predicting physiological aging rates from quantitative traits using machine learning methods.

### Authors:

J. Ding <sup>1</sup>; E.D. Sun <sup>1</sup>; T.J. Butler <sup>1</sup>; G.R. Abecasis <sup>2</sup>; F. Cucca <sup>3</sup>; I.G. Goldberg <sup>1</sup>; L. Ferrucci <sup>1</sup>; D. Schlessinger <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Institute on Aging, Baltimore, Maryland.; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan Ann Arbor, MI; 3) Institute of Genetic and Biomedical Research, National Research Council, Monserrato, Cagliari, Italy

---

It is widely thought that individuals age at different rates. A method that measures physiological age independently of chronological age could therefore be a first step to understand relevant mechanisms; but searches for individual biochemical markers of physiological age have had limited success. In this study, we assessed the extent to which an individual's physiological age could be determined as a composite score inferred from a broad range of biochemical and physiological trait data. Data were collected in population studies in Sardinia ("SardiNIA") and Tuscany ("InCHIANTI"). We used machine learning strategies on data for ~6,000 Sardinian participants, who ranged in age from 12 to 81. The best predictive models were determined from combinations of dimensionality reduction, classification, and regression algorithms. They reached very strong correlations ( $R > 0.9$ ) between predicted and actual ages. However, some individuals showed differences between predicted and chronological age, which were maintained in repeated visits. Similar results were also seen for participants in the InCHIANTI study. We then defined an Effective Rate of Aging (ERA) for each participant, a continuous trait measured as the ratio of an individual's predicted physiological age to his/her chronological age. The inference that individuals have characteristic rates of aging is supported by findings that in the SardiNIA cohort, values of ERA showed genetic heritability of 40%. Thus, young individuals already showed a rate of aging, and its heritability was sufficient to sustain a genome-wide association study. It pointed to genetic variants influencing the rate of aging in genes that affect telomere length and metabolic activity.



## PgmNr 2856: Assessing age, BMI, and genetic effects on serum IGF-1 in the long life family study.

### Authors:

R. Sherlala<sup>1</sup>; C. Kammerer<sup>1</sup>; A. Kuipers<sup>2</sup>; M. Wojczynski<sup>3</sup>; S. Ukraintseva<sup>4</sup>; M. Feitosa<sup>3</sup>; J. Mengel-From<sup>5</sup>; J. Zmuda<sup>2</sup>; R. Minster<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh., Pittsburgh, Pennsylvania.; 2) Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh., Pittsburgh, Pennsylvania.; 3) Department of Genetics, Washington University School of Medicine in St. Louis; St. Louis, MO; 4) Social Science Research Institute, Duke University, Durham, NC; 5) University of Southern Denmark, Odense, Denmark

---

Serum levels of insulin-like growth factor 1 (IGF-1) decrease with increasing age and are associated with susceptibility to age-related diseases. Measures of adiposity, such as body mass index (BMI) also are associated with susceptibility to age-related diseases. Previous reports of the relationship between IGF-1 and BMI ranged from positive to negative to no relationship, perhaps because previous reports studied different age cohorts. Using data on 4270 participants (aged 24-110 years old [yo]) from the Long Life Family Study (LLFS), we investigated the relationship between IGF-1 and BMI overall and by age groups, and also estimated heritability, and performed linkage analyses on IGF-1. IGF-1 and BMI were positively correlated in the total sample ( $\beta=0.161$ ,  $r^2=0.0038$ ,  $P<0.0001$ ). However, further analyses revealed that the relationship between IGF-1 and BMI varied by age quartile: in the 1st quartile (24-58 yo) the relationship was negative ( $\beta=-0.204$ ,  $r^2=0.011$ ,  $P=0.0008$ ); in the 2nd quartile (59-66 yo) the relationship was negative but non-significant ( $\beta=-0.069$ ,  $r^2=0.0012$ ,  $P=0.28$ ); in the 3rd quartile (67-86 yo) the relationship was positive but non-significant ( $\beta=0.106$ ,  $r^2=0.002$ ,  $P=0.13$ ); and in the 4th quartile (87-110 yo) the relationship was positive ( $\beta=0.388$ ,  $r^2=0.019$ ,  $P<0.0001$ ). This pattern did not differ by sex. We also detected a similar age-related pattern between IGF-1 and BMI using an independent dataset, NHANES III. In LLFS, after adjusting for age and sex, the estimated residual heritability of IGF-1 was 0.40 ( $P<0.0001$ ). Multipoint linkage analysis of IGF-1 (using maximum likelihood methods) revealed a potentially novel quantitative trait locus on chromosome 11 at 78 cM (maximum LOD-score=3.48). In summary, we find age-group effects that may clarify some of the inconsistency in previous literature about the relationship between IGF-1 and BMI. In addition, we identified a possible novel locus for IGF-1 levels. Additional studies of IGF-1 and adiposity measures, as well as additional genetic studies of IGF-1, are needed to better understand the underlying mechanisms involved.

## PgmNr 2857: Heritable determinants of mitochondrial heteroplasmy.

### Authors:

P. Nandakumar <sup>1</sup>; J. O'Connell <sup>1</sup>; C. Tian <sup>1</sup>; A. Paterson <sup>2</sup>; N. Sondheimer <sup>2</sup>; D. Hinds <sup>1</sup>; 23andMe Research Team

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) 23andMe, Inc., 899 W Evelyn Avenue, Mountain View, CA, USA; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada

---

Mitochondrial (Mt) heteroplasmy, or the existence of multiple mitochondrial sequences within an individual, is frequently identified in mitochondrial sequencing. For pathogenic variants, the severity of associated symptoms correlates well to the degree of heteroplasmy. Heteroplasmy can be *de novo* or inherited, and inherited heteroplasms can quantitatively shift during mother-to-child transmission. Our goal was to evaluate heteroplasmy and identify its underlying genetic determinants in the 23andMe database. Our large sample size and the high density of mitochondrial variants on the most commonly used 23andMe array allow us to quantify heteroplasmy and identify nuclear genes in which variation is associated with appreciable statistical power.

We performed Mt variant-level quality control in a subset of 278,196 unrelated European-ancestry samples, which consisted of discarding (1) variants with a high proportion of samples with low intensity, and (2) variants for which we observed extensive heteroplasmy, including those in hypervariable regions of the mtDNA. We retained 327 variants for downstream analyses. As described in previous work (Sondheimer et al. HMG 2011), we computed heteroplasmy from the allele-specific intensity ratios for each of these variants, and to validate the observed heteroplasmy in our data, we compared these variants' intensities in maternal-child pairs and noted a high correlation. Next, to quantify heteroplasmy per-individual and at a large scale, we considered both the mean and the maximum values across this high-quality set of variants as quantitative traits in approximately 1 million unrelated 23andMe European-ancestry samples. Preliminary GWAS of both heteroplasmy definitions identified multiple genes of interest, including the mitochondrial transcription factor *TFAM*, and the mitochondrial deacetylase *SIRT3*. Additional work will be carried out to replicate these findings in other data and also to further study the physical location, potential consequences, and prevalence of heteroplasmy across the Mt genome in the large-scale 23andMe dataset.

## PgmNr 2858: Incidence estimation of monogenic disorders caused by *de novo* variants.

### Authors:

J.A. Lopez Rivera<sup>1,2,3</sup>; E. Perez-Palma<sup>2,4</sup>; J. Symonds<sup>5,6</sup>; S. Zuberi<sup>5,6</sup>; A. Brunklaus<sup>5,6</sup>; R.S. Møller<sup>7,8</sup>; D. Lal<sup>1,2,3,4,9,10</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomic Medicine Institute, Cleveland Clinic Lerner Research Institute, Cleveland, OH, USA; 2) Department of Molecular Medicine, Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH, USA; 3) Epilepsy Center, Neurological Institute, Cleveland Clinic, Cleveland, OH, USA; 4) Cologne Center for Genomics, University of Cologne, Cologne, NRW, Germany; 5) The Paediatric Neurosciences Research Group, Royal Hospital for Children, Glasgow, United Kingdom; 6) School of Medicine, University of Glasgow, Glasgow, United Kingdom; 7) Department of Epilepsy Genetics and Personalized Medicine, Danish Epilepsy Centre, Dianalund, Denmark; 8) Department of Regional Health Research, University of Southern Denmark, Odense, Denmark; 9) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 10) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

**Purpose** Epidemiological disease estimates are not available for the vast majority of rare single-gene disorders due to phenotypic heterogeneity and the absence of nation-wide genomic screens. Here, we use a genetic data based statistical method to predict the incidences of all *de novo* variant-associated, rare, and severe single-gene disorders.

**Methods** We used a well-established mutational model to calculate *de novo* variant rate and predict incidence in 100,000 births for all selected rare single-gene disorders.

**Results** We predicted the disorder incidence for 103 known and 3,104 variant intolerant putative *de novo* variant associated genes. Greater predicted incidences correlated with larger numbers of pathogenic variants annotated in patient variant databases for 1,108 known and putative single-gene disorders ( $p\text{-value} = 7.2 \times 10^{-06}$ ), supporting the validity of the incidence prediction. Furthermore, our predicted incidence rates were in the range of reported epidemiological estimates identified in the literature for 86% (six out of seven) *de novo* variant associated single-gene disorders (*SCN1A*, *SLC2A1*, *SALL1*, *TBX5*, *KCNQ2*, and *CDKL5*).

**Conclusion** In the absence of epidemiological data, our catalog of incidence estimates for 103 established and 3,104 putative *de novo* variant associated single-gene disorders can guide patient advocacy groups, clinicians, researchers, and policymakers in strategic decision making.

## PgmNr 2859: Imputed gene expression identifies significant contribution of common regulatory variants to congenital malformations in a medical biobank population.

### Authors:

T.W. Miller-Fleming <sup>1,2</sup>; D. Velez Edwards <sup>2,3</sup>; G. Chenevix-Trench <sup>4</sup>; N.J. Cox <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA; 3) Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN, USA; 4) QIMR Berghofer Medical Research Institute, Brisbane, Queensland, 4029, Australia

---

Congenital malformations (CM) are structural defects present from birth. CM are detected in 3% of births and are the leading cause of infant mortality; however, studies suggest that a large number of CM cases go undetected by medical professionals. Most defects arise due to genetic variation, in combination with environmental exposures, nutritional deficiency, and parental age. Identifying the genetic risk factors, and the individuals at greatest risk for CM, could provide insight into the prevention or treatment of these disorders. In this study we demonstrate the utility of medical biobanks paired with imputed gene expression to identify the genetic and phenotypic correlates of CM.

We set out to test the hypothesis that genetic variation in the regulatory regions of known CM genes is associated with congenital outcomes in a biobank population. We tested this prediction across 23,000 individuals of European ancestry in the de-identified electronic health record-linked biobank at Vanderbilt University, BioVU. In addition to the rich phenotypic information, genotyping data on these individuals gave us the opportunity to assess genetic variation in regulatory regions of the genome through imputed gene expression. Models for gene expression were built using the large eQTL dataset, GTEx. We find that imputed gene expression of known CM genes is significantly associated with congenital phenotypes in the BioVU population (DNMT3B (DNA Methyltransferase-3 Beta) with congenital anomalies of esophagus  $p = 9.51e-7$ , POLR1A (RNA Polymerase I Subunit A) with chromosomal anomalies  $p = 9.60e-6$ ). Additionally, we generated a phenotype risk score (PheRS) for CM and performed a transcriptome-wide association study to identify all imputed genes correlated with the CM PheRS. This approach resulted in the confirmation of known congenital genes (DHODH (Dihydroorotate dehydrogenase)  $p = 2.27e-11$ , and UBE2W (Ubiquitin Conjugating Enzyme E2 W)  $p = 8.39e-18$ ).

This study examined the effects of regulatory variation on CM phenotypes in a large biobank population. We find that common genetic variation in known CM genes results in imputed gene expression that is significantly correlated with both individual CM phenotypes and the CM PheRS. These results suggest that genomic and phenomic tools paired with large biobank datasets provide an opportunity to confirm known CM genes, in addition to identifying new CM genes and pathways.

## PgmNr 2860: Noninvasive prenatal screening of uniparental disomy.

### Authors:

H. Xu; Y. Guan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC.

Uniparental disomy (UPD) is the inheritance of two copies of a chromosome from one parent and no copy from the other parent. Its prevalence is estimated to be one in 2000 births. UPD may cause imprinting disorders or recessive diseases. Current available methods to detect fetal UPD are invasive, requiring chorionic villus sampling or amniocentesis samples. Here we present a noninvasive method that can reliably detect fetal UPD (isodisomy) and its parental original, which can be an add-on item for noninvasive prenatal screening (NIPS).

We obtain both cell-free DNA (cfDNA) and white blood cell DNA (wbcDNA) from peripheral blood of a women at early pregnancy (gestation age 9-12 weeks) and perform whole genome sequencing for each DNA sample at a depth of about 1X. We define read heterozygosity at SNPs covered by two reads such that AB is read heterozygosity. If fetus has maternal UPD, the read heterozygosity is smaller than that when fetus is normal. If fetus has paternal UPD, the read heterozygosity is larger than that when fetus is normal. This difference depends on fetal fraction, the amount cfDNA that is of fetal origin.

To demonstrate the power of our method, we simulated data of 4 UPD syndromes using 1000 Genome Project genotypes. For fetal fraction as 0.04, 0.06, 0.08, 0.10, the power of maternal UPD7 is 0.934, 0.995, 1.000 and 1.000 (at the type I error rate of 0.001), the power of maternal UPD14 is 0.696, 0.968, 1.000 and 1.000, the power of paternal UPD15 is 0.379, 0.570, 0.899 and 0.931, the power of paternal UPD6 is 0.626, 0.915, 0.983, 0.995. Our simulations suggest that the power of our method critically depends on chromosomal size and fetal fraction.

Our method for noninvasive prenatal screening of uniparental disomy has several merits. First, it is effective screening; Second, it can easily be integrated into the routine NIPS, with no additional experimental cost; Third, our method can directly infer the parental original of UPD, a critical information that affects the clinical consequences of UPD. For example, maternal UPD7 causes Silver-Russell syndrome, but paternal UPD7 is clinically unapparent.

## PgmNr 2861: Gut microbiome and virome profiles associated with lean mass traits among menopausal women.

### Authors:

W.Q. Lv <sup>1</sup>; X. Lin <sup>2</sup>; Z.S. Hu <sup>3</sup>; Z.F. Zhu <sup>4</sup>; H.M. Liu <sup>1</sup>; B.Y. Li <sup>1</sup>; W.D. Shen <sup>1</sup>; K.J. Su <sup>5</sup>; C.L. Ge <sup>6</sup>; F.Y. Lv <sup>6</sup>; F.Z. Sun <sup>4</sup>; J. Shen <sup>2</sup>; H.M. Xiao <sup>7</sup>; H. Shen <sup>5</sup>; H.W. Deng <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Systems Biology and Data information and Reproductive Health, Central South University, Changsha, Hunan, China; 2) Department of Endocrinology and Metabolism, The Third Affiliated Hospital of Southern Medical University, Guangzhou, China; 3) Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, TN, USA; 4) Quantitative and Computational Biology Program, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA; 5) Tulane Center of Bioinformatics and Genomics, Department of Biostatistics and Data Science, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA, United States; 6) LC-BIO Technologies CO., LTD., Hangzhou, 310018, China.; 7) Institute of Reproductive & Stem Cell Engineering, School of Basic Medical Science, Central South University, Changsha, 410000, China

**Background** Sarcopenia, the age-related loss of muscle mass and strength, is a major health problem with the rapid growth of the aging older population. Increasing evidence suggested that gut microbial dysbiosis might contribute to the development of sarcopenia. Here, we investigated the relationship between human gut microbiome and virome, and their associations with lean mass among menopausal women.

**Methods** We conducted a metagenome-wide association study for lean mass in 497 Chinese women, aged 41-65 years. Lean mass for arms, legs, trunk and the total body were measured using dual-energy X-ray absorptiometry. Shotgun metagenomics sequencing was performed on the HiSeq 4000 platform (Illumina), targeting ~7.6 Gb of raw reads per sample. After filtering the raw reads and sequence quality check, we measured the taxa of human gut microbiome and virome using software Centrifuge, and aligned the entire NCBI nonredundant nucleotide sequence database. The taxa present in > 10% samples were subjected to further analysis. We estimated the alpha diversity using Shannon index and performed Spearman's rank correlation tests between each microbial taxa and lean tissue mass. Multiple hypothesis testing was adjusted using the false discovery rate method, with a q value < 0.2 as the significant level.

**Results** We found 24 of the 434 microbiome species to be associated with various lean mass traits, and 2 of the 40 virome species to be associated with trunk lean mass. Of specific note, we identified *Enterobacteria phage P4* (Spearman correlation coefficient SCC = 0.12, q = 0.14) and *Sfi* (SCC = 0.12, q = 0.14) to be positively associated with trunk lean mass, while *Enterobacteria phage P4* was significantly associated with its host *Escherichia coli* (SCC = 0.23, p =  $2.4 \times 10^{-7}$ ) and *Escherichia virus P2* (SCC = 0.21, p =  $1.6 \times 10^{-6}$ ).

**Conclusion** We identified microbiome and virome species which were significantly associated with body lean mass traits and the potential interaction between microbiome and virome. These findings suggested novel biological mechanisms underlying the regulation of human muscle mass.

## PgmNr 2862: Dissecting polygenic risk score for Alzheimer's disease by pathway.

### Authors:

G. Leonenko <sup>1</sup>; M. Shoai <sup>2</sup>; J. Williams <sup>1,3</sup>; J. Hardy <sup>2,4</sup>; V. Escott-Price <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, UK; 2) Reta Lilla Research Laboratories, Department of Neurodegeneration and UK Dementia Research Institute, UCL Institute of Neurology, London, UK; 3) UK Dementia Research Institute at Cardiff University, Cardiff, UK; 4) UK Dementia Research Institute at UCL, London, UK

---

Alzheimer's disease (AD) is the most common form of dementia and is responsible for a huge and growing health care burden in the developed and developing world. Majority of drugs development have been focused on the amyloid cascade hypothesis based on implication of APP metabolism and A $\beta$  production in the etiology of the disease. So far clinical trials for this hypothesis have been unsuccessful showing far more complex nature of the disease. Genome wide, exome and genome sequencing studies have implicated with varying degrees of confidence lipid metabolism, the innate immune system and endosomal vesicle recycling gene networks in disease pathogenesis. Gene networks construction is an imperfect art both because of the knowledge base used in the generation of the pathways and because proteins may have more than one function in more than one cell type. Polygenic Risk Score (PRS) analysis has been successfully applied to different case-control AD datasets in order to predict individual risk of developing AD. This study exploits PRS analysis further to divide polygenic risk by the most recently published biologically relevant gene-sets that were identified as associated with AD.

We examine association of gene-sets polygenic risk scores in ADNI dataset (Alzheimer's Disease Neuroimaging Initiative) comparing a) AD cases to controls (174/224), b) samples with cognitive impairment (MCI) which is pre-clinical stage of AD to controls (344/224) and c) individuals with positive amyloid deposition to amyloid negative individuals (357/304).

This study is important because in terms of modeling the disease through iPSC technologies one might like to assign high or low risk by gene-sets and, eventually, it is possible that one might wish to tailor therapies to gene-sets deficits. Therefore in these analyses we grasp the nettle of dissecting gene-sets analyses to the prediction of disease risk in the largest publicly available dataset with extensive phenotypic data: (ADNI).



## **PgmNr 2863: Genome-wide rare variant imputation in 60,178 subjects identifies novel rare variant candidate loci in late-onset Alzheimer's disease (LOAD): The International Genomics of Alzheimer's Project (IGAP).**

### **Authors:**

A.C. Naj<sup>1,2</sup>; J. Sha<sup>1</sup>; G. Leonenko<sup>3</sup>; X. Jian<sup>4</sup>; B. Grenier-Boley<sup>5</sup>; M. Leber<sup>6</sup>; S. van der Lee<sup>7</sup>; R. Sims<sup>8</sup>; V. Chouraki<sup>5</sup>; J.C. Bis<sup>9</sup>; Y. Zhao<sup>2</sup>; A. Kuzma<sup>2</sup>; B.W. Kunkle<sup>10</sup>; H. Comic<sup>11</sup>; A. Pitsillides<sup>12</sup>; R. Xia<sup>13</sup>; P. Holmans<sup>8</sup>; L.-S. Wang<sup>2</sup>; R.P. Mayeux<sup>14</sup>; L.A. Farrer<sup>12,15</sup>; C. van Duijn<sup>11,16</sup>; J.L. Haines<sup>17</sup>; A.L. DeStefano<sup>12</sup>; M.A. Pericak-Vance<sup>10,18</sup>; A. Ramirez<sup>6</sup>; P. Amouyel<sup>5</sup>; J.-C. Lambert<sup>5</sup>; S. Seshadri<sup>19</sup>; J. Williams<sup>8</sup>; G.D. Schellenberg<sup>2</sup>; International Genomics of Alzheimer's Project (IGAP)

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Biostatistics, Epidemiology, and Informatics/Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 2) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Wales, UK; 4) Brown Foundation Institute of Molecular Medicine, University of Texas Health Sciences Center at Houston, Houston, TX, USA; 5) Epidemiology and Public Health Unit, University of Lille, Lille, France; 6) Department of Psychiatry & Psychotherapy, University of Cologne, Cologne, Germany; 7) Alzheimer Center, VU University Medical Center, Amsterdam, the Netherlands; 8) Dementia Research Institute, School of Medicine, Cardiff University, Wales, UK; 9) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; 10) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 11) Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands; 12) Department of Biostatistics, School of Public Health, Boston University, Boston, MA, USA; 13) Imm Center for Human Genetics, University of Texas Health Sciences Center at Houston, Houston, TX, USA; 14) The Taub Institute for Research on Alzheimer's Disease and the Aging Brain and The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA; 15) Department of Neurology, School of Medicine, Boston University, Boston, MA, USA; 16) Nuffield Department of Population Health, St. Cross College, Oxford University, Oxford, England, UK; 17) Department of Population & Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH, USA; 18) Dr. John T. Macdonald Foundation, Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 19) Alzheimer's Institute, UT Health San Antonio, San Antonio, TX, USA

---

IGAP meta-analyses of genome-wide association studies (GWAS) have previously identified 30 susceptibility LOAD loci in addition to *APOE*, the majority common (minor allele frequency (MAF)>0.02). To identify novel rare variant, gene, and pathway associations, IGAP consortia used using Minimac3 on the Michigan Imputation Server to impute variants of MAF>0.0008 in 42 GWAS datasets (24,466 cases /39,951 controls) to the Haplotype Reference Consortium (HRC) r1.1 reference panel (64,976 haplotypes/39,235,157 SNPs) (MAF>0.0008) for discovery association testing. For variants with MAF>0.01, logistic regression was performed using imputed genotype probabilities in SNPTEST (generalized linear mixed model in R for family-based variants) and fixed-effects meta-



analysis performed using METAL. Variants with  $MAF \leq 0.01$  were meta-analyzed using score-based tests with the program SeqMeta/R. Both analyses adjusted for age, sex, and population substructure. Gene-based and pathway associations were examined using VEGAS2. Analyses of ~39.2M genotyped or imputed SNVs confirmed single variant associations in 26 of 30 known IGAP LOAD loci at suggestive levels of significance ( $P < 10^{-5}$ ), with 12 known loci attaining genome-wide statistical significance ( $P < 5 \times 10^{-8}$ ). Newly observed associations included variation at *APP* (rs112547745, OR[95% CI]=1.10 [1.06, 1.14],  $P=8.70 \times 10^{-7}$ ), an early-onset AD gene; variants in genes associated with cardiovascular traits [*IGSF5* (rs16998166, OR[95% CI]=0.74 [0.65-0.84],  $P=4.23 \times 10^{-6}$ ), *ACE* (rs116112765, OR[95% CI]=1.28 [1.15-1.42],  $P=2.95 \times 10^{-6}$ ), *LIPG* (rs2156500, OR[95% CI]=0.94 [0.91-0.96],  $P=2.59 \times 10^{-6}$ ), and *SCARB1* (rs12229555, OR[95% CI]=0.93 [0.90-0.96],  $P=5.28 \times 10^{-6}$ ), which encodes an HDL cholesterol receptor]; and variants in genes involved in neurodegenerative processes, like *TRIT1* (rs61781270, OR[95% CI]=1.08 [1.04-1.11],  $P=3.54 \times 10^{-6}$ ), encoding an amyloid fiber-forming protein. Gene-based analyses identified several genes involved in innate immunity, including *BTRC* and *DCUN1D5* (both  $P < 10^{-6}$ ). Pathway analyses implicated neuronal development [GO:0030182 ( $P_{\text{empirical}}=10^{-6}$ ) and GO:0048666 ( $P_{\text{empirical}}=1.20 \times 10^{-5}$ )]. Replication of these signals is ongoing. Several novel LOAD candidate loci, including those with prior associations with cardiovascular traits and early-onset AD, were identified using high-quality imputation of rare and low-frequency variants in IGAP, highlighting a multitude of genetically-influenced pathologic processes in LOAD development.

## PgmNr 2864: Genetics of handedness in Alzheimer's disease.

### Authors:

L. Qu <sup>1</sup>; Y. Zhao <sup>1</sup>; E.E. Mlynarski <sup>1</sup>; M. Tang <sup>1</sup>; W. Wang <sup>1</sup>; J.M. Ringman <sup>2</sup>; L.S. Wang <sup>1</sup>; G.D. Schellenberg <sup>1</sup>; A.C. Naj <sup>1,3</sup>; Alzheimer's Disease Genetics Consortium (ADGC)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA.; 2) USC Memory and Aging Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 3) Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

---

The skewed distribution of handedness, or laterality, with 87-90% of individuals favoring their right hand for complex manual tasks and 10-13% favoring their left hand, is believed to be a product of differences in brain morphology, and several studies have suggested that lefthandedness (LH) correlates with elevated risk of Alzheimer's disease (AD) and earlier onset of cognitive deficits, while others have suggested a protective effect of lefthandedness against AD and slower disease progression. In this study, we examined handedness in individuals with and without AD to identify genetic contributors to handedness which may affect Alzheimer's disease risk. We tested genome-wide association (GWA) with handedness in 7,541 Caucasian subjects in the AD Genetics Consortium (ADGC) (3,553 cases/3,988 controls) with genotyping imputed to the HRC r1.1 panel (~39.2 million SNPs) and available handedness and AD data; primarily analyses adjusted for population substructure and AD case-control status; secondary control-only analyses examined handedness in a sample more representative of the general population. Examining LH vs. righthandedness (RH), multiple variants upstream the chromosome 7q32.1 gene *TSPAN33* (encoding tetraspanin 33) demonstrated genome-wide significant association (GWS;  $P < 5 \times 10^{-8}$ ), including rs112574546 (minor allele frequency (MAF)=0.047; OR[95% CI]=2.21 [1.76, 2.77],  $P = 9.78 \times 10^{-12}$ ). Associations near GWS were also observed upstream of the 7q11.23 gene *HIP1* (encoding huntingtin interacting protein 1) and the 12q21.31 genes *SLC6A15* (encoding a solute carrier family 6 transporter protein) and *TSPAN19* (encoding tetraspanin 19). Secondary analyses using only cognitively normal controls also observed the regional LH association with *TSPAN33* (OR[95% CI]=2.29 [1.69, 3.10],  $P = 7.07 \times 10^{-8}$ ), although similar evidence for association was not observed for *HIP1*, *SLC6A15*, or *TSPAN19*. Similar patterns of association were observed in analysis comparing those with LH or ambidexterity (AMBI) against RH. *TSPAN33* is a member of the TspanC8 group of tetraspanins that have been shown to regulate *ADAM10* (encoding a disintegrin and metalloprotease 10), which cleaves amyloid precursor protein and is known to affect AD risk. Additional analyses, including polygenic risk score and Mendelian randomization analyses are on-going. This work suggests that genetic contributors to handedness may also influence the pathology of and affect risk of Alzheimer's disease.

## PgmNr 2865: Whole-exome sequencing in African American cohort confirms previously identified Alzheimer's disease genes and suggest new risk variants in *SLC25A39* and *PHF1* genes.

### Authors:

J. Ramos<sup>1</sup>; B. Vardarajan<sup>3</sup>; F. Rajabli<sup>1</sup>; P. Whitehead<sup>1</sup>; K. Hamilton-Nelson<sup>1</sup>; L. Adams<sup>1</sup>; T. Starks<sup>1</sup>; J. Vance<sup>1,2</sup>; A. Griswold<sup>1,2</sup>; G. Schellenberg<sup>4</sup>; J. Haines<sup>5,6</sup>; L. Farrer<sup>7</sup>; L.S. Wang<sup>8</sup>; G. Byrd<sup>9</sup>; E. Martin<sup>1,2</sup>; R. Mayeux<sup>10</sup>; G. Beecham<sup>1,2</sup>; M. Pericak-Vance<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 3) Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University, New York, New York; 4) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; 5) Department of Population & Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 6) Cleveland Institute for Computational Biology, Cleveland, OH; 7) Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, Massachusetts; 8) Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 9) Maya Angelou Center for Health Equity (MACHE) / Wake Forest School of Medicine, Winston Salem, NC; 10) Department of Neurology and Sergievsky Center, Columbia University, New York, New York

Individuals of African American (AA) ancestry are more likely to develop Alzheimer's disease (AD) compared to those of non-Hispanic white (NHW) ancestry. However, AA remain underrepresented in genomic studies of the etiology of AD.

To begin addressing this we have assessed whole-exome sequencing (WES) data of 3,810 unrelated AA individuals (1,715 AD cases and 2,095 controls) obtained from the Alzheimer's Disease Genetics Consortium (ADGC) and the Washington-Heights/Inwood Columbia Aging project (WHICAP). To our knowledge, this is the largest WES-AD study in AA to date.

Cases and controls were at least 60 years old. Cases met NINCDS-ADRA criteria for AD on clinical assessment or presence of AD upon neuropathological exam. We conducted single-variant (with PLINK) and gene-based (with SKAT-O) association analyses, adjusting for population substructure (PC1, PC2 and PC3), age and sex covariates.

Our results confirmed the association of AD with variants in *APOE* and *ABCA7* genes and in the *APOE* region two previously identified risk factors in AA. Two new variants, rs41267647, an intron variant in the *PHF1* gene and rs79808124, an intron variant in the *SLC25A39* gene were significantly associated with AD but not genome wide ( $p < 1.00E-5$ ). Suggestive odds ratios and 95% CI were 0.4771 (0.351-0.647) and 1.637 (1.316-2.036) respectively.

Recent experiments revealed that mutations in the *SLC25A39* Drosophila homolog resulted in accumulation of reactive oxygen species, mitochondrial dysfunction, and neurodegeneration. rs41267647, located in the *PHF1* gene exerted a protective effect in AD. This gene encodes a protein involved in epigenetic regulation of gene expression. rs41267647 is also a strong eQTL for *CUTA*, a

gene that affects A $\beta$  generation through beta-secretase 1-mediated *APP* processing. In summary, our study suggests two new variants associated with AD in AA that have not been detected in NHW.

# PgmNr 2866: Polygenic risk scores as screening tools for Alzheimer's disease clinical trials.

## Authors:

C. Whelan; S. Bian; J.Z. Liu; C.Y. Chen; W. Wang; M.D. Laloti; F. Gao; H. Runz; for the Alzheimer's Disease Neuroimaging Initiative (ADNI)

[View Session](#) [Add to Schedule](#)

**Affiliation:** Research and Early Development, Biogen Inc., Cambridge, MA.

---

## Background

Polygenic risk scores (PRS) can be used to predict the diagnosis of Alzheimer's disease (AD), yielding predictive area under the ROC curve (AUC) estimates of 0.8+, even after accounting for *APOE* e4. PRS may also serve as a useful screening tool for AD clinical trials, enhancing the prediction of pathological beta-amyloid ( $A\beta$ ) levels prior to screening via CSF lumbar puncture or PET imaging. Here, we aimed to retrospectively investigate the utility of PRS as a screening tool for ENGAGE and EMERGE, two large-scale Phase III trials of aducanumab, an  $A\beta$ -targeting therapeutic candidate for AD.

## Methods

We stratified samples from the Alzheimer's Disease Neuroimaging Initiative using the screening protocol for ENGAGE/EMERGE (<https://clinicaltrials.gov/ct2/show/NCT02484547>). Samples were further stratified by  $A\beta$  pathology, defined by a PET SUVR  $>1.11$ , or CSF phospho-tau/ $A\beta$ 42 ratio  $>-3.6$ . In total, we included 678 prodromal AD patients (268  $A\beta$ -negative, 410  $A\beta$ -positive) and 69 mild AD patients (7  $A\beta$ -negative, 62  $A\beta$ -positive). PRS were constructed using summary statistics from the most recent GWAS of clinically-confirmed AD (Kunkle et al., 2019), applying LD clumping ( $r^2$  cutoffs 0.2, 0.5 and 0.8) and the PRS-CS algorithm (shrinkage parameters of 0.01, 0.0001 and auto-select) to genome-wide SNPs. PRS distributions were compared between  $A\beta$ -positive and  $A\beta$ -negative individuals using logistic regressions in R.

## Results

In mild AD, prediction accuracy was moderate in a model using weights derived via auto-selected parameters (AUC = 0.80; 95% CI = 0.64-0.97), and lower after *APOE* exclusion (AUC = 0.49; 95% CI = 0.22-0.75). In prodromal AD, the best prediction accuracy for  $A\beta$ -positivity was achieved using auto-tuned parameters (AUC = 0.67; 95% CI = 0.62-0.72), which lowered when *APOE* was excluded (AUC = 0.54; 95% CI = 0.48-0.60). Prodromal AD patients with a PRS of  $\geq 0.02$  (representing 12% of all prodromal patients) had a 3.3 fold increased likelihood of PET- or CSF-defined amyloid positivity compared to patients with a PRS  $<0.02$ .

## Conclusion

The predictive power of PRS for  $A\beta$  positivity in AD is moderate when scores are constructed from genetic studies of AD susceptibility, and currently uninformative when *APOE* is removed. Construction of PRS using alternative training data (such as GWAS of PET- or CSF-defined  $A\beta$ /tau load), integration of exome sequencing data, and combination with blood biomarkers should be considered to further evaluate the clinical utility of PRS.

## PgmNr 2867: Genome-wide linkage analysis of African-American Alzheimer disease families.

### Authors:

P.L. Whitehead Gay <sup>1</sup>; J. Jaworski <sup>1</sup>; B.W. Kunkle <sup>1</sup>; F. Rajabli <sup>1</sup>; L.D. Adams <sup>1</sup>; T. Starks <sup>2</sup>; A. Goate <sup>3</sup>; M. Cuccaro <sup>1</sup>; J.M. Vance <sup>1</sup>; E.R. Martin <sup>1</sup>; T. Foroud <sup>4</sup>; R. Mayeux <sup>5</sup>; C. Reitz <sup>5</sup>; G.S. Byrd <sup>2</sup>; G.W. Beecham <sup>1</sup>; M.A. Pericak-Vance <sup>1</sup>; The NIA-LOAD Family Based Study

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Hussman Institute for Human Genomics, University of Miami, Miami, Florida.; 2) Maya Angelou Center for Health Equity, Wake Forest University, Winston-Salem, N.C.; 3) Icahn School of Medicine at Mount Sinai, New York City, New York; 4) Indiana Alzheimer's Disease Center, Indianapolis, IN; 5) Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY

---

Genetic studies of African-Americans (AA) with late-onset Alzheimer's disease (LOAD) have been limited, despite evidence suggesting an increased risk for Alzheimer disease (AD), and differential genetic effects across ancestral populations. In this study, we assessed 44 multi-generational AAAD families from the Research in African American Alzheimer Disease Initiative (REAAADI) and LOAD using genome-wide linkage analysis to prioritize regions containing risk or protective genes for AAAD. As part of REAAADI and LOAD, genotype data were generated for 44 families (137 affected and 282 unaffected) and whole genome sequence data for 33 families. We performed linkage analyses in 44 families using genotype data. Three models were assessed using Merlin software: 1) two-point parametric dominant affected-only and age-dependent penetrance models 2) multipoint parametric dominant affected-only dominant model (MPT-AO) and 3) Non-parametric linkage (NPL) multipoint. WGS data were used to prioritize variants in the consensus regions based on segregation with disease among affected individuals, rarity, and annotation for putative function. The MPT-AO model had 4 peaks over an HLOD of 1.2 with the strongest peak seen on Chromosome 7 (HLOD 2.49, four families contributing >0.58). Family specific peaks over 1.4 were seen in two families in MPT-AO model. The strongest signal localized to a region on Chromosome 12 (LOD=3.18) in the parametric dominant model. This overlaps with a previously reported linkage peak in Non-Hispanic White (NHW) families in (Scott et al., 2000) and in the ADSP Linkage analysis (Beecham et al., 2019). The region on chromosome 12 includes *LRRK2* and *CNTN1* genes. Missense mutations in *LRRK2* reported as a major cause of inherited and sporadic Parkinson disease, and *CNTN1* identified as a new candidate locus for Lewy Body dementia. However, in the segregation analysis using WGS data we didn't observe any variants in *LRRK2* and *CNTN1* genes that segregate with the disease in the family that support linkage peak. Linkage analysis of AA families confirmed a previously identified family specific peak on chromosome 12. Our results didn't show any risk variants in known dementia genes such as *LRRK2* and *CNTN1*. The overlapping of the linkage peak on chromosome 12 with previously reported linkage peaks and known dementia genes emphasize the need for additional studies of all genes at the region by using segregation analysis and bioinformatics tools.

## **PgmNr 2868: Developing statistical models to study gene-gene interactions underlying autism spectrum disorder.**

**Authors:**

Q. Li

[View Session](#) [Add to Schedule](#)**Affiliation:** 2500 University Dr NW, Calgary, AB, Canada

---

Through many years' efforts, thousands of genes are identified to be associated with the Autism Spectrum Disorder (ASD). However, the interpretation of the genetic variants tailoring to specific molecular mechanism is largely unsolved, leaving gaps between the statistical association and functional follow-ups. This is partly due to that fact that the current statistical methods do not integrate biological knowledge and the statistical learning and the biological interpretation are done sequentially. So, the development of statistical models specific to molecular mechanics is urgently needed to bridge this gap. In this project, as a proof-of-concept, we have developed a novel statistical model leveraging information of 3-D genomic conformation assessed by Hi-C experiments and the power of machine learning algorithms to further elucidate the genetic underpinning of complex behaviors of ASD. By applying the novel model on benchmark datasets of more than seven thousand individuals from the MSSNG consortium, we have identified interesting results indicating the association between the 3-D structure and ASD risk. The outcome is naturally interpretable by default as the 3-D genomic conformation is driven by the design of the statistical models. This project can further basic understanding of 3-D genetics of diseases and provide a new way to understand the pathology of other diseases.

## PgmNr 2869: Transmission patterns of ultra-rare variants in autism spectrum disorder.

### Authors:

A.B. Wilfert<sup>1</sup>; T.N. Turner<sup>1</sup>; A. Sulovari<sup>1</sup>; S.C. Murali<sup>1,2</sup>; T. Wang<sup>1</sup>; H. Guo<sup>1,3</sup>; B.P. Coe<sup>1</sup>; E.E. Eichler<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genome Sciences, University of Washington, Seattle, Washington, 98195, USA.; 2) Howard Hughes Medical Institute, University of Washington, Seattle, Washington, 98195, USA.; 3) Center for Medical Genetics & Hunan Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, 410078, China

---

Autism spectrum disorder (ASD) is highly heritable but genetically and phenotypically heterogeneous. Although *de novo* mutations (DNMs) and copy number variants (CNVs) have a well-known role in ASD pathogenesis, collectively these mutations are estimated to account for only 19-30% of cases. To further understand the heritability of ASD we combined whole-exome and whole-genome sequencing data from 9,309 families with ASD from five distinct studies and identified over 2.9 million private, coding variants (heterozygous single-nucleotide variants and indels observed only once across all families and transmitted to at least one affected or unaffected child). Using aggregate burden tests we find an increased burden of private, likely-gene disrupting variants in probands as compared to siblings for genes showing an enrichment of DNMs in autism individuals (OR = 4.93,  $p = 8.24e^{-3}$ ). Previous analyses by our group found that private variant burden increases with increasing gene constraint. When we exclude known DNM-enriched genes from this analysis, we find that at least 92.3% of the burden remains and the percent of remaining burden increases with gene constraint, suggesting that private variants are acting on a distinct set of genes that have not yet been identified by DNM enrichment analyses. Although still underpowered to identify specific genes, we developed a gene-specific binomial testing framework and have identified a dozen high-likelihood candidates. As a group, we confirm that these genes are more intolerant to mutation (OR = 6.47,  $p = 4.24e^{-3}$ ). Modeling the relationship between private variants, DNMs, CNVs, as well as common risk variants, we find an inverse relationship between *de novo* and private mutations versus common variant burden among individuals with ASD, pointing to a highly complex genetic model of autism with multiple sources of disease etiology.



## PgmNr 2870: Genome-wide association analysis demonstrates that the genetic risk burden for multiple sclerosis confers earlier onset.

### Authors:

F.B.S. Briggs <sup>1</sup>; M.F. Davis <sup>2</sup>; S. Brugger <sup>2</sup>; J. Beales <sup>2</sup>; E. Misicka <sup>1</sup>; J. Denny <sup>3</sup>; S. Loomis <sup>4</sup>; P.G. Bronson <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio.; 2) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah.; 3) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee.; 4) Translational Biology, Biogen, Boston, Massachusetts

---

**Objective:** To characterize the genetic component of age of onset (AOO) in person with multiple sclerosis (MS).

**Background:** AOO is a strong predictor of MS evolution independent of disease duration. Factors influencing AOO are not well known though *HLA-DRB1\*15:01* (DR2; the primary genetic risk factor), female sex, and relapsing subtypes confer earlier onset.

**Design/Methods:** We conducted a genome-wide association meta-analysis and evaluated the narrow-sense heritability ( $h^2$ ) of MS AOO in cases of European ancestry. The meta-analysis (N=2,376) included the Accelerated Cure Project (ACP1; N=1,076), ACP2 (N=191), and Vanderbilt University Medical Center BioVU (N=1,109) cases. There were  $1.6 \times 10^6$  SNPs in  $\geq 2$  datasets (MAF  $\geq 0.01$ , HWE  $p > 0.0001$ ). A tagging SNP rs3135388A determined DR2. A genetic risk score (GRS) was constructed based on 200 risk SNPs outside 6p21 where DR2 resides. Linear regression analyses were conducted adjusting for ancestry and sex using PLINK v1.9. Meta-analyses were conducted using METAL weighting by standard errors.  $h^2$  was estimated using GCTA, partitioning by MS risk and non-risk loci.

**Results:** The 6p21 locus was the most significant region associated with AOO ( $p < 3 \times 10^{-6}$ ). MS cases homozygous for DR2 were 2.6 years younger at onset compared to non-carriers ( $p = 0.0008$ ). The largest impact on AOO was a splice donor *SCN7A* (a voltage-gated sodium channel protein) variant ( $\beta = -7.5$  years,  $p < 3 \times 10^{-6}$ ). Additional top ranking associations ( $p < 4 \times 10^{-5}$ ) were observed for variants in *MTG2*, *SLC30A10*, *MIR548H4*, *STXBP6*, *PDE11A*, *PLAR2R1*, *CPAMD8*, *A2ML1*, *PLXDC2*, *MMP7*, and *CIITA*.

10% of MS risk SNPs outside 6p21 were associated with AOO ( $p < 0.05$ ), including the *CLEC16A-CIITA* variants. Increases in GRS was associated with earlier onset across datasets ( $p < 0.05$ ). Furthermore, AOO  $h^2$  ranged between 21-26%, and established MS risk loci explain a portion of this heritability. Exploratory analyses in ACP1 showed, the predicted AOO was 3.8 years earlier for DR2 carriers with the highest GRS quantile burden (30.5 years) versus DR2 non-carriers with the lowest GRS quantile burden (34.3 years).

**Conclusions:** The MS genetic risk component significantly predisposes susceptible individuals for earlier onset, reducing the time from initiation of pathogenetic processes (particularly those

associated with DR2 and DR2 regulation [*CIITA*]) to the appearance of clinical symptoms. Replication analyses are ongoing in Biogen Phase 3 trials (N=1,702; ADVANCE, ASCEND, and DECIDE).

## PgmNr 2871: Pharmacogenetic analysis of known variant associated with IFN- $\beta$ -induced liver injury replicated in African Americans with multiple sclerosis.

### Authors:

S.W. Brugger <sup>1</sup>; N. Xinnan <sup>2</sup>; J.C. Denny <sup>2</sup>; M.F. Davis <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, Utah, USA.; 2) Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

---

Multiple sclerosis (MS) is a debilitating disease characterized by the demyelination of the axons of the central nervous system. Disease-modifying therapies (DMT), such as interferon- $\beta$  (IFN- $\beta$ ), have been successfully used to reduce relapses, but its use has been observed to induce liver damage in some patients. rs2205986 has previously been associated with IFN- $\beta$ -induced liver injury in cohorts of MS patients of European ancestry ( $P = 2.3 \times 10^{-8}$ ). The aim of this study was to replicate this result in a cohort of MS patients of African American ancestry. Of the 691 African American MS patients in Vanderbilt University Medical Center's BioVU database, 151 had genotyping information available for analysis. Manual review of these patients' electronic health records (EHR) identified 82 patients with clearly defined IFN- $\beta$  treatment periods. Laboratory values for liver enzymes alkaline phosphatase (AlkP), bilirubin, aspartate aminotransferase (AST), and alanine aminotransferase (ALT) that were recorded within the IFN- $\beta$  treatment period were extracted from the patients' EHRs using R v3.4.3, leaving 68 patients for the final analyses. The liver enzyme values for each test were normalized to the upper limit of normal. One patient was excluded from the analyses for AST and ALT values exceeding three standard deviations from the mean. A linear regression analysis was performed using PLINK v1.90 with age, sex, and first four principal components included as covariates. No significant associations were identified between rs2205986 and elevated AlkP, bilirubin, and AST levels ( $P = 0.634$ ,  $P = 0.701$ ,  $P = 0.197$ ). However, rs2205986 was significantly associated with elevated ALT levels ( $P = 0.015$ ,  $\beta = 3.61$ ). Additionally, results for each test trended in the direction with those previously reported. A genome-wide association study was performed for each liver enzyme. Four SNPs of genome-wide significance were identified. rs77134863 ( $P = 6.01 \times 10^{-10}$ ,  $\beta = -3.80$ ), rs58470026 ( $P = 3.93 \times 10^{-9}$ ,  $\beta = -3.75$ ), and rs200464088 ( $P = 5.97 \times 10^{-9}$ ,  $\beta = -3.75$ ) were associated with decreased levels of ALT, indicating a reduced risk for IFN- $\beta$ -induced liver injury. By contrast, rs254276 ( $P = 2.56 \times 10^{-8}$ ,  $\beta = 1.17$ ) was associated with increased levels of AST, indicating an elevated risk for IFN- $\beta$ -induced liver damage. These results may prove valuable in informing MS clinicians and patients about specific risks associated with IFN- $\beta$  use and guide treatment to suit the needs of the individual patient.

## PgmNr 2872: Sex-specific genetic drivers in common diseases with susceptibility sex bias: The case of multiple sclerosis.

### Authors:

Z. Fang<sup>1,2,3,4</sup>; Sex and X/Y Working Group, on behalf of the International Multiple Sclerosis Genetics Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Systems Biology and Computer Science Program, Ann Romney Center for Neurological Diseases, Department of Neurology, Brigham & Women's Hospital, Boston, MA; 2) Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA; 3) Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, MA; 4) Harvard Medical School, Boston, MA

---

**Background:** Several common diseases exhibit sex bias, i.e. differential risk between females and males. The sex bias can be evident in any facet of a given disease, among which disease susceptibility is of upmost importance, since we can translate genetic knowledge into potentially actionable prediction models. Multiple sclerosis (MS) is a common neurodegenerative inflammatory disease that exhibits a well-described sex bias. Females are two-fold or higher risk to develop MS with little understanding of the drivers. We have recently compiled the genetic map of MS, describing 32 genome-wide associations ( $p\text{-value} < 5 \times 10^{-8}$ ) in the major histocompatibility complex (MHC) region, 200 within the autosomal non-MHC, and one in chromosome X. Overall, we can explain almost half of MS's heritability. In this study, we aimed to identify any genetic drivers of the susceptibility's sex bias in MS and develop a framework of in-depth investigation that can be applied to any common disease.

**Methods:** We leveraged data from 15 genome-wide associations studies (GWAS) that included overall 42,575 individuals: 10,966 female MS cases and 14,262 female healthy controls, and 4,266 male MS cases and 13,081 male healthy controls. We applied the same quality check pipeline and imputed all data with the most recent version of 1,000 Genomes. We meta-analyzed post-imputation probabilities across the 15 GWAS genetic variants with minor allele frequency (MAF) of at least 1% and imputation quality score of at least 0.5.

**Results:** We analyzed 8,456,964 genetic variants that fulfilled all inclusion criteria. Applying our sex-specific framework, we identified a total of 508 autosomal non-MHC and 1,445 MHC sex-specific effects ( $p\text{-value} < 10^{-4}$ ). These putatively sex-specific variants represented at least 81 autosomal non-MHC sex-specific loci, and 3 MHC loci. We further dissected our sex-specific findings to 10 categories based on magnitude and direction of effects. Most strikingly, we identified that 55 out of 81 autosomal non-MHC loci predisposed males to MS, among which 27 were protective in female, and 25 had no effect in female ( $p\text{-value} > 0.05$ ). Pathway analysis of the sex-specific variants identified 148 pathways at a  $p\text{-value} 0.05$ , the vast majority of which were immune-related pathways.

**Discussion:** We developed a framework for in-depth investigation of sex-specific genetic associations in common diseases and applied it in MS, identifying several sex dimorphic genetic variants.

## PgmNr 2873: Epilepsy and epilepsy subsyndromes: Findings from the Epi25 Consortium GWASs.

### Authors:

A.W. Smith <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, Massachusetts.; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts.

---

Genome-wide association studies (GWAS) in epilepsy and the two largest clinical subgroups - genetic generalized epilepsy (GGE) and focal epilepsy - have had only limited success in identifying risk loci. The largest GWAS in epilepsy up to date, comprising 8,696 individuals with epilepsy and 26,157 controls, identified two susceptibility loci shared by all epilepsies, one susceptibility locus for genetic generalized epilepsy, and no association signal for focal epilepsy. Here we report a new large-scale genome-wide meta-analysis in epilepsy in the largely independent Epi25 cohort, currently the largest sequencing project in the epilepsies, with 14,671 sequenced and genotyped epilepsy samples. In the European subsample of the Epi25 cohort comprising 10,497 individuals with epilepsy and 15,610 controls, we replicated the susceptibility locus for all epilepsies at 4p15.1 in proximity to *PCDH7* (rs4692500,  $P=2.2 \times 10^{-11}$ ), and a previously replicated susceptibility locus for GGE at 2p16.1 in a 706kb region encompassing *VRK2* and *FANCL* (rs2717004,  $P=2.0 \times 10^{-9}$ ). In addition, we identified three credible new susceptibility loci at 5q22.1 near *NREP* (rs112832534,  $P=4.2 \times 10^{-8}$ ), 16p12.1 near *TNRC6A* (rs2112783,  $P=1.7 \times 10^{-8}$ ), and 19q13.33 near *GLTSCR1* (rs2914006,  $P=2.2 \times 10^{-8}$ ). Subgroup analysis identified two new susceptibility loci for GGE at 2q12.1 in proximity to *TMEM182* (rs1379430,  $P=7.7 \times 10^{-12}$ ), and at 12q13.13 near *ACVR1B* (rs114131287,  $P=1.6 \times 10^{-12}$ ). Consistent to previous results, no genome-wide significant association signal could be observed for focal epilepsy. Our study provides further insights on the shared etiology of all common epilepsies, despite clinically different manifestations. We further provide additional evidence for a significant genotypic and phenotypic heterogeneity of the focal epilepsies, and a challenge for improvement of the actual clinical classification.

## PgmNr 2874: Evidence supports causal association between genetic variation in vitamin D binding and susceptibility to multiple sclerosis.

### Authors:

C. Adams <sup>1</sup>; A. Manouchehrinia <sup>2,3</sup>; N. Clark <sup>4</sup>; X. Shao <sup>1</sup>; H. Quach <sup>1</sup>; S. Roalstad <sup>5</sup>; C. Casper <sup>6</sup>; T. Lotze <sup>7</sup>; A. Waldman <sup>8</sup>; M. Gorman <sup>9</sup>; L. Krupps <sup>10</sup>; J. Graves <sup>11</sup>; T. Olsson <sup>2,3</sup>; I. Kockum <sup>2,3</sup>; C. Cotsapas <sup>12</sup>; E. Waubant <sup>11</sup>; C. Schaefer <sup>13</sup>; C. Ponting <sup>4</sup>; L. Alfredsson <sup>2,3</sup>; L.F. Barcellos <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Epidemiology, School of Public Health, UC Berkeley, Berkeley, California, USA; 2) Karolinska Institutet, Stockholm, Sweden; 3) Centre for Occupational and Environmental Medicine, Stockholm County Council, Sweden; 4) MRC Human Genetics Unit, The Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, UK; 5) Network of Pediatric Multiple Sclerosis Centers, University of Utah, Department of Pediatrics, Salt Lake City, UT, USA; 6) Department of Pediatrics, University of Utah, Salt Lake City, UT, USA; 7) The Blue Bird Circle Clinic for Multiple Sclerosis, Texas Children's Hospital, Baylor College of Medicine, Houston, TX, USA; 8) Children's Hospital of Philadelphia, Philadelphia, PA, USA; 9) Pediatric Multiple Sclerosis and Related Disorders Program at Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; 10) New York University, Pediatric MS Center, Neurology, New York, NY, USA; 11) UCSF Regional Pediatric MS Center, University of California, San Francisco, San Francisco, CA, USA; 12) Department of Neurology, Yale University, New Haven, CT, USA; 13) Kaiser Permanente Division of Research, Kaiser Permanente Northern California, Oakland, California, USA

---

There is strong evidence for a causal association between 25-hydroxyvitamin D (25OHD) serum levels and multiple sclerosis (MS; MIM 126200) susceptibility, but the mechanisms underlying these associations are unknown. It is well established that 25OHD signals through the nuclear vitamin D receptor (VDR), a ligand-regulated transcription factor. Testing for associations between VDR binding and traits can be difficult; however, single nucleotide polymorphisms (SNPs) associated with VDR binding can be used as instrumental variables for Mendelian Randomization (MR) studies. These SNPs must be independent of confounders, and can only influence the trait via the exposure. Pleiotropy can violate this last assumption and bias MR estimates. To address this, we used a weighted genetic risk score (wGRS) for 25OHD levels, a proxy for vitamin D bioavailability, as a second instrumental variable. Observed interaction between VDR binding and the wGRS would provide evidence that the association between VDR binding and MS is not entirely due to pleiotropy, and is causal. Data from an international cohort of MS cases and controls from the US and Sweden (7,881 cases, 16,069 controls) were studied. Genome-wide genetic data were subjected to QA/QC protocols prior to analysis. All participants were of European ancestry. VDR binding variants (VDR-BVs) were previously identified using ChIP-exo data from 16 calcitriol-stimulated Lymphoblastoid Cell-Lines followed by Allele-seq. VDR-BVs were regressed against ChIP-exo read counts; 305 were significant ( $p < 0.05$ ) and used as VDR binding instruments. Associations from these regressions were used as weights. 136/305 VDR-BVs were present in the data after QA/QC. The wGRS for 25OHD level was comprised of 6 SNPs identified through recent GWAS. Logistic regression was used to estimate associations between each VDR-BV instrument and MS susceptibility, and interaction between each VDR-BV instrument and

25OHD wGRS. Analyses were adjusted for genetic ancestry. Four VDR-BVs (rs17160772, rs1011817, rs4047774, and rs4413108) were independently associated with MS after multiple testing correction ( $q < 0.05$ ). There was also evidence for interaction between 4 VDR-BVs (rs9493768, rs3793786, rs7309003, and rs13098781) and 25OHD wGRS ( $p < 0.05$ ). rs17160772 is in an enhancer region for *ZC3HAV1*, a zinc finger protein that may help prevent infection by retroviruses. This study is the first to provide evidence that VDR binding contributes to MS susceptibility.

## PgmNr 2875: Genetics of multiple sclerosis disease severity.

### Authors:

S.J. Loomis<sup>1</sup>; T. Plavina<sup>1</sup>; E. Fisher<sup>1</sup>; A. Herman<sup>2</sup>; R. Rudick<sup>1</sup>; S. John<sup>1</sup>; T. Bhangale<sup>2</sup>; P. Bronson<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biogen, Cambridge, Massachusetts; 2) Genentech, South San Francisco, California

---

**Introduction:** Numerous genetic variations have been linked to Multiple Sclerosis (MS) risk, but the genetic contribution to highly variable disease progression rates and severity is undefined. Rate of brain atrophy and level of serum neurofilament light (sNfL) have been independently linked to progression rates and/or disease severity. The goal of this study is to evaluate the impact of common genetic variants on sNfL levels and rate of brain atrophy in MS patients from randomized controlled trials (RCTs).

**Methods:** We performed genome-wide association studies (GWAS) of percent brain volume change (PBVC) in each of six RCTs enrolling MS patients (N=3,389): Biogen RCTs ASCEND (secondary progressive MS [SPMS]), ADVANCE (relapsing MS [RMS]), and DECIDE (RMS); and Roche RCTs OPERA1 (RMS), OPERA2 (RMS), and ORATORIO (primary MS [PMS]). Annualized PBVC from baseline to last time point was analyzed as a continuous trait. Each GWAS was adjusted for age, sex, genetic ancestry and treatment. Mean follow-up was 2 years (range 1.7-3 years). We also performed GWAS of cross-sectional sNfL levels at baseline in two Biogen RCTs (N=974): ASCEND (SPMS) and ADVANCE (RMS), adjusting for age, sex and genetic ancestry. For each trait (PBVC and sNfL), GWAS were pooled together in a meta-analysis across RCTs.

**Results:** For PBVC, a region near *RAB4A/SPHAR/CCSAP* on chromosome 1 ( $p=7.8 \times 10^{-8}$ ) approached genome-wide significance. For sNfL levels, we found genome-wide significant loci on chromosomes 17 near *MEOX1* and 18 near *RPRD1A* ( $P_{\text{meta}}=1.7 \times 10^{-8}$ ,  $2.4 \times 10^{-9}$ , respectively). Thus far, we have not found overlap in signal between MS susceptibility, sNfL and brain atrophy.

**Discussion:** We identified several genetic regions of potential interest in two measures of MS progression, sNfL and PBVC. We will expand our sNfL GWAS by adding Roche RCTs (N=1,056) to our meta-analysis. To identify causal genes and potential mechanisms, we are conducting systematic colocalization analysis across our sNfL and PBVC results, along with other phenotypic traits (e.g. MS risk) and molecular traits (e.g. gene or protein expression). We will perform additional analysis of brain atrophy in the placebo subset (N=518) to separate treatment effect from disease severity. Our results represent a step toward using objective, quantitative traits to examine the genetics of MS progression.



## PgmNr 2876: Genome-wide association analyses of progressive supranuclear palsy identifies novel loci.

### Authors:

Y. Zhao <sup>1</sup>; E.E. Mlynarski <sup>1</sup>; W. Wang <sup>1</sup>; A. Kuzma <sup>1</sup>; L.-S. Wang <sup>1</sup>; G.D. Schellenberg <sup>1</sup>; A.C. Naj <sup>1,2</sup>; PSP Genetics Study Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 2) Department of Biostatistics, Epidemiology, and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

---

Progressive supranuclear palsy (PSP) is a rare neurodegenerative disease characterized by tau protein aggregation in the brain and clinical features including increased frequency of falls, postural instability, and supranuclear ophthalmoplegia. While this disease shares some features with other well-studied neurodegenerative diseases, including Alzheimer's and Parkinson's diseases, unlike those diseases much of the underlying genetic component of the disease remains to be identified. To identify novel common and rare variant associations, we imputed genome-wide genotyping on 2,238 PSP cases and 3,228 controls to the Haplotype Reference Consortium (HRC) r1.1 reference panel (64,976 haplotypes/39,235,157 SNPs) (MAF>0.0008) for discovery association testing using Minimac3 on the Michigan Imputation Server. For variants with MAF>0.01, logistic regression was performed using imputed genotype probabilities in PLINK. Variants with MAF≤0.01 were analyzed using score-based tests with the program SeqMeta/R. Analyses adjusted for population substructure and genotyping platform. We confirmed known genome-wide significant (GWS;  $P<5\times10^{-8}$ ) associations on chromosome 17q21.31 near *MAPT* and *KANSL1*, and the top signals correlated with associations in the H1/H2 haplotype (lead SNP rs9468, OR[95% CI]=4.96 [3.95, 6.22],  $P=9.37\times10^{-44}$ ). GWS associations were also observed at variants in *STX6* (SNP rs9468, OR[95% CI]=1.51 [1.32, 1.72],  $P=1.79\times10^{-9}$ ), which encodes the syntaxin 6 gene previously shown to be associated with PSP. Additional associations were observed at *TOM1* (lead SNP rs138760, OR[95% CI]=1.49 [1.28, 1.74],  $P=1.88\times10^{-7}$ ) and *PTPRS* (lead SNP rs73919781, OR[95% CI]=1.73 [1.41, 2.14],  $P=2.08\times10^{-7}$ ). Although these associations did not attain GWS, these genes are strong biological candidates: *TOM1*, which encodes a target of the v-myb oncogene, has recently been shown to regulate neuronal accumulation of beta amyloid oligomers in Alzheimer's Disease whereas *PTPRS*, which encodes protein tyrosine phosphatase receptor type S, has been implicated in axon guidance during embryogenesis and molecular control of adult nerve repair. Replication of these associations is ongoing. In a GWAS case-control analysis of the largest collection of PSP cases to date, we confirmed associations at several known loci and identified novel associations at several potential biological candidate genes, helping to further illuminate the complex genetics of neurodegeneration.

## PgmNr 2877: Metabolome-wide association study identifies altered cerebrospinal fluid metabolites in Alzheimer's disease.

### Authors:

D.J. Panyard <sup>1</sup>; K.M. Kim <sup>1</sup>; B.F. Darst <sup>2</sup>; Y.K. Deming <sup>3</sup>; C.D. Engelman <sup>1</sup>; Q. Lu <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Population Health Sciences, University of Wisconsin, Madison, WI.; 2) Keck School of Medicine, University of Southern California, Los Angeles, CA; 3) Wisconsin Alzheimer's Disease Research Center, School of Medicine and Public Health, University of Wisconsin, Madison, WI; 4) Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI

---

**Background:** Despite progress in understanding the etiology of Alzheimer's disease (AD), an effective treatment remains elusive. One source of novel biomarkers and potential therapeutic targets is the cerebrospinal fluid (CSF) metabolome. However, the invasive procedure used to acquire CSF samples makes large-scale, direct study of the CSF metabolome difficult. Here, we demonstrate the use of genetic prediction models trained on a smaller set of CSF samples from the Wisconsin Registry for Alzheimer's Prevention (WRAP) cohort to conduct a study of CSF metabolites in AD that leverages the sample sizes from large-scale AD GWAS consortia.

**Methods:** Data for 338 metabolites were measured in CSF samples from 136 cognitively healthy adults (ages 45-74) with genotype data in the WRAP cohort. We conducted a genome-wide association study (GWAS) to identify predictive single nucleotide polymorphisms (SNPs) for each CSF metabolite and replicated our findings in 156 cognitively healthy adults (ages 55-86) from the Wisconsin Alzheimer's Disease Research Center (W-ADRC) study.

Using the GWAS results in WRAP, we trained a polygenic score (PGS) predictive model for each CSF metabolite and assessed the performance using the W-ADRC samples. These models were used with the BADGERS method to conduct a CSF metabolome-wide association study (MWAS) for AD using three independent GWAS data sets: the International Genomics of Alzheimer's Project (N = 54,162), an independent subset of the Alzheimer's Disease Genetics Consortium (N = 7,050), and the UK Biobank (parental AD diagnosis; N = 114,564).

**Results:** At the Bonferroni-corrected significance threshold ( $P < 1.5e-4$ ), we identified 19 AD-associated CSF metabolites in the MWAS meta-analysis, of which 13 associated metabolites were lipid-related. The most strongly associated metabolite was 1-palmitoyl-2-arachidonoyl-GPC (16:020:4n6;  $P = 9.13e-9$ ), a lipid from the phosphatidylcholine pathway.

**Conclusion:** Taken together, these findings provide novel biological insights into the genetic regulation of the CSF metabolome, demonstrate the utility of genetic prediction models in studying metabolome-disease associations in rare sample sources like the CSF, and add to the growing literature implicating altered lipid metabolism in AD. We believe the CSF MWAS framework developed in our study will have wide applications in genetic studies of neurological disorders.

## PgmNr 2878: Analysis of the X chromosome for modifiers of Huntington's disease age-at-onset.

### Authors:

E.P. Hong <sup>1,2</sup>; M.J. Chao <sup>1,2</sup>; L. Jones <sup>3</sup>; P. Holmans <sup>3</sup>; S. Kwak <sup>4</sup>; M. Orth <sup>5</sup>; D.G. Monckton <sup>6</sup>; J.D. Long <sup>7</sup>; M.E. MacDonald <sup>1,2</sup>; J.F. Gusella <sup>1,8</sup>; J.M. Lee <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Neurogenetics Unit, Center for Genomic Medicine, Massachusetts General Hospital Boston MA 02114, USA; 2) Department of Neurology, Harvard Medical School, Boston MA 02115, USA; 3) Medical Research Council (MRC) Centre for Neuropsychiatric Genetics and Genomics Division of Psychological Medicine and Clinical Neurology, School of Medicine, Cardiff University, Cardiff CF14 4YS, UK; 4) CHDI Foundation, Princeton NJ 08540 USA; 5) Department of Neurology, University of Ulm, Helmholtzstraße 16, 89081 Ulm, Germany; 6) Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, UK; 7) Department of Biostatistics, College of Public Health, and Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City 52242, USA; 8) Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston MA 02115, USA

---

Huntington's disease (HD) is a dominantly inherited neurodegenerative disease, caused by an expansion of unstable CAG trinucleotide repeat in *HTT*, which encodes huntingtin. HD involves various neurological changes (such as motor, cognitive, and psychiatric), whose onset ages can be largely accounted for by CAG repeat size, but the substantial residual variance in age-at-onset of HD symptoms shows heritability. Our recent HD genome-wide association study (GWAS) revealed that age-at-onset in HD is determined by the length of the uninterrupted CAG repeat and is modified by numerous genes involved in DNA maintenance processes. Here, we performed an X chromosome-wide association study (XWAS) to supplement the GWA analysis which had focused on autosomal SNPs. We imputed SNPs on the X-chromosome for approximately 8,900 HD subjects with European ancestry using Haplotype Reference Consortium data as the reference panel, and subsequently analyzed residual age-at-onset as the phenotype to evaluate genetic association. In single SNP association analysis, no variant revealed genome-wide significant association with residual age at onset, but two loci (including Xq12) showed suggestive significances. In gene-based analysis, genes in the Xq12 region generated false discovery rate significant association, suggesting a potential role for Xq12 in modifying HD age-at-onset. Genetic association analysis of the X-chromosome is challenging due to its inactivation in females and haploidy in males. Thus, it may require a large sample size to achieve genome-wide significance in genetic association analysis. Nevertheless, our data suggest a potential role for variants on the X-chromosome in modifying the course of HD, and additional genetic analysis with increased power may reveal a new avenue for therapeutic development.

## PgmNr 2879: Colocalization analysis of genes linked to Huntington's disease age at onset and disease progression.

### Authors:

M. Ward <sup>1</sup>; A. Vishnu <sup>2</sup>; G. Hemani <sup>1</sup>; G. Davey Smith <sup>1</sup>; T. Gaunt <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MRC Integrative Epidemiology Unit, University of Bristol, UK; 2) CHDI Management/CHDI Foundation, 155 Village Boulevard, Suite 200, Princeton, NJ

---

Huntington's disease (HD) is a Mendelian disorder, however other genetic loci also influence the disease trajectory. GWAS have identified loci that modify the age at onset (AAO) and the rate of disease progression (DP) of HD. Taking a list of genes of interest derived from these GWAS findings, we performed colocalization analysis to assess which genes causally influence these HD phenotypes. Colocalization tests whether the exposure and outcome share a common genetic effect, rather than the association being due to linkage disequilibrium.

We applied three methods to test for colocalization of gene expression with AAO and DP: coloc, SMR/HEIDI and S-PrediXcan. Each method has different limitations hence we used multiple methods and compared the results. We also performed 2-sample Mendelian randomization (MR) of gene expression levels on AAO and DP. We studied gene expression in blood and brain tissues, using expression data from GTEx (whole blood, brain caudate basal ganglia and brain cortex), eQTLGen (whole blood) and Brain-eMeta (meta-analysis of brain tissues).

Our coloc results found that DP colocalized with *DHFR* in blood, caudate and cortex, and AAO colocalized with *MSH3* in blood. In our Brain-eMeta coloc results, DP and AAO both colocalized with *FAN1* but not *DHFR* or *MSH3*, perhaps due to the Brain-eMeta sample including other brain tissues. Results from SMR/HEIDI suggest that AAO colocalizes with *DHFR* in blood and DP colocalizes with *MSH3* in brain. Our S-PrediXcan results suggest that AAO colocalizes with *MSH3* in blood and DP colocalizes with *DHFR* in blood and brain and *MSH3* in brain cortex.

MR analysis of genetically predicted gene expression on AAO estimated that AAO decreases by 1.13 years (95%CI=0.53-1.74,  $p=2.5 \times 10^{-4}$ ) per unit increase in *MSH3* expression in blood (measured in Transcripts Per Million) when using eQTLs from GTEx, and by 0.78 years (95%CI=0.34-1.23,  $p=6.1 \times 10^{-4}$ ) when using eQTLs from eQTLGen. MR analysis of genetically predicted gene expression on DP estimated that the HD DP score increases by 0.20 (95%CI=0.14-0.27,  $p=4 \times 10^{-9}$ ) per unit increase in *DHFR* expression in blood (using GTEx eQTLs), by 0.18 (95%CI=0.09-0.26,  $p=3.5 \times 10^{-5}$ ) in blood (using eQTLGen eQTLs), and by 0.13 (95%CI=0.08-0.18,  $p=1.2 \times 10^{-7}$ ) in brain cortex (using GTEx eQTLs).

Overall, our results imply that *MSH3* expression colocalizes with AAO and *DHFR* expression colocalizes with DP. Increased *MSH3* expression predicts a lower AAO, and increased *DHFR* expression predicts faster DP.

## PgmNr 2880: An exon deletion polymorphism in the haptoglobin (HP) gene influences neurocognitive impairment in people with HIV infection.

### Authors:

H. Bai <sup>1,2</sup>; H. Kaur <sup>3</sup>; T. Hulgan <sup>4</sup>; S. Letendre <sup>5</sup>; D.R. Cookson <sup>5</sup>; D.R. Franklin Jr <sup>6</sup>; R. Ellis <sup>7</sup>; A.R. Kallianpur <sup>3</sup>; W.S. Bush <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Cleveland Institute for Computational Biology, Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, United States; 2) System Biology and Bioinformatics, Department of Nutrition, Case Western Reserve University School of Medicine, Cleveland, OH, United States; 3) Department of Genomic Medicine, Lerner Research Institute and Department of Medicine, Cleveland Clinic, Cleveland, OH, United States; 4) School of Medicine, Vanderbilt University, Nashville, TN, United States; 5) Department of Medicine, University of California San Diego, San Diego, CA, United States; 6) Department of Psychiatry, University of California San Diego, San Diego, CA, United States; 7) Department of Neuroscience, University of California San Diego, San Diego, CA, United States

---

Haptoglobin (HP) captures free hemoglobin and delivers it to degradation. We previously reported that higher cerebrospinal fluid (CSF) HP levels confer an increased risk of HIV-associated neurocognitive impairment (NCI). Beyond that, the *HP* gene contains a common copy number variant—deletion of two exons—that separates the two alleles: *HP1* (deletion) and *HP2* (normal). It also influences the multimerization of the encoded HP protein. These *HP* alleles are not captured by typical genome-wide arrays and their impact on NCI in people living with HIV is unknown. To assess this, we imputed the *HP* deletion in a large, observational neuroHIV study--CHARTER, using experimentally validated reference panels and IMPUTE2 in 558 European- and African-descent samples. CSF (n=202) HP levels were quantified by immunoassay. Within each subpopulation, *HP* genotypes were examined for associations with CSF HP levels and NCI, assessed by the global deficit score (GDS), with adjustments for age, sex, neuro-influential comorbidity, plasma viral load, nadir CD4+ T-cell count, and ancestry-based principal components. Increasing *HP* deletions were associated with progressively higher CSF ( $p=1.22e-5$ ) HP levels. The *HP1/HP2* heterozygous group has the lowest GDS values (multivariable-adjusted  $\beta=-0.1314$ ,  $p=0.006$ ) and a 2.13 times lower risk of NCI (defined by a  $GDS \geq 0.5$ ;  $p=0.0013$ ). Although the significance of the effects decreased when adjusted for CSF HP levels, an *HP*-by-CSF-HP interaction showed no statistically significant effect suggesting the impact is not through HP protein levels. Moreover, when involving apolipoprotein E (APOE), having one *APOE4* allele increases 0.34 in GDS ( $p=0.027$ ,  $n=202$ ) in the *HP2* homozygous group compared to *HP1* homozygous and it is predominantly driven by the African-descent population. Given prior reports that HP is involved in *beta*-amyloid aggregation and acts as an anti-oxidant of APOE, and that all three proteins co-immunoprecipitate, our findings implicate the *HP* deletion variant may result in abnormal protein processing in the brain and play a role in the development of NCI in people living with HIV. While the mechanisms remain to be clarified, the multimeric structural change of HP protein complex influenced by the *HP1* is likely to disrupt its function, potentially also altering the amyloid processing and clearance together with APOE in people

living with HIV.

## PgmNr 2881: Analysis of interaction and association of genetic variants with ischemic stroke risk in five Brazilian cities.

### Authors:

L.E. Ferreira <sup>1</sup>; R. Secolin <sup>2</sup>; I. Lopes-Cendes <sup>2</sup>; E. Santos <sup>1</sup>; N.L. Cabral <sup>1</sup>; P.H.C. França <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Universidade da Região de Joinville, Joinville, Santa Catarina, Brazil and Joinville Stroke Biobank; 2) University of Campinas – UNICAMP and Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, Brazil

---

Stroke is a severe disorder with high morbidity, mortality and disability-adjusted life years rates, which is considered a major healthcare problem and a worrisome economic burden. In Brazilian population, the adjusted annual incidence rate of IS ranges from 62 to 92 per 100,000 inhabitants, with a mixed pattern of cardiovascular risk factors, due to the high prevalence of cardiovascular diseases, dyslipidemia, hypertension, and diabetes. Risk factors have been associated with IS, including advanced age, hypertension, diabetes mellitus, smoking, and dyslipidemia. However, around 10% of IS events remain unclassified, which had reinforced to the hypothesis of a multi-factorial model disease, including genetic and small-effect environmental risk factors and even possible interactions among them. We aimed demonstrated the genetic association and genetic variants interaction for large- artery atherosclerosis and cardioembolic ischemic stroke. Were genotyped common genetic variants in individuals from five Brazilian regions: Campo Grande, Central region (774,202 inhabitants), Canoas, outh region (323,827 inhabitants), Joinville, South region (515,288 inhabitants), Sertãozinho, South-east region (101,784 inhabitants) and Sobral, North-east region (147,135 inhabitants), distributed in Large Artery Atherosclerosis Ischemic Stroke and Cardioembolic Ischemic Stroke (only for group from Joinville, SC – Brazil) and healthy controls. Genetic variants previously described in association with stroke were selected to this work: rs2107595 (NC\_000007.13:g.19049388G>A) and rs2383207 (NC\_000009.11:g.22115959A>G) for LAAS-IS, and rs879324 (NC\_000016.10:g.73034779G>A), rs1396476 (NC\_000005.10:g.60100815T>C), rs2910829 (NC\_000005.10:g.60174072G>A), rs966221 (NC\_000005.10:g.60206693A>G), rs6843082 (NC\_000004.11:g.111718067G>A), and rs152312 (NC\_000005.10:g.60491989G>A) for CE-IS. We reported a significant association of *rs2383207* genetic variant localized near *CDKN2B* gene with the susceptibility to Large Artery Vessel Ischemic stroke in Brazilian population *rs2383207*\*A allele [OR 2.35 (95% CI= 1.79 - 3.08)  $p=4.66 \times 10^{-10}$ ]. In addition, the genetics variants investigated for CE-IS, were not demonstrated significative association with CE-IS. The genetic contribution of stroke risk is not clear until, however studies of gene-gene and gene-environmental interactions may be elucidated the genetic role in pathogenesis of stroke



## PgmNr 2882: Variant prioritization by pedigree-based haplotyping: method characterization using simulation data.

### Authors:

R.A. Nafikov<sup>1</sup>; E.M. Wijsman<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 2) Department of Biostatistics, University of Washington, Seattle, WA

---

Rare genetic variants have recently been implicated in the genetic basis of complex diseases. It has been suggested that multiple rare, rather than common, variants might play a role in the etiology of complex diseases and that these rare variants could account for the “missing” heritability of diseases. Population-based association analyses are underpowered to detect disease-associated rare variants of small size effects. On the other hand, multipoint linkage analysis is one of the best approaches for localizing trait loci that segregate rare variants associated with complex diseases. The presence of large numbers of variants within linkage regions however complicates the identification of disease-associated variants. To begin to address the problem of prioritizing these variants, we developed a pedigree-based haplotyping method and applied it to a real-data situation. We assumed that the key rare variants should have originally occurred on a single risk haplotype thus reducing the number of variants under consideration to <0.25% in the region of interest.

To validate our method further, we now use a 52-member pedigree with simulated chromosomal descent patterns based on different reduced-penetrance trait models. We randomly assigned phased real WGS data to founders and, using simulated chromosomal descent patterns in the rest of the pedigree members, dropped the individual WGS alleles through the pedigree. Following identification of a linkage signal for the simulated trait, we are performing pedigree-based haplotyping to identify the risk-haplotype responsible for the linkage signal, as well as the subjects who carry vs. those who do not carry this haplotype. As before, our goal is to determine boundaries of the risk-haplotype and to identify rare variants that are unique to the risk-haplotype. The performance of our method will be characterized by accuracy of (1) detection of boundaries of a risk haplotype within the linkage region, (2) identification of subjects carrying at least one copy of the risk haplotype, and (3) determination of the phase of variants relative to the risk haplotype. Based on previous application of this method to real data, pedigree-based haplotyping with subsequent identification of a haplotype driving a linkage signal may be a useful tool for WGS variant prioritization and for follow up in other samples.



## PgmNr 2883: Parkinson's disease polygenic risk scores in a South American cohort.

### Authors:

D. Loesch<sup>1</sup>; A. Horimoto<sup>2</sup>; C. Velez-Pardo<sup>3</sup>; M. Jimenez-Del-Rio<sup>3</sup>; S. Moreno<sup>3</sup>; F. Lopera<sup>3</sup>; M. Cornejo-Olivas<sup>4,5</sup>; L. Torres<sup>5,6</sup>; M. Inca-Martinez<sup>4,7</sup>; P. Mazzetti<sup>4,6</sup>; C. Cosentino<sup>5,6</sup>; G. Arboleda<sup>8</sup>; H. Arboleda<sup>8</sup>; E. Dieguez<sup>9</sup>; V. Raggio<sup>10</sup>; V. Tumas<sup>11</sup>; V. Borges<sup>12</sup>; H. Ferraz<sup>12</sup>; C. Rieder<sup>13</sup>; A. Shumacher-Schuh<sup>14</sup>; A. Medina<sup>15</sup>; P. Chana-Cuevas<sup>16</sup>; T. Thornton<sup>17</sup>; T. O'Connor<sup>1</sup>; I. Mata<sup>7</sup>; LARGE-PD Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Genome Sciences, University of Maryland Baltimore, Baltimore, MD, USA; 2) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; 3) Neuroscience Research Group, Medical Research Institute, Faculty of Medicine, Universidad de Antioquia (UdeA), Medellín, Antioquia, Colombia; 4) Neurogenetics Research Center, Instituto Nacional de Ciencias Neurológicas, Lima, Peru; 5) Movement Disorders Unit, Instituto Nacional de Ciencias Neurológicas, Lima, Peru; 6) Universidad Nacional Mayor de San Marcos, Lima, Peru; 7) Lerner Research Institute, Genomic Medicine, Cleveland Clinic Foundation, Cleveland, OH, USA; 8) Neuroscience and Cell Death Research Group, Medical School and Genetic Institute, Universidad Nacional de Colombia, Bogotá, Colombia; 9) Neurology Institute, Universidad de la República, Montevideo, Uruguay; 10) Department of Genetics, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay; 11) Ribeirão Preto Medical School, Universidade de São Paulo, Ribeirão Preto, Brazil; 12) Movement Disorders Unit, Department of Neurology and Neurosurgery, Universidade Federal de São Paulo, São Paulo, Brazil; 13) Departamento de Neurologia, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil; 14) Serviço de Neurologia, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; 15) Universidad Nacional del Altiplano, Puno, Peru; 16) CETRAM, Facultad de ciencias Medicas, Universidad de Santiago de Chile, Chile; 17) Department of Biostatistics, University of Washington, Seattle, WA, USA

---

**Background:** Polygenic risk scores (PRS) have been shown to efficiently sum weighted risk loci identified via genome-wide association analyses (GWAS) and improve predictive models. However, until recently, GWAS have been focused on individuals of European ancestry (EA), bringing in to question the generalizability of a PRS to non-Europeans. A recent study found that a PRS constructed using EA GWAS data for Hispanics/Latinos (HL) can vary in performance (Grinde et al.). We present a PRS in a South American Parkinson's disease (PD) cohort constructed using EA summary statistics in order to determine if an EA-based PRS is associated with PD in a HL cohort.

**Methods:** PD GWAS summary statistics originate from a meta-analysis of EA cohorts (Chang et al.), with a total of 9,830 variants. We estimated local heritability using HESS with an effective sample size of 77,195. The LARGE-PD cohort consists of 808 cases and 691 controls. Subjects were genotyped using the MEGA chip with standard quality control steps. Imputation was performed using Beagle 5.0 and a custom reference panel comprised of 1000 Genomes subjects and Native Americans from the Peruvian Genome Project. The PRS was calculated using R and PLINK 1.9's clump procedure. The PRS was tested via logistic regression, adjusting for age, sex, and the first five PCs; associations with age at onset were similarly tested using linear regression. PRS parameters were evaluated using AUC,

Nagelkerke's pseudo R<sup>2</sup>, and cross-validation (CV) error.

Results: The PRS calculated from EA GWAS was significantly associated with PD status across all PRS parameters. The PRS calculated with an R<sup>2</sup> of 0.8 and a 1E-04 p-value threshold had the strongest association with PD status (5.0E-17) and an AUC of 0.721, a 7% improvement over the base model. The PRS explained 7% of the trait variance as compared to the 8.7% estimated from the summary statistics. The PRS calculated with parameters of 0.2 and 0.01 had the strongest association with age at onset (1.8E-04), suggesting differences in trait architecture.

Conclusion: Without a second independent HL cohort, we are limited to CV. Nevertheless, the results here indicate that a EA-derived PD PRS performed reasonably well. This finding suggests that the genomic architecture of PD is similar in EA versus HL cases; the extent of this similarity warrants further study. Greater diversity in PD cohorts is needed to elucidate this question and improve the predictive power of the PRS in Latinos.

## PgmNr 2884: Mutation burden and oligogenic inheritance in a large inherited axonopathy cohort.

### Authors:

D. Bis-Brewer<sup>1</sup>; F. Tao<sup>1</sup>; Z. Gan-Or<sup>3</sup>; L. Abreu<sup>1</sup>; P. Sleiman<sup>2</sup>; H. Hakonarson<sup>2</sup>; G. Rouleau<sup>3</sup>; S. Zuchner<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dr. J.T. MacDonald Department for Human Genetics, Hussman Institute for Human Genomics, University of Miami, Miami, Florida, USA.; 2) Center for Applied Genomics, the Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA.; 3) Montreal Neurological Institute and Hospital, Department of Neurology and Neurosurgery, Department of Human Genetics, McGill University, Montreal, Quebec, Canada.

---

Inherited axonopathies include the clinically distinct phenotypes, Charcot-Marie-Tooth (CMT) and Hereditary Spastic Paraplegia (HSP), which both cause slowly-progressing, length-dependent axonal degeneration. Both phenotypes are genetically and phenotypically diverse with close to 100 Mendelian genes involved for each thus far. Whole-exome sequencing of axonopathy patients may identify more than one rare variant within known disease genes. Occurrence of additional rare variation, also referred to as a 'mutation burden', has been reported in two independent CMT cohorts (n £40) supported by functional zebrafish assays. The data indicate that mutation burden may influence clinical heterogeneity and severity of disease. We sought to replicate a mutation burden across inherited axonopathies in a WES cohort 10-fold larger than the original observations (CMT cases = 357, HSP cases = 515, controls = 931). We tested the mutation burden in cases compared to controls for both non-synonymous and loss-of-function variants at ExAC MAF £0.1% and 1%. For each tested variant set, cases harbored a higher average number of qualifying variants (Mann-Whitney, p-value £0.05). The significance of this difference was further evaluated by permutating case/control status over 10,000 iterations (p-value £0.05). Next, we evaluated the possibility of di- and oligogenic inheritance within each cohort. Cases carrying a qualifying variant in <sup>3</sup>2 genes were classified as di/oligogenic and in <sup>3</sup>3 genes as oligogenic. We observed a difference in the proportion of cases and controls carrying variants for both di/oligogenic and oligogenic inheritance for non-synonymous variation (Chi-squared, p-value £0.05). Neither HSP nor CMT showed evidence of oligogenic inheritance for loss-of-function variation; however, HSP cases were enriched for digenic inheritance (Chi-squared, p-value £0.05). In this study, we provide further evidence of a mutation burden in CMT cases, demonstrate a mutation burden in HSP cases, and explore potential oligogenic inheritance patterns in a large cohort.

## PgmNr 2885: P-value combination methods for integrative meta-analysis of genome-wide associations.

### Authors:

Z. Wu <sup>1</sup>; H. Zhang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Mathematical Sciences, WPI, Worcester, Massachusetts.; 2) Biostatistics and Research Decision Sciences, Merck Research Laboratories, Rahway, NJ

---

Many summary data of genome-wide association studies are publicly available now. They contain p-values of SNPs from various genotyping platforms (array or sequencing technologies), studies, and phenotypes. The combination of these data together provides large sample size, high density SNPs, and a systematic view of multiple related phenotypes of a disease. For best mining the rich data to detect novel disease genes we desire 1) statistical tests that are optimal to various signal patterns; and 2) strategies that fully incorporate correlation information due to LD, shared samples, phenotypic similarity, etc.

We provide a general framework of combining SNP p-values by two families of tests that are optimal for addressing a broad spectrum of signal patterns. The first family, referred as the gGOF, is a generalized goodness-of-fit tests, which take the supremum of the contrasts between ordered p-values and their null expectations. The gGOF is optimal for detecting weak and sparse signals. The second family, referred as the tFisher, which is a general Fisher type cumulation of p-values with proper truncation and weighting of them. The tFisher is optimal for both sparse and dense signals. Data-adaptive omnibus tests in both families can automatically select the best test statistics for given data. We developed analytical calculation for controlling the type I error under correlations. We also studied linear transformations of data in order to increase the signal strength of true genes for more powerful detection.

To illustrate the application of these tests, we analyzed big genetic data of Osteoporosis (GEnetic Factors for OSteoporosis Consortium) and ALS (ALS Variant Server). We combines p-values and incorporate correlations at three levels 1) SNP sets; 2) meta studies; 3) phenotypes. Promising novel genes for these diseases were found and reported.

## PgmNr 2886: Transparent deep learning incorporating causal inference for genomic disease risk prediction.

### Authors:

S. Liu <sup>1,2</sup>; Z. Li <sup>2</sup>; Y. Liu <sup>2</sup>; Q. Ge <sup>2</sup>; M. Xiong <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Jiangxi Science and Technology Normal University, Nanchang; 2) The University of Texas Health Center at Houston

---

Deep learning, a rapidly resurged subfield of machine learning, outperforms many classical ML approaches and is emerging as a major analytic platform in machine learning. Deep learning with massive amounts of computational power has produced a revolution in driverless cars, speech recognition and imaging analysis, and demonstrated great potential for diagnosis and risk prediction of complex diseases. There is a growing interest in application of deep learning to healthcare and medicine. Despite its great progresses in computer vision, natural language processing, control, decision making, diagnosis and early detection of complex diseases, deep learning is also well known as a 'black box' due to its low interpretability to humans and still has a serious opacity problem. Overcoming the limitation of the lack of transparency and interpretation remains a great challenge for deep learning. In this talk, we develop a novel general framework that integrates deep learning and causal inference for disease classification. We will arrange millions of SNP and other genomic data into imaging-like two dimensional or three dimensional data structures. Therefore, all machine learning methods for imaging classification can be applied to SNPs data for disease risk prediction. The new framework for genomic disease risk prediction consists of two stages: (1) develop deep convolutional neural networks (CNN) to classify disease status based on genomic data and use occlusion map to find SNPs that are most distinctive for disease status and (2) the state-of-the-art causal inference tools to determine if the identified genetic variants are causal for complex diseases. Specifically, the convolutional neural network (CNN) model VGG (Visual Geometry Group) that won the first and the second places in the localization and classification tracks, respectively, in the ImageNet Challenge 2014 was chosen for disease risk prediction. Prediction difference analysis for visualizing the response of CNN to a specific input was used to select genetic variants for disease risk prediction. Wasserstein conditional generative adversarial networks (WCGAN) will be used to discover causal relationships between the genetic variants and diseases. The proposed algorithm is applied to the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset with 2379855 SNPs typed in 498 individuals. The results show that the intelligent algorithms substantially outperform the state-of-the-art methods.

## PgmNr 2887: Transcriptome-wide transmission disequilibrium analysis identifies novel risk genes for autism spectrum disorder.

### Authors:

K. Huang <sup>1,2</sup>; Y. Wu <sup>2</sup>; Z. Ni <sup>1</sup>; J. You <sup>3</sup>; Y. Lin <sup>3</sup>; Q. Lu <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistics, University of Wisconsin-Madison, Madison, WI; 2) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI; 3) Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI

---

Large-scale genetic studies have identified numerous point mutations and copy number variants (CNVs) implicating risk of autism spectrum disorder (ASD). However, these genetic variations are only found in a moderate proportion of ASD probands. Recent advances in consortium-scale genome-wide association studies (GWAS) have highlighted the critical involvement of common genetic variants in the etiology of ASD, but our understanding of their mechanistic roles is far from complete. Here, we introduce a novel statistical framework to quantify the transmission disequilibrium of genetically regulated transcriptomic activities from parents to offspring. We generate three pseudo-sibling controls for each proband based on phased parental genotypes, apply genetic prediction models to impute gene expression levels and alternative splicing events in each tissue, and use conditional logistic regression to identify over- or under-transmission of imputed molecular activities. We applied our method to conduct a transcriptome-wide association study (TWAS) in 11 brain tissues using 7,805 ASD proband-parent trios from the Autism Genome Project, Simons Simplex Collection, and Simons Foundation Powering Autism Research for Knowledge, and replicated our findings using independent samples from the iPSYCH project (N=35,740). We identified significant transmission disequilibrium of *POU3F2* and *MSRA* (splicing) in the discovery stage and replicated both associations in iPSYCH ( $P=2.1e-7$  and  $5.7e-9$ , respectively). Meta-analysis identified 16 additional transcriptome-wide significant associations. No associations were identified in 3,245 unaffected sibling-parent trios, suggesting a well-controlled type-I error rate in our approach. Among the identified genes, transcription factor *POU3F2* is a master regulator of a brain gene expression module associated with psychiatric disorders. *SOX7* and *NKX2-2* ( $P=4.7e-6$  and  $1.5e-10$ ) are two transcription factors involved in the regulation of embryonic development. We also identified the tau protein *MAPT* ( $P=3.6e-7$ ) and amyloid precursor protein secretase *CTSB* ( $P=1.5e-6$ ), both of which are critical genes for Alzheimer's disease. The identified genes showed minimal overlap with loss-of-function intolerant genes or known ASD genes enriched for pathogenic mutations, hinting at distinct biological processes underlying common and rare genetic variations in ASD. These results provide fundamental new insights into the genetic basis of ASD.

## PgmNr 2888: Determining the clinical epidemiology and genetic architecture of psychogenic nonepileptic seizures.

### Authors:

S.B. Goleva<sup>1, 2</sup>; K.F. Haas<sup>5</sup>; L.K. Davis<sup>1, 2, 3, 4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN.; 2) Vanderbilt Genetics Institute; Vanderbilt University Medical Center, Nashville, TN; 3) Division of Medical Genetics, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Psychiatry and Behavioral Sciences, Vanderbilt University Medical Center, Nashville, TN; 5) Department of Neurology, Vanderbilt University Medical Center, Nashville, TN

---

Psychogenic nonepileptic seizures (PNES) are characterized by episodes clinically similar to epileptic seizures. 80% of PNES patients are initially misdiagnosed with epilepsy. PNES do not arise due to aberrant brain electrical signaling but are psychology- or stress- related. PNES patients have higher rates of psychiatric disorders and obesity, and 75% of patients with PNES are females. PNES are understudied and studying PNES using electronic health records (EHR) is difficult because there is no International Classification of Diseases (ICD) code for PNES. Nothing is known about the genetic contribution to PNES. The goal of this project was to characterize the clinical and genetic characteristics of PNES in a hospital population.

We identified 4,267 PNES cases in the VU-EHR out of 2,346,808 people (prevalence 0.18%) in the Vanderbilt University EHR (VU-EHR) by developing an algorithm using convulsion ICD codes and natural language processing of keywords in patient charts. 205 of these were genotyped. We performed chart review on 50 identified cases along with a clinical neurologist.

We characterized the pattern of comorbidities associated with PNES across the lifespan. We identified VU-EHR cases of 13 common psychiatric disorders using ICD codes, and tested them for comorbidity with PNES within males and females using the Phi correlation coefficient. We then compared the likelihood of association in females vs. males using odds ratios. We found associations with PNES in a host of psychiatric disorders confirming previous reports. We found that several of these psychiatric disorders were more likely to be comorbid with PNES in females vs. males (e.g. PTSD, OR 2.72). Other phenotypes in the VU-EHR associated with PNES case/control status were determined using a Phenome Wide Association Study. Notably, cerebrovascular disease was associated with PNES (OR 1.08,  $p = 6.8 \times 10^{-19}$ ).

We determined which disorders shared underlying genetic architecture with PNES. We calculated polygenic risk score (PRS) for PNES cases ( $n = 205$ ) and controls ( $n = 48492$ ) using GWAS summary statistics for commonly comorbid psychiatric phenotypes, epilepsy, and type 2 diabetes. To determine variables' contribution to the variance in PNES case/control status, we ran a multivariable logistic regression using the PRSs. The PRS for T2D ( $p = 0.0033$ ), suicide attempts ( $p = 0.0021$ ), focal epilepsy ( $p = 0.036$ ), and insomnia ( $p = 0.037$ ) were significantly correlated with PNES case/control status.



## PgmNr 2889: Tractor: A framework for well-calibrated genomic analysis of complex traits in admixed individuals.

### Authors:

E.G. Atkinson<sup>1,2,3</sup>; A.X. Maihofer<sup>4</sup>; C.M. Nievergelt<sup>4</sup>; K.C. Koenen<sup>2,3</sup>; B.M. Neale<sup>1,2</sup>; M.J. Daly<sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Analytical and Translational Genomics Unit, Massachusetts General Hospital, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA; 4) Department of Psychiatry, University of California San Diego, La Jolla, CA; 5) Institute for Molecular Medicine Finland, Helsinki, Finland

---

Admixed individuals are routinely removed from medical genetic studies due to the paucity of methodological approaches that account for their genomic complexity such that population substructure can infiltrate analyses and bias results. Admixed populations, including African American and Latino individuals, make up more than a third of the US populace yet face severe disparities in medical research and treatment due their underrepresentation in genomic studies.

Here we present a novel analytical framework, distributed as a software package named '*Tractor*,' which precisely accounts for subtle differences in admixture at the genotype level, allowing admixed samples to be readily included alongside homogenous ones in statistical genomics efforts. Our pipeline incorporates local ancestry in addition to global, which takes into account subtle differences in individual-level admixture patterns that may differ among case and control cohorts even if their global ancestry fractions are the same. *Tractor* further leverages the information in ancestral chromosome painting to correct phase errors and recover long-range haplotypes in admixed individuals, which we find to be severely disrupted by statistical phasing.

We test our framework in simulation modeling off of African American cohorts from the Psychiatric Genomics Consortium PTSD working group as well as apply it to empirical data from African-descent individuals of the UK Biobank. We observe a significant gain in power to detect risk loci across sample sizes and disease prevalences using the *Tractor* framework, which incorporates a novel local-ancestry aware GWAS method, compared to the traditional GWAS model. We further demonstrate that this framework gives increased fine-mapping precision by leveraging the disrupted linkage disequilibrium blocks visible with ancestral chromosome painting in recently admixed groups and boosts signal to discover GWAS loci across genetic and demographic contexts.

This framework could be applied to solve the statistical issues related to admixture across many medical and population genetics activities, such as evolutionary studies using genome-wide selection scans or in the construction of polygenic risk scores. *Tractor* dramatically advances the existing methodologies for studying admixed individuals and allows for significantly better calibrated study of the genetics of complex disorders in underrepresented populations.



## PgmNr 2890: A novel statistical method to identify susceptibility genes from copy number variations.

### Authors:

M. Qiao <sup>1,2</sup>; G. Wang <sup>2</sup>; X. He <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 2) Department of Human Genetics, The University of Chicago, Chicago, IL

For over a decade, genome-wide association studies (GWAS) have been the main strategy to uncover genetic architecture of complex diseases. GWAS typically focus on single nucleotide polymorphisms (SNPs). Despite its success in a wide range of disease applications, SNPs thus discovered explain limited heritability in cognitive disorders such as Schizophrenia (SCZ). Copy number variations (CNVs), on the other hand, are believed to play a critical role in complex disease etiology. CNVs are large genomic duplication or deletion events. Previous studies have suggested that CNVs are an important source of genetic variation affecting many neurodevelopmental disorders. However, since CNVs often span multiple genes, distinguish susceptible gene(s) from other genes in the same CNV event is difficult. Results of CNV associations are thus difficult to interpret. This hurdles the use of CNV data to unravel the genetics of complex diseases.

To address the challenge we developed a new approach that exploits large-scale genome-wide CNV data in case-control studies to map genes. It is inspired by statistical fine-mapping of causal variants in linkage-disequilibrium blocks from GWAS. Unlike existing approaches that directly test for CNV associations, our method seeks to identify in CNV events true susceptible genes in a rigorous statistical framework. Genome-wide CNV data are first clustered into disjoint analysis blocks, i.e. no CNV spans between any two blocks. For genes within a block we test for disease associations while accounting for correlations among genes induced by CNV in the same block. We accomplish this by extension of a recently developed Bayesian variable selection method, SuSiE (Wang et al 2018). Our method thus selects a small number of putative risk genes among multiple correlated ones that best explain the CNV-phenotype data. Furthermore, we leverage knowledge of known biological pathways to set prior probabilities of genes in CNV events to increase power. Our model estimates posterior probabilities of all putative risk genes as well as 95% credible sets (i.e. the set of genes that cover all risk genes with high probability). Using this new approach we perform gene-level analysis in several case-control CNV datasets in SCZ. Since our method reports the statistical confidence of genes, it can be integrated with other gene-level datasets, e.g. results from exome-sequencing studies. This provides a powerful strategy to integrate data from independent sources.

## PgmNr 2891: Most rare and high-risk CNV carriers do not have major health, cognitive, or socioeconomic consequences.

### Authors:

E. Saarentaus<sup>1</sup>; N. Mars<sup>1</sup>; A. Ahola-Olli<sup>1</sup>; T. Kiiskinen<sup>1</sup>; J. Partanen<sup>1</sup>; S. Ruotsalainen<sup>1</sup>; M. Kurki<sup>1,2,3</sup>; L. Urpa<sup>1</sup>; A.S. Havulinna<sup>1,4</sup>; M. Perola<sup>4</sup>; V. Salomaa<sup>4</sup>; M. Peltonen<sup>4</sup>; J. Kaprio<sup>1,5</sup>; O. Pietiläinen<sup>2,6</sup>; M. Daly<sup>1,2,3</sup>; A. Palotie<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland; 2) Stanley Center for Psychiatric Research, The Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) National Institute for Health and Welfare, Helsinki, Finland; 5) Department of Public Health, University of Helsinki, Finland; 6) Stem Cell and Regenerative Biology, Harvard University, Cambridge, USA

---

CNVs are associated with syndromic and severe neurodevelopmental and psychiatric disorders (SNPDs), such as intellectual disability (ID), epilepsy, bipolar disorder (BD), and schizophrenia (SZ). Although considered high-impact, CNVs exist in the general population as well. This presents a diagnostic challenge in evaluating their clinical impact. To estimate the impact of CNVs to general health and well-being, we analyzed CNV burden alongside three genome-wide Polygenic Risk Scores (PRS; IQ, Educational Attainment [EA], and SZ) in a Finnish working-age population (FINRISK, n=23,053).

In carriers of high-risk CNVs (Susceptibility CNV, Risk Gene Deletion, or Large [ $>1\text{Mb}$ ] CNV), 96.1% (537/559) had no SNPD diagnosis. The remaining 3.9% (n=22) had been diagnosed with ID (4/22, 18%), Schizophrenia (4/22, 18%), Epilepsy (13/22, 59%), Bipolar Disorder (3/22, 14%), and Behavioral and Emotional Disorders (1, 5%). ID in particular was associated with Large Deletions (OR=7.2 [1.7-31]), Large Duplications (OR=5.4 [1.2-23]), ID Gene Deletions (OR=11.9 [1.5-91]), and Susceptibility CNVs (OR=13.4 [1.7-102]). In comparison, the 559 individuals with the highest PRS<sub>SZ</sub> were enriched for Schizophrenia (OR=6.1 [3.3-12]), ID (OR=2.9 [1.1-7.6]) and Bipolar Disorder (OR=2.6 [1.2-5.3]).

We hypothesized that even if CNV carriers might not have a diagnosed SNPD, some of them might have subclinical features that could be associated with their general health, well-being or socioeconomic status. In individuals without SNPD (n=22,210), we observed lower educational attainment in individuals carrying Susceptibility CNVs (1.5 years [0.4-2.6]), Large Deletions (0.7 years [0.1-1.3]), and High pLI Gene Deletions (0.6 years [0.1-1.0]). In the 559 lowest PRS<sub>EA</sub> individuals, the effect was more severe (1.9 years [1.6-2.2]) than any high-risk CNV type. The 559 lowest PRS<sub>IQ</sub> individuals had 1.2 years [0.8-1.5] lower EA.

Household income was impacted by large deletions (ordinal logit OR=0.52 [0.37-0.72]) and High pLI Gene Deletions (OR=0.69 [0.54-0.86]). Subjective Health was lower in individuals harbouring ID Gene Deletions (OR=0.37 [0.19-0.68]).

Our analyses indicate that high-risk CNVs contribute to risk for developing SNPDs. Despite this, the vast majority of working-age individuals carrying high-risk CNVs have no disease. High-risk CNVs other than large duplications are associated with lower subjective health, educational attainment, or income, yet no more than common variation as measured by PRS.

## PgmNr 2892: GWAS of alcohol consumption in the UK Biobank identifies variants that affect tissue-specific alcohol disease.

### Authors:

D. Kim <sup>1,2</sup>; X. Du <sup>2</sup>; Y. Chen <sup>2</sup>; Y.K. Li <sup>1</sup>; S. Handelman <sup>2,3</sup>; J. Mellinger <sup>2</sup>; E.K. Speliotes <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Statistical Genetics, Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA; 2) Division of Gastroenterology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA; 3) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA

---

**Introduction:** Alcohol consumption is associated with a wide range of human diseases and is a significant contributor to worldwide morbidity and mortality via alcohol use disorder (AUD), psychiatric disorders, and liver/pancreas-related sequelae. The genetic contributions to these traits is not fully known.

**Methods:** To better understand the etiology of alcohol use, we performed a genome-wide association study of self-reported alcohol intake (N=288,919) and alcohol-related sequelae (N=407,713) in > 390,000 Caucasian ancestry individuals from the United Kingdom (UK) Biobank. Genome wide significant SNPs were assessed for effects on targeted alcohol related diseases and traits with effects that reached an FDR < 5% reported.

**Results:** After conditional analyses adjusting for previously identified loci, we identified 2 novel and independent loci associated with alcohol consumption via fine mapping: rs109536-C, an expression quantitative trait locus (eQTL) variant for BRD3 and LINC00094 (B=0.018 and Pconditional=1.77'10-8) and rs13066050-T, an intergenic SNP 213kb downstream of GBE1 (B=0.022 and Pconditional=2.92'10-10). Analyses of the effects of genome-wide significant variants for alcohol-intake on alcohol related sequelae identified variants at GCKR and ADH1B that associated with AUD, delirium secondary to alcohol withdrawal, and alcoholic hepatitis and/or cirrhosis.

**Conclusions:** These results further elucidate and refine the genetic architecture of alcohol intake and its sequelae.

## PgmNr 2893: Testing gene-environment interactions without measuring the environment.

### Authors:

J. Miao<sup>1</sup>; Y. Wu<sup>1,2</sup>; K. Huang<sup>1</sup>; Z. Ni<sup>1</sup>; Q. Lu<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistics, University of Wisconsin-Madison, Madison, WI; 2) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI

---

Both genetic variations and environmental factors play key roles in shaping the etiology of complex disease. Dissecting the interplay between genetics and environmental factors may provide new insights into disease development and progression and shed light on prevention and treatment strategies. However, studies on gene-environment (GxE) interactions have only had limited success to date. This is in part due to a lack of large datasets that provide all the components required in a successful GxE study: genome-wide genetic data, robust measurements of environmental risk factors, and a sufficient number of cases for the disease of interest. We propose a novel statistical framework to test GxE interactions by using polygenic score (PGS) as a proxy for the environmental factor. PGS has quickly gained popularity in GxE research in the past few years. In fact, most recent GxE studies used PGS as the 'G' component. Through theoretical derivations and numerical analyses, we demonstrate that the inference of interaction effect remains statistically valid using our method. We use extensive simulations to show that our method provides accurate estimates for GxE effects without inflating the type-I error. In addition, our method can be applied to diverse types of GxE study designs. We applied our method to three large, independent genetic datasets for autism spectrum disorder (ASD), i.e. the Autism Genome Project (AGP; n=2,188 probands), Simons Simplex Collection (SSC; n=1,794 probands), and Simons Foundation Powering Autism Research for Knowledge (SPARK; n=3,823 probands), to investigate the interaction effect of genetic variants and birth weight on ASD risk. We used a PGS based on a GWAS conducted on the UK biobank (n=205,475) as the genetic proxy for birth weight in our analyses. Meta-analysis of three independent datasets revealed a significant negative interaction between the PGS of ASD and the PGS of birth weight ( $p=4.64E-4$ ), suggesting that a higher birth weight may buffer the genetic risk of ASD. As a negative control, we also applied the same approach to 3,243 unaffected siblings of ASD probands from SPARK and SSC cohorts and did not identify any interaction ( $p=0.65$ ). Our novel approach makes it possible to perform rigorous GxE inference on large GWAS datasets even if the environmental factor is not directly measured. We believe this method has great potential for accelerating findings in GxE research and advancing our understanding of complex disease.

## **PgmNr 2894: Model-based assessment of replicability of association signals: Application to genome-wide association meta-analysis of smoking and drinking phenotypes in 1.2 million individuals.**

### **Authors:**

D. McGuire <sup>1</sup>; Y. Jiang <sup>1</sup>; Q. Li <sup>2</sup>; S. Vrieze <sup>3</sup>; D. Liu <sup>1</sup>; GSCAN

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA; 2) Statistics Dept, Penn State University, State College, PA; 3) Psychology Department, University of Minnesota, Minneapolis, MN

---

Genome-wide association meta-analysis (GWAMA) is an effective approach to enlarge sample size and empower association studies. However, as current GWAMA often aggregate all available datasets, it becomes impossible to find a large enough independent dataset for replication. Un-replicated signals are much more likely to be false positives, and although some ad-hoc procedures may be applied to examine the validity of the signals, e.g. whether the signal is supported by multiple studies, they are often based upon heuristics and hard to be interpreted and generalized.

To overcome this limitation, we developed MAMBA for assessing the "posterior-probability-of-replicability" (PPR) for identified associations. MAMBA is a two-level mixture model which leverages both the strength and consistency of association signals between contributing studies. We model the genetic effects distribution as a mixture of normally behaving SNPs and outlier SNPs with inflated effect sizes. By fitting the model to genome-wide SNPs, the model will assign PPR to each SNP, measuring how likely the SNP has real non-zero effects. To facilitate the comparison with frequentist methods, we also developed a bootstrap procedure to calculate p-values.

We compared MAMBA with fixed effect, random effect, RE2 (a more powerful random effect method under genetic heterogeneity), and binary effect meta-analysis. We showed through extensive simulations that MAMBA is robust to the presence of outlier studies, and is almost always the most powerful or close to the most powerful even when the data is generated to favor alternative methods.

We also applied MAMBA to a meta-analysis of smoking/drinking phenotypes in 1.2 million individuals. Using 23andMe dataset as an independent replication, we observed that MAMBA yielded considerably higher p-value correlation and effect size correlation between discovery and replication estimates across all phenotypes, particularly for rare variants (e.g. for cigarettes per day phenotype, the correlation of effect sizes for rare variant association goes from .05 for fixed effect meta-analysis to .33 for MAMBA). Combining the discovery and replication datasets, we identified multiple novel rare coding variants for smoking traits with high PPR. These associations point to novel immunity or cancer related genes, including GZZM, SPOCK3, SLC7A9. We expect our method will be highly valuable for assessing the replicability of association signals in large GWAMA.

## PgmNr 2895: *In silico* enhanced GWAS for schizophrenia.

### Authors:

J. Wen <sup>1</sup>; G. Li <sup>2</sup>; J. Szatkiewicz <sup>1</sup>; P. Sullivan <sup>3,4</sup>; Y. Li <sup>1,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of North Carolina at Chapel Hill; 2) Department of Statistics and Operations Research, University of North Carolina at Chapel Hill; 3) Departments of Genetics and Psychiatry, University of North Carolina at Chapel Hill; 4) Department of Medical Epidemiology and Biostatistics and Department of (Clinical) Genetics, Karolinska Institutet; 5) Department of Biostatistics, University of North Carolina at Chapel Hill; 6) Department of Computer Science, University of North Carolina at Chapel Hill

---

Schizophrenia (SCZ), affecting ~1% of the world population, is a highly heritable, debilitating psychiatric disorder associated with high morbidity, mortality, personal and public health costs. The largest SCZ genome-wide association studies (GWAS) published to date have revealed the polygenic genetic architecture underlying SCZ, identifying >140 genome-wide significant loci with 40,000 cases and 100K controls. Taken together with earlier GWAS, these studies indisputably demonstrate the value of increased sample sizes for genetic dissection of SCZ. However, unlike many other complex diseases and traits, it is challenging to increase the sample size for SCZ GWAS by another order of magnitude. Therefore, enhanced GWAS on existing samples via computational approaches may be particularly valuable for SCZ research.

Here, we attempt to enhance power of uncovering SCZ risk loci by leveraging GWAS summary statistics from the largest published GWAS for SCZ (discovery  $n = 35476$  cases + 46839 controls), as well as brain related functional genomic information. Specifically, we train models to predict SCZ associated genetic variants via multiple statistical or computational methods including classic statistical (logistic regression), machine learning (random forest, and XGBoost) and deep learning (deep neural network) methods. Applying the trained models to GWAS summary statistics from the largest published GWAS, we performed an enhanced GWAS for SCZ and compared with our latest/unpublished results with  $n = 40675$  cases + 64643 controls (treated as the truth). Our preliminary results from chromosome 6 showed that our enhanced GWAS improved sensitivity by 20% (from 75.6% to 95.6%), at the cost of a 0.0018% false positive rate. Our enhanced GWAS also revealed 3 novel loci not detected in the large published GWAS. Our study suggests a powerful new paradigm for genetic studies of SCZ and other disorders for which necessary increases in sample sizes are difficult or expensive to attain in practice.

## PgmNr 2896: Longitudinal variant-set retrospective association test.

**Authors:**

W. Wu; Z. Wang

[View Session](#) [Add to Schedule](#)**Affiliation:** Department of Biostatistics, Yale School of Public Health, New Haven, CT.

---

Set-based tests have become popular for the identification of rare genetic variants that are associated with disease traits. Burden test and variance component test are two widely used set-level tests for single time measurement. Longitudinal repeated measures have been increasingly used in genome-wide association studies. The repeated measures provide an opportunity to study the temporal development of traits and also increase the statistical power in association tests. Most of the existing variants-set association tests are based on a population model in which ascertainment is ignored. Prospective inference with longitudinal traits and rare variants can have inflated type I error when the trait model is misspecified. Here, we propose LSRAT (Longitudinal variant-Set Retrospective Association Tests) and RSMMAT (Retrospective variant-Set Mixed Model Association Tests), two groups of retrospective variant-set tests that are constructed based on the genotype model given the phenotype and covariates. RSMMAT can be viewed as a retrospective version of the recently proposed variant-set mixed model association tests (SMMAT) and the LSRAT tests are derived under the generalized estimation equation framework. These tests have several advantages: (1) they are robust against trait model misspecification; (2) they are able to adjust both static and time-varying covariates; (3) they allow for related subjects and account for population structure; and (4) they are computationally more efficient than existing prospective approaches. Simulation studies showed that our proposed tests are robust to the trait model misspecification and gain power compared to SMMAT. We illustrated our method in the Veterans Aging Cohort Study to evaluate the association of repeated measures of cocaine use with rare variants.



## PgmNr 2897: Gene-based association testing for GWAS summary data using the general optimally weighted combination method.

### Authors:

J. Zhang <sup>1</sup>; S. Gonzales <sup>2</sup>; J. Liu <sup>1</sup>; X. Wang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Mathematics, University of North Texas, Denton, Texas; 2) Department of Biomedical Engineering, University of North Texas, Denton, Texas

---

To identify genetic susceptible variants for complex traits and disease, genome-wide association studies (GWAS) have been widely performed with immense, publicly available summary data. However, the identified genetic variants can only explain a small fraction of the overall heritability for most complex traits and disease. To explain the missing heritability for complex traits and disease, powerful statistical methods need to be developed to utilize GWAS summary data to identify novel disease susceptible genes. The power of different gene-based tests rely on the sample size and the underlying genetic architecture, which can differ in number, effect size, and effect direction of the causal variants in different genes. A statistical challenge is that, due to unknown true association patterns, there is no uniformly most powerful test to detect traits and/or disease susceptible genes; an association test may perform well for one dataset, but not necessarily for another. In this study, we propose an optimal weighted combination (OWC) test which is a general weighted combination framework utilizing GWAS summary statistics. Some popular methods including burden tests, the weighted sum of squared score (SSU), weighted sum statistic (WSS) and score test are special cases of OWC. We analytically prove that aggregating the variants in one gene is the same as using the weighted combination of Z-scores for each variant based on score test. Extensive simulation studies indicate that the type I error rates of the OWC are well controlled in different scenarios. We also numerically illustrate that our proposed test outperforms other several comparable methods through extensive simulation studies. Further, the application of the OWC and other comparison methods on two schizophrenia GWAS datasets obtained from the Psychiatric Genomics Consortium demonstrates that the new method can identify more biology plausible genes than the comparison methods. We identified that genes affecting music-related creative behavior are associated with schizophrenia. Also, our results support that schizophrenia is caused by a combination of multiple genetic factors, with each showing a relatively small effect size. Although our application was focused on schizophrenia, the proposed method is quite general and applicable to other traits and/or diseases based on either individual-level or summary GWAS data. In addition, we have made the R program of OWC freely and publicly available.

## PgmNr 2898: No causal effect of alcohol and coffee consumption on perceived stress: A Mendelian randomization study.

### Authors:

T. Nishiyama <sup>1</sup>; A. Hishida <sup>2</sup>; M. Nakatochi <sup>3</sup>; R. Fujii <sup>4</sup>; K. Wakai <sup>2</sup>; S. Suzuki <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Nagoya City University Graduate School of Medicine, Nagoya, Japan; 2) Department of Preventive Medicine, Nagoya Univ. Graduate School of Medicine, Nagoya, Japan; 3) Department of Nursing, Nagoya Univ. Graduate School of Medicine, Nagoya, Japan; 4) Department of Medical Technology, Fujita Health University, Toyoake, Japan

---

The few cohort studies that examined association between alcohol consumption and the risk of depression produced inconsistent results. Therefore, we undertook the present study to confirm causal effect of alcohol consumption on perceived stress, instead of depressive symptoms, using a Mendelian randomization (MR) approach. Perceived stress contributes to the development of depressive disorders, and substantial genetic overlap was observed between perceived stress and depressive severity: the genetic correlation between them was 0.78 in a twin study [Rietschel L 2014]. In previous studies, the variant, rs671 in ALDH2 revealed a strong genome-wide association for alcohol consumption, coffee consumption and serum uric acid (SUA) in East Asians [Jorgenson E 2017, Nakagawa-Senda H 2018, Nakatochi M 2019]. Therefore, if this variant was used for an instrumental variable, horizontal pleiotropy was a particular concern. To address this highly probable problem, we employed a multivariable MR framework, where multiple exposures can be incorporated. We conducted the present study using a general population sample of 13,618 Japanese participated in the Japan Multi-Institutional Collaborative Cohort (J-MICC) Study.

Since SUA was not associated with perceived stress in a linear regression analysis, horizontal pleiotropy via the uric acid pathway was unlikely. In previous cohort studies, coffee consumption showed a protective effect for depression. Therefore, we incorporated alcohol consumption (g/day) and coffee consumption (cups/day) as exposure variables in a multivariable MR. For the analysis, we used nine SNPs reported to be associated with alcohol or coffee consumption as instrumental variables (rs671, rs1229984, rs1957553, rs17685, rs4410790, rs12094032, rs1260326, rs6265, rs573194563) based on three analysis methods: a two-stage least-squares regression, an inverse-variance weighted (IVW) and MR-Egger methods. In the MR-Egger analysis, the intercepts were not statistically different from zero, which means that the average horizontal-pleiotropic effect across the genetic instruments were likely to be null. Then, both the two-stage least-squares regression and IVW analysis revealed no significant causal effect of alcohol consumption ( $P$ -value = 0.246 and 0.107) and coffee consumption ( $P$ -value = 0.970 and 0.716) on perceived stress, respectively. Thus, the present MR study confirmed that alcohol and coffee consumption have no influence on perceived stress.

## PgmNr 2899: A phenome-wide association study identifies the consequence of smoking derived methylation changes.

### Authors:

D. Zhou<sup>1,2</sup>; E. Gamazon<sup>1,2,3,4</sup>; N. Cox<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN; 3) Data Science Institute, Vanderbilt University Medical Center, Nashville, TN; 4) Clare Hall, University of Cambridge, Cambridge, United Kingdom

---

The tobacco epidemic is one of the biggest public health threats. As the leading cause of disease and death in the US, smoking accounting for about 1 in 5 deaths. Studies have supported that smoking affects DNA methylation, and these effects may contribute to the development and progression of various diseases. In the current study, we systematically scanned the phenotypic consequences of smoking-induced methylation changes in an EHR-based dataset BioVU (~23,000 EA samples). The methylation level and gene expression level were not directly measured in BioVU samples. Instead, we used common variants to mimic the effect of smoking derived methylation changes on gene expression. We first identified regions in which smoking significantly affected methylation status at sites that are variably methylated (Joehanes et al 2016 Circ Cardiovasc Genet). In independent data, we identified genes in which measured expression was significantly affected by methylation status at the previously identified regions (Liu et al 2013 Hum Mol Genet). For each of the genes, an expression prediction model was trained from the GTEx reference panel (Gamazon et al 2015 Nat Genet). PheWAS was performed for the predicted score aggregated from affected genes for each individual in BioVU. In brain tissues (frontal cortex BA9, anterior cingulate cortex BA24), we observed an increased risk of addiction-related phenotypes, consistent with the hypothesis that smoking-induced methylation changes may work within reward system pathways to reinforce nicotine addiction. *FCRLB* (Fc Receptor-Like B) is the leading gene for the association. Increased risk of multiple sclerosis, PTSD, primary pulmonary hypertension, etc. was also observed to be associated with the smoking-associated methylation and expression changes. Taking advantage of the wealth of traits from an EHR-linked biobank, the current study extends our understanding of the phenome-wide consequences of methylation-induced, potentially long-term consequences of smoking.

## PgmNr 2900: Inferring causal effects of neuroproteins on psychiatric disorders.

### Authors:

L. Repetto <sup>1</sup>; P. Navarro <sup>2</sup>; J.F. Wilson <sup>1,2</sup>; X. Shen <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, United Kingdom; 2) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, United Kingdom; 3) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 4) School of Life Sciences, Sun Yat-sen University, Guangzhou, China

---

Protein quantitative trait loci (pQTL) are essential tools to study the molecular basis of complex diseases, as they provide insights on the role of genetic variation in determining the protein levels that modulate the metabolic state of an individual. We quantified 184 proteins involved in neurological processes in the Olink (c) Neuro-exploratory and Neurology panels in 1070 individuals of the Orkney Complex Disease Study (ORCADES), for whom we had genotypic information (350k genotyped SNPs on the Illumina Hap300 and Omni panels, and data imputed to the Haplotype Reference Consortium panel 11 million SNPs). For each protein, we performed a genome-wide association analysis (GWAS) looking for loci associated with protein levels, both in the proximity of the protein coding gene (using gene position plus and minus 50kb in cis) and distantly (1mb in trans).

We discovered 48 cis- and 59 trans-pQTL for 95 of these neuroproteins. We performed enrichment analyses on level of expression of these proteins in different tissues, and show that the neuroproteins with cis-pQTL display enriched expression in the brain. We subsequently investigated the potential causal effect of variation in proteins levels on different psychiatric disorders, including Depression, Schizophrenia, and Bipolar Disorder for the proteins with cis-pQTL. For example, using our pQTL study and GWAS summary-level data for major depressive disorder (Wray et al, 2018), we discovered 12 additional loci associated with major depressive disorder. Two of these 12 loci we uncovered were also discovered in the latest 2019 meta-analysis for major depression, confirming our results. Further investigation of the role of the remaining 10 novel loci in depression and neural functions (including functional validation) is required, but our study highlighted novel loci with potentially relevant roles in psychiatric disorders.

## PgmNr 2901: Challenges of predicting health outcomes from gene expression levels in individuals with anorexia nervosa.

### Authors:

L.E. Blake <sup>1</sup>; J.H. Baker <sup>2,3</sup>; R. Guerra <sup>2</sup>; L.M. Thornton <sup>2</sup>; C.M. Bulik <sup>2-4</sup>; Y. Gilad <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Chicago, Chicago, Illinois.; 2) Center for Excellence in Eating Disorders, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 3) Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 4) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 5) Section of Genetic Medicine, University of Chicago, Chicago, Illinois

---

Anorexia nervosa (AN) is a heritable psychiatric disorder characterized by low body weight and a fear of gaining weight. Individuals with AN at critically low weight are often hospitalized for medical stabilization and to promote weight gain. Repeated hospitalizations are a major concern: over 20% of patients are re-hospitalized within one year. Improved prediction of clinical outcomes post-discharge are needed to facilitate earlier, more personalized interventions.

Subjective behavioral, social, and psychological factors vary in their ability to predict weight change post-discharge or rehospitalization. Furthermore, no robust, biological predictors of these outcomes exist. Our goal is to identify genes whose expression levels in blood may be predictive of weight change post-hospital discharge. To do so, we collected clinical, demographic, and biological information at admission, discharge, and 3 months post-discharge from 41 females with AN at the UNC School of Medicine.

Previous studies have focused on predicting body mass index (BMI) post-discharge. However, an individual's change in BMI over time also impacts rehospitalization. We found that BMI at discharge is strongly correlated with BMI post-discharge (Pearson's  $r = 0.60$ ,  $P < 0.01$ ), but not with rate of BMI change ( $r = 0.19$ ,  $P = 0.29$ ). Therefore, we attempted to predict both outcomes from discharge gene expression levels. For each outcome, we identified 1000 genes with the most highly associated expression levels and 8 covariates, and then trained a support vector machine. The training set comprised 29 individuals (for whom we had information at multiple timepoints post-discharge). Since an independent set was not available, the test set was comprised of the additional 12 individuals in our study. We could not reliably predict either outcome in the test set ( $R^2 < 0$ ).

This study highlights the clinical and logistical challenges of predicting treatment response in individuals hospitalized for AN. For example, further characterization of outcome—including how BMI and rate of BMI change influence rehospitalization—is needed. Additionally, future studies must collect substantially larger cohorts and include an independent replication set. Power analyses will allow us to estimate a lower bound for the number of individuals needed to predict these types of outcomes. Overall, our pilot study will guide the design of future studies predicting treatment response in individuals with AN.

## PgmNr 2902: Prediction of pathological gambling and problematic gambling behaviours by Bipolar Disorder and Parkinson's disease: A polygenic risk score analysis.

### Authors:

P.A. Lind <sup>1</sup>; A. Campos <sup>2</sup>; L. Colodro-Conde <sup>1</sup>; S.E. Medland <sup>1</sup>; W.S. Slutske <sup>3</sup>; N.G. Martin <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia; 3) Department of Psychological Sciences and Center of Excellence in Gambling Research, University of Missouri, MO

---

Increasing evidence has emerged that there is (i) an increased prevalence of problem gambling in individuals with Bipolar Disorder (BP); and (ii) a subset of Parkinson's disease (PD) patients who are at higher risk for developing pathological gambling when they undergo dopamine agonist and/or dopamine replacement therapy. We examined the extent to which individual differences in the age of gambling initiation (Initiation), gambling versatility (Versatility), DSM-IV pathological gambling (PG), and severity of problematic gambling (Severity), can be predicted by a polygenic risk scores (PRS) derived from the latest Psychiatric Genomics Consortium BP (leave out QIMR; Stahl et al 2019) and International PD Genomics Consortium (leave out 23andMe; Nalls et al 2018) analyses. Gambling data and genotyping were available for ever gamblers in the 25UP (N=1644) and GA (N=1948) studies undertaken at QIMR. In the GA study, Severity was calculated as the number of DSM-IV PG and South Oaks Gambling Screen symptoms endorsed (30-point scale; N=1948). In the 25UP study, Severity was calculated as the Problem Gambling Severity Index score (27-point scale; N=555). Initiation and Versatility (a 13-point scale of the number and types of gambling activities engaged in) were available for N=3,592. PRS were calculated using the PLINK profile score method for clumped SNPs. Linear and logistic regressions on the profile scores were performed, controlling for relatedness, ancestry, sex, age and age<sup>2</sup> at survey time, sex\*age, cohort and imputation status. Between 2.2% and 3.2% of GA subjects were diagnosed with DSM-IV PG or as probable pathological gamblers respectively. Around 7% of 25UP subjects screened positive for problem gambling. After correcting for multiple testing, genetic risk for Parkinson's disease predicted DSM-IV PG and Severity (all PRS calculated from  $p \leq 0.05$  to  $p \leq 1$ ; explaining ~0.5% of variance). There was no evidence that the genetic risk for Bipolar Disorder predicted DSM-IV PG or Severity. Genetic risk for Parkinson's disease and Bipolar Disorder did not predict Initiation or Versatility. These preliminary findings suggest a possible shared genetic aetiology between pathological gambling and Parkinson's disease. Furthermore, the development of pathological gambling risk prediction models that incorporate genetic information from future gambling GWAS could enable us to identify at-risk PD patients undergoing dopamine agonist and/or dopamine replacement therapy.

## PgmNr 2903: A new robust method uncovers significant context-specific heritability in diverse polygenic traits.

### Authors:

A. Dahl <sup>1,2</sup>; K. Nguyen <sup>2</sup>; N. Cai <sup>3,4</sup>; M. Gandal <sup>5</sup>; J. Flint <sup>6</sup>; N. Zaitlen <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neurology, University of California Los Angeles, Los Angeles, California.; 2) Department of Medicine, University of California San Francisco, San Francisco, California; 3) Wellcome Trust Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 4) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 5) Department of Psychiatry, Semel Institute, University of California Los Angeles, Los Angeles, CA; 6) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA

---

Gene-environment interaction (GxE) is a conjectured source of substantial heritability and can be fundamental in applications ranging from functional genomics to precision medicine. However, unbiased methods to profile GxE genome-wide are nascent and, as we show, cannot accommodate general environment variables, heterogeneous noise, or binary traits. To address this gap, we propose a simple, unifying mixed model for gene-environment interactions (GxEMM). Our scalable approach identifies changes in both genetic and environmental variance, and is implemented in the LDAK software package.

In simulations and theory, we show GxEMM can dramatically improve estimates and eliminate false positives when the assumptions of existing methods fail. We then apply GxEMM to a range of human and model organism data sets and find broad evidence of context-specific genetic effects. First, in a major depression (MD) cohort, we find Bonferroni-significant evidence for genetic heterogeneity across several measures of lifetime adversity; in particular, this directly implies the existence of polygenic MD subtypes. Second, we examine psychiatric disease-specific functional genomic architecture in prefrontal cortex and find substantial signal, both on average across the transcriptome (e.g. ~20% increase in heritability in schizophrenia) and at 9/63 known candidate genes. Finally, we analyze 115 traits in outbred rats and find clear evidence of sex-specific genetic and non-genetic variability in many traits at Bonferroni thresholds, including bone density and glucose tolerance. These results can help pinpoint the mechanisms underlying sexual dimorphism in complex traits. Overall, GxEMM is broadly useful for quantifying genome-wide genetic interactions, which can be useful for explaining heritability and is invaluable for determining biologically relevant environments.



## PgmNr 2904: Genome-wide causation studies of complex diseases.

### Authors:

R. Jiao <sup>1</sup>; X. Chen <sup>2</sup>; M. Xiong <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics and Data Science, School of Public Health, University of Texas Health Science Center at Houston, Houston, Texas, USA; 2) Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, USA

---

The mainstream of research in genetics and epigenetics data analysis focuses on statistical association or exploring statistical dependence between variables. Despite significant progress in dissecting the genetic architecture of complex diseases by genome-wide association studies (GWAS), the signals identified by association analysis can only explain a small proportion of the heritability of complex diseases. A large fraction of risk genetic variants is still hidden. Finding causal SNPs only by searching the set of associated SNPs may miss many causal variants. Using association analysis as a major analytical platform for the complex data analysis is a key issue that hampers the theoretic development of genomic science and its application in practice. Causality shapes how we view and understand mechanism of complex diseases.

To shift the current paradigm of genetic analysis from association analysis to causal discovery, we develop novel causal inference methods for genome-wide causal studies (GWCS). Permutation is used to calculate the P-values of the causal tests. Large simulation studies show satisfactory Type I error rates and high power of the proposed method in four scenarios: no association and no causation, having association but no causation, no association but having causation, having both association and causation. Excitingly, linkage disequilibrium has little impact on identification of causal SNPs.

The proposal methods have been applied to CATIE-MGS-SWD schizophrenia dataset with 8,421,111 common SNPs typed in 13,557 individuals for GWCS of schizophrenia. At the significance level of , 245 SNPs show causation. Among them, 62 causal SNPs can be confirmed from the literature and four of them are on the typical 108 schizophrenia-associated genetic loci (Nature, 511 (2014), pp. 421-427). We also conduct GWAS for this dataset. A total of 5,917 SNPs are associated with SCH at the significance level of and only 89 of them show causation.



## PgmNr 2905: An implementation of adaptive p-value thresholding to improve power in detecting bipolar disorder and schizophrenia associations.

### Authors:

R. Yurko <sup>1</sup>; M. G'Sell <sup>1</sup>; K. Roeder <sup>1,2</sup>; B. Devlin <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA; 2) Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA; 3) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA

---

The continual growth of “omics” technologies creates exciting opportunities to merge information from multiple studies to enhance power to understand the etiology of disease. Yet, such explorations are challenged by the need to limit detection of false positives and spurious biological conclusions, while maintaining power. Classical approaches for addressing this challenge rely on simple, albeit many hypothesis tests; recent methodologies, however, can improve power by enabling exploration of auxiliary information to weight hypotheses while retaining statistical guarantees. Such approaches can utilize prior research results and provide insights into key biological processes. We illustrate an approach, called adaptive p-value thresholding (AdaPT; Lei and Fithian, 2018) for false discovery rate (FDR) control, by analyzing bipolar disorder (BD) and schizophrenia (SCZ) genome-wide association studies (GWAS) results from the Psychiatric Genomics Consortium (PGC). Specifically, we apply AdaPT to both the BD and SCZ p-values from a set of eSNPs (gene expression single nucleotide polymorphisms) identified from the recent BrainVar study (Werling et al., 2019), which uses human dorsolateral prefrontal cortex samples across a developmental span from 176 individuals. Analyzing this set of candidate eSNPs, we demonstrate an improvement in power by exploring multiple types of information external to the specific GWAS results for the disorder: (1) due to the genetic correlation between BD and SCZ, we use independent GWAS tests from one phenotype to inform the other; (2) we explore how estimated eQTL slopes inform BD and SCZ discoveries; (3) using eSNP-gene pairs and gene-gene coexpression, we determine if module assignments from weighted gene co-expression network analysis are informative; and (4) we utilize additional SNP information gleaned from PsychENCODE. How can we use such diverse sources of information? Within the AdaPT framework, we use this auxiliary information in the context of flexible gradient boosting trees to model conditional local FDR. This analysis yields a substantial increase in the number of discoveries with each additional piece of information; for example, revealing BD and SCZ enrichment for particular gene modules. Importantly, the process underlying this analysis can be implemented for many different settings to improve power and by utilizing a variety of complementary data sources.

## PgmNr 2906: Using Mendelian randomization to evaluate the causal relationship between serum C-reactive protein levels and age-related macular degeneration.

### Authors:

Han <sup>1,2</sup>; J.S. Ong <sup>1</sup>; J. An <sup>1</sup>; A.W. Hewitt <sup>3,4</sup>; P. Gharahkhani <sup>1</sup>; S. MacGregor <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Austr, Brisbane, Queensland, Australia; 2) School of Medicine, University of Queensland, St Lucia, Brisbane, Australia.; 3) Menzies Institute for Medical Research, University of Tasmania, Australia.; 4) Centre for Eye Research Australia, University of Melbourne, Australia.

---

Serum C-reactive protein (CRP), an important inflammatory marker, has been associated with age-related macular degeneration (AMD) in some observational studies; however, the findings are inconsistent across studies. Genetic studies have the potential to shed light on this but studies to date find no evidence for the association between genetic variants in the CRP gene and AMD risk. It remains unclear whether elevated circulating CRP levels are causally related to AMD risk. We used recently released UK Biobank (UKBB) CRP data to dramatically expand our understanding of CRP genetics, allowing powerful Mendelian randomization (MR) analysis to evaluate the potential causal relationship between genetically predicted elevated serum CRP levels and the risk of AMD.

Two-sample MR analyses were performed using data from UKBB and the International AMD Genomics Consortium (IAMDGC). We derived 526 (nearly all novel) independent lead variants (genetic instruments) for serum CRP levels in 418,642 participants of European descent from UKBB. We conducted a genome-wide association study (GWAS) for 12,711 advanced AMD cases and 14,590 controls of European descent from IAMDGC. Genetic variants which predicted elevated serum CRP levels were associated with advanced AMD (odds ratio [OR] for per standard deviation [SD] increase in serum CRP levels: 1.31, 95% confidence interval: 1.19 to 1.44,  $P = 5.2 \times 10^{-8}$ ). Our results were unchanged in sensitivity analyses using MR models which make different modelling assumptions. Our findings were broadly similar across the different forms of AMD (intermediate AMD, choroidal neovascularization, and geographic atrophy). We used multivariable MR to adjust for the effects of other potential AMD risk factors including smoking, body mass index (BMI), systolic blood pressure (SBP), high-density lipoprotein cholesterol (HDL-C), and glycated haemoglobin (HbA1c); this did not alter our findings.

Our study provides strong genetic evidence that higher circulating CRP levels lead to increases in risk for all forms of AMD. This finding highlights the potential utility for using circulating CRP as a biomarker in future trials aimed at modulating AMD risk via systemic therapies.

## PgmNr 2907: Linkage analyses of Midwest Amish families uncover novel loci for age-related macular degeneration on chromosomes 8q21 and 18q21.

### Authors:

A.R. Waksmunski<sup>1,2,3</sup>; R.P. Igo, Jr.<sup>3</sup>; Y.E. Song<sup>3</sup>; J.N. Cooke Bailey<sup>2,3</sup>; R. Laux<sup>3</sup>; D. Fuzzell<sup>3</sup>; S. Fuzzell<sup>3</sup>; L.D. Adams<sup>4</sup>; L. Caywood<sup>4</sup>; M. Prough<sup>4</sup>; W.K. Scott<sup>4</sup>; M.A. Pericak-Vance<sup>4</sup>; J.L. Haines<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, Ohio; 2) Cleveland Institute for Computational Biology, Case Western Reserve University, Cleveland, Ohio; 3) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio; 4) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida

---

Age-related macular degeneration (AMD) is the most common cause of blindness in adults over 60 in the developed world. Although many common genetic risk variants are associated with AMD, a few rare variants also contribute to disease risk. The Amish have a lower genetic burden from known AMD variants than the general population, which suggests that rare variants may be contributing to their AMD risk. We genotyped 175 Amish individuals from Ohio and Indiana for the Illumina HumanExome panel, including 86 AMD cases, 77 controls, and 12 related Amish individuals with unknown AMD diagnosis. We determined that these individuals are interrelated through a 2,118-person pedigree and therefore were well-suited for discovery of novel rare genetic loci for AMD. Because analyzing large complex pedigrees for linkage is computationally intractable, we divided this pedigree into 16 sub-pedigrees and performed multipoint linkage analyses of 5,668 autosomal variants passing extensive quality control. We utilized affecteds-only dominant and recessive models (disease allele frequency = 0.10) allowing for locus heterogeneity. We found significant evidence of linkage ( $HLOD > 3.6$ ) on chromosomes 8q21 (maximum recessive  $HLOD = 4.03$ ) and 18q21 (maximum dominant  $HLOD = 3.87$ ; maximum recessive  $HLOD = 4.27$ ). These are independent of susceptibility loci identified by the International AMD Genomics Consortium and other previous AMD linkage screens. To determine whether genes in the 1-HLOD support intervals of our significant loci are functionally related to one another, we performed gene ontology (GO) enrichment analyses with ClueGO in Cytoscape. *P*-values were calculated with right-sided hypergeometric tests and corrected using the Benjamini-Hochberg method. The 1-HLOD support interval of our chromosome 8 peak ( $n = 80$  genes) is enriched (corrected  $p < 0.0005$ ) in genes involved in lipid catabolic processes and fatty acid binding, and the support interval of the chromosome 18q21 locus under the dominant model ( $n = 102$  genes) harbors genes that contribute to the positive regulation of epithelial to mesenchymal transition and endopeptidase regulator and inhibitor activities. The 1-HLOD support interval of our chromosome 18 peak under the recessive model contains 46 genes that are not overrepresented in GO terms. Our results implicate novel AMD loci in the Midwest Amish and highlight the power of investigating genetic polymorphisms for a complex trait, like AMD, in an isolated population.

## PgmNr 2908: Variation in *GSAP* identified as a novel risk factor for pigment dispersion syndrome and pigmentary glaucoma.

### Authors:

M. Simcoe<sup>1,2</sup>; N. Weisschuh<sup>3</sup>; B. Bernd Wissinger<sup>3</sup>; D. Garway-Heath<sup>4</sup>; H. Choquet<sup>5</sup>; E. Jorgenson<sup>5</sup>; P. Hysi<sup>1,2</sup>; C. Hammond<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ophthalmology, King's College London, London, London, United Kingdom; 2) Department of Ophthalmology, King's College London, UK; 3) Institute for Ophthalmic Research, Centre for Ophthalmology, University of Tuebingen, Tuebingen, Germany; 4) NIHR Biomedical Research Centre at Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; 5) Division of Research, Kaiser Permanente Northern California (KPNC), Oakland, CA, 94612, USA

---

Pigmentary glaucoma (PG) is an uncommon form of open angle glaucoma that can result in irreversible blindness. The average age of onset for PG is much younger than other forms of glaucoma, meaning it has a high burden of disease. Pigment dispersion syndrome (PDS) is a precursor and cardinal characteristic of PG. Although PG and PDS have a significant genetic component, to date only rare risk variants have been identified in familial studies. We performed a genome-wide analysis to identify genetic risk present at the population level.

We conducted separate GWAS using Firth logistic regression in two cohorts with PG cases (cohort 1: N cases=227, N controls=291; cohort 2: N cases=66, N controls=50,821), and one with PDS cases (cohort 3: N cases=137, N controls=1370), all of European ancestry. An inverse variance meta-analysis of these cohorts was just underpowered to identify genetic variants with genome-wide significant association, but did identify some strongly suggestive loci. Results from the meta-analysis were used to perform a gene-based analysis in S-Predixcan to identify association between gene expression and PDS.

After strict Bonferroni-adjustment for all gene-tissue pairs, statistically significant association was identified with *GSAP* ( $p=2.48 \times 10^{-7}$ ) which encodes the gamma-secretase activating protein. Gamma-secretase is a protease that has been shown to regulate the trafficking of TYR and TYRP1 in melanosomes. Mutations within both *TYR* and *TYRP1* are both responsible for ocular pigment defects (oculocutaneous albinism), and DBA/2J mice with *Tyrrp1* mutations present with PG. The gamma-secretase complex also eliminates the C-terminal fragment of the premelanosome protein (PMEL) and mutations within *PMEL* have been recently shown to be associated with PG in a familial study. This study identifies a novel association between *GSAP* and PDS that is present at a population level. Not only is this result compatible with previously published data, but it identifies a common pathway that accounts for the locus heterogeneity seen between familial studies.

## PgmNr 2909: GWAS-based machine learning approaches for predicting AMD progression.

### Authors:

Y. Ding <sup>1</sup>; T. Sun <sup>1</sup>; W. Chen <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biostatistics, Univ of Pittsburgh, Pittsburgh, Pennsylvania.; 2) Pediatrics, Univ of Pittsburgh, Pittsburgh, Pennsylvania.

---

Recent advances in machine learning have made extraordinary achievements in establishing flexible and powerful prediction models. The genome-wide association study (GWAS) of Age-related Macular Degeneration (AMD), a polygenic and progressive eye disease, is the first and most successful GWAS research among many diseases, where the massive GWAS data provide unprecedented opportunities to study disease risk and progression. Motivated by the need to establish a flexible and reliable prediction model for AMD progression, such as the progression-time-to-late-AMD, we developed a novel framework, which builds deep neural networks on time-to-event outcomes (referred as “NNsurv”) to effectively extract features from the wealthy GWAS data. Using data from two large randomized clinical trial on AMD progression, Age Related Eye Disease Study (AREDS) and AREDS2, we applied our method to develop and evaluate three machine-learning-based prediction models, including the NNsurv, Cox Lasso and random survival forest, to predict the risk of progression to late-AMD given the patient’s clinical and genetic profiles. The c-index on the independent test data from these three methods ranged from 0.74 to 0.77, implying satisfactory prediction performance. Time-dependent ROC curves from all prediction models were provided. Furthermore, the prediction model under NNsurv was able to classify patients into different progression risk categories based on their genetic and clinical characteristics. The result provides valuable insights for early prevention and clinical management of AMD.

## PgmNr 2910: Large-scale analysis for age-related macular degeneration reveals new risk loci.

### Authors:

R.P. Igo<sup>1,2</sup>; C.W. Halladay<sup>3</sup>; D.C. Crawford<sup>2,4,5</sup>; T. Hadi<sup>6</sup>; P. Greenberg<sup>7,8</sup>; J.M. Sullivan<sup>9</sup>; S.J. Fliesler<sup>9</sup>; S.M. Damrauer<sup>10,11</sup>; W.-C. Wu<sup>12,13</sup>; P.E. Konicki<sup>1,14</sup>; H. Choquet<sup>15</sup>; J. Yin<sup>15</sup>; K. Patasova<sup>16</sup>; E. Jorgenson<sup>15</sup>; P. Hysi<sup>16</sup>; A. Lotery<sup>17</sup>; N.S. Peachy<sup>1,18</sup>; S.K. Iyengar<sup>1,2,4</sup>; International AMD Genomics Consortium; UKBiobank Eye & Vision Consortium; VA Million Veteran Program

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Louis Stokes Cleveland VA Medical Center, Cleveland, OH; 2) Population and Quantitative Health Sciences, Case Western Reserve University, Ohio; 3) Center of Innovation in Long Term Services and Supports, Providence VAMC, Providence, RI; 4) Genetics and Genome Sciences Case Western Reserve University, Cleveland, OH; 5) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 6) Ophthalmology and Visual Sciences, Case Western Reserve University, Cleveland, OH; 7) Ophthalmology, Providence VAMC, Providence, RI; 8) Ophthalmology, Alpert Medical School, Brown University, Providence, RI; 9) Research Service, VA Western NY Healthcare System, Buffalo, NY; 10) Vascular and Endovascular Surgery, University of Pennsylvania, Philadelphia, PA; 11) Surgery, Corporal Michael Crescenz VA Medical Center, Philadelphia, PA; 12) Cardiology, Medical Service, Providence VAMC, Providence, RI; 13) Cardiology, Alpert Medical School, Brown University, Providence, RI; 14) Psychiatry, Case Western Reserve University, Cleveland, OH; 15) Division of Research, Kaiser Permanente Northern California, Oakland, CA; 16) Twin Research and Genetic Epidemiology, King's College London, London, UK; 17) Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK; 18) Cole Eye Institute and Ophthalmology, Lerner College of Medicine, Cleveland Clinic, Cleveland, OH

---

Age-related macular degeneration (AMD) is a leading cause of blindness in the elderly with multifactorial etiology. Genome-wide association studies (GWAS) have identified > 50 independent risk variants for AMD but nearly half the heritability remains to be explained. To identify additional variants, we meta-analyzed four European ancestry GWAS, totaling 40,952 cases and 247,517 controls, using samples linked to electronic health records or ocular fundus examination from the Million Veteran Program (MVP) ( $n = 157,750$ ), the International AMD Genomics Consortium (IAMGDC) ( $n = 33,976$ ), the Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort ( $n = 28,588$ ) and the UK Biobank ( $n = 68,155$ ). Genotypes were imputed to the 1000 Genomes Phase 1 or Phase 3 reference panel. Fixed-effects and random-effects meta-analysis were conducted, combining GWAS results from logistic regression, adjusting for age (except MVP), sex and principal ancestry components.

GWAS results from 10,158,774 autosomal variants with minor allele frequency  $\geq 0.004$  revealed 54 genome-wide significant (GS) loci (index  $p < 5e-08$ ), 16 of which meeting genome-wide significance for the first time. Conditional association analysis revealed 72 independent GS variants. A total of 12,711 markers with  $p < 5e-08$  included 181 nonsynonymous and two stop-gain variants, with 47 coding variations classified as deleterious by SIFT or Polyphen score. The MVP also offered the ability to examine these loci in African Americans (AA, 1,247 AMD cases, 26,720 controls) and Hispanic

Americans (HA, 544 cases, 7,425 controls). While none of these loci met genome-wide significance in the GWAS of AA and HA, notably the *CFH* and *ARMS2* loci, 54% and 78% of the index markers showed consistent effect estimates in AA and HA, respectively.

Rare-variant analysis of the MVP cohort demonstrated that coding variants in the complement system occur more frequently in Europeans, whereas mutational profiles in AA and HA are not as strong for complement but do highlight new loci. Thus, our results were dominated by deleterious variants in complement genes in European ancestry individuals, but the two new ethnic groups provided novel insights into AMD biology.

This effort is the largest multi-ethnic genetic analysis for AMD. In summary, we have revealed fundamental differences in genetic architecture across different ethnic populations of US Veterans, a finding of potential clinical significance.

## PgmNr 2911: Novel score test to increase power in association test by integrating external controls.

**Authors:**

Y. Li; S. Lee

[View Session](#) [Add to Schedule](#)**Affiliation:** Department of Biostatistics, University of Michigan, Ann Arbor, MI.

---

Recent advances in genotyping and sequencing technologies have enabled genetic association studies to leverage high-quality genotyped data to identification of variants accounting for a substantial portion of disease risk. However, identifying potentially pathogenic less common variants requires a large sample size. The usage of external controls, whose genomes have already been sequenced and are publicly available, could be a cost-effective approach to increase the power of association testing. There has been a recent effort to integrating external controls while adjusting for possible batch effects. For example, the integrating External Controls into Association Test (iECAT) method, which uses an empirical Bayesian approach to combine external controls, has been shown to improve power while controlling type I error rates. The original iECAT test, however, cannot adjust for covariates such as age, gender, etc. Hence, based on the insight of iECAT, we propose a novel score-based test that allows for covariate adjustment and constructs a shrinkage score statistic that is a weighted sum of the score statistics using exclusively internal samples and using both internal and external control samples. Similar to the original iECAT test, we assess the existence of batch effect at a variant by comparing control samples of internal and external sources. We show by simulation studies that our method has increased power over the original iECAT while controlling for type I error rates. We present the application of our method to the association studies of age-related macular degeneration (AMD) utilizing data from the International AMD Genomics Consortium (IAMDGC) and Michigan Genomics Initiative (MGI). Through incorporation of the score test approach we extend the use of iECAT to adjust for covariates and improve power, further honing the statistical methods needed to identify disease-causing variants within the human genome.



## PgmNr 2912: Meta-analysis of GWAS of the comorbidity of asthma plus eczema.

### Authors:

P. Margaritte-Jeannin<sup>1</sup>; A. Budu-Aggrey<sup>2</sup>; M. Ege<sup>3</sup>; L. Pain<sup>4</sup>; E. von Mutius<sup>3</sup>; R. Granell<sup>2</sup>; F. Demenais<sup>1</sup>; C. Laprise<sup>4</sup>; E. Bouzigon<sup>1</sup>; M.H. Dizier<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) INSERM, UMRS 1124, Université de Paris, Paris, France; 2) Medical Research Council (MRC) Integrative Epidemiology Unit, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK; 3) Dr von Hauner Children's Hospital, Ludwig Maximilian University; Comprehensive Pneumology Center Munich (CPC-M), German Center for Lung Research, Munich, Germany; 4) Département des sciences fondamentales, Centre intersectoriel en santé durable (CISD), Université du Québec à Chicoutimi, Saguenay, QC, Canada

Numerous susceptibility genes have been associated with the three most common allergic diseases (asthma, allergic rhinitis, eczema) but these genes explain only a part of the heritability. In the vast majority of genetic studies, complex phenotypes, such as the co-occurrence of at least two of these diseases in the same subjects (comorbidity), has not been considered. This may explain the missing heritability.

Our goal was to identify genetic variants specifically associated with the comorbidity of asthma plus eczema.

In a first step, we conducted a meta-analysis of four GWAS of the combined asthma plus eczema phenotype based on the Illumina 610K chip (total of 8807 European-ancestry subjects of whom 1208 subjects had both asthma and eczema). To assess whether the association with SNP(s) detected at the first step (threshold  $P=10^{-4}$ ) was specific to the asthma/eczema comorbidity, we conducted at a second step, a meta-analysis of homogeneity test of association according to disease status (presence of the two diseases "asthma plus eczema" versus the presence of only one disease "asthma only or eczema only"). On another hand, we used a joint test by combining the two test statistics from the disease-SNP association meta-analysis and the phenotypic homogeneity meta-analysis. Altogether, these analyses detected nine SNPs (of which seven were independent) showing significant association ( $P<3.10^{-7}$ , the significance threshold after correction for multiple testing).

These SNPs were located in seven genes, not previously detected for asthma or eczema. The most significant SNP ( $P=8.10^{-9}$ ) was located in *TBC1D14* associated with monocyte, leucocyte and eosinophil counts by previous GWAS. Moreover, eQTL data showed strong association between this SNP (and proxies) with *TBC1D14* expression in the blood, the skin and the esophagus mucosa.

We identified a SNP located in *TBC1D14* specifically associated with the comorbidity of asthma plus eczema. This gene is involved in both immune response and epithelial barrier mechanisms. Our study underlies the importance of studying sub-phenotypes as the comorbidity to detect new susceptibility genes.

## PgmNr 2913: Genome-wide association study of fluorescent oxidation products in asthma-ascertained families of the French EGEA study.

### Authors:

L. Orsi <sup>1,2</sup>; P. Margaritte-Jeannin <sup>3,4</sup>; M. Andrianjafimasy <sup>1,2</sup>; F. Demenais <sup>3,4</sup>; R. Matran <sup>5,6</sup>; F. Zerimech <sup>5,6</sup>; R. Nadif <sup>1,2</sup>; M.H. Dizier <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) INSERM, UMR 1168, Villejuif, France; 2) Université Versailles Saint-Quentin en Yvelines, France; 3) INSERM, UMR 1124, Paris, France; 4) Université de Paris, Paris, France; 5) CHU de Lille, EA4483, Lille, France; 6) Université Lille Nord de France, Lille, France

---

Oxidative stress is a major pathophysiological mechanism involved in asthma. Fluorescent oxidation products (FIOPs), a global and stable biomarker of damage due to oxidative stress, are of interest for epidemiological studies. Recently, they were found to be associated with expression and poor control of asthma in adults from the French Epidemiological study on the Genetics and Environment of Asthma (EGEA) (Andrianjafimasy et al, ERJ 2017). To date, no genetic study of FIOPs has yet been published.

The aim of the present study was to identify genetic loci associated with FIOPs.

We conducted a GWAS of FIOPs levels based on the Illumina 610 Quad array (Illumina, San Diego, CA) in 1216 adults (mean age 43 years old, 51% women, 37% with current asthma) from the second survey of the EGEA cohort study. FIOPs levels (geometric mean (Q1-Q3) = 92 (80-105) RFU/mL) were log-transformed and adjusted for age, sex and smoking prior to the GWAS. Association analysis between adjusted-FIOPs and SNPs used a generalized linear model to take into account the EGEA family structure; principal components to capture population ancestry were included in the model. This analysis was conducted in the whole sample and by stratifying according to asthma status (current asthma vs non-asthmatic).

In the whole sample, the two SNPs that showed the strongest associations, although not reaching the significance threshold of  $2.10^{-7}$ , were located in *BMP6* on chromosome 6 ( $P=3.10^{-6}$ ) and in *BMPER* on chromosome 7 ( $P=9.10^{-6}$ ). Interestingly, both genes were found to be strongly associated with lung function in previous GWAS and to interact with each other in a functional study. In the stratified analyses, among the top 10 SNPs showing suggestive evidence for heterogeneity according to asthma ( $P=9.10^{-6}$ ), two variants were located in *RRM2B*, a gene involved in the 'Response to oxidative stress' pathway. This gene has been previously associated with chronic obstructive pulmonary disease.

Overall, this first GWAS of FIOPs suggest promising genetic determinants previously associated with respiratory outcomes. Further replication is needed to confirm these findings.

## **PgmNr 2914: Integrated analysis of GWAS and mRNA expression array identified IFNG and CD40L as the most significant upstream-regulators in primary biliary cholangitis.**

### **Authors:**

K. Ueno <sup>1,8</sup>; Y. Aiba <sup>2</sup>; Y. Hitomi <sup>3,8</sup>; S. Shimoda <sup>4</sup>; O. Gervais <sup>5</sup>; Y. Kawai <sup>1,8</sup>; M. Kawashima <sup>6</sup>; N. Nishida <sup>1</sup>; K. Kojima <sup>7</sup>; M. Nagasaki <sup>5</sup>; K. Tokunaga <sup>1,8</sup>; M. Nakamura <sup>2,9</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Genome Medical Science Project, National Center for Global Health and Medicine (NCGM), Tokyo, Japan; 2) Clinical Research Center, National Hospital Organization (NHO) Nagasaki Medical Center, Omura, Japan; 3) Department of Microbiology, Hoshi University School of Pharmacy and Pharmaceutical Sciences, Tokyo, Japan; 4) Department of Medicine and Biosystemic Science, Kyushu University Graduate School of Medical Sciences, Fukuoka, Japan; 5) Human Biosciences Unit for the Top Global Course Center for the Promotion of Interdisciplinary Education and Research, Kyoto University, Kyoto, Japan; 6) Japan Science and Technology Agency (JST), Tokyo, Japan; 7) Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 8) Department of Human Genetics, Graduated School of Medicine, The University of Tokyo, Tokyo, Japan; 9) Department of Hepatology, Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki, Japan

---

Genome-wide association studies (GWAS) in European descents and East Asian populations have identified more than 40 disease-susceptibility genes in primary biliary cholangitis (PBC). However, the disease-pathways and their upstream-regulators remain to be elucidated. The aim of this study is to computationally identify the disease-pathways and their upstream-regulators in PBC by integrated analysis of GWAS and mRNA microarray in the Japanese population. Disease-pathways and their upstream-regulators were analyzed by Ingenuity Pathway Analysis (IPA) using the dataset 1 of GWAS (1920 PBC cases and 1770 controls) which included 261 annotated genes derived from 6760 SNPs ( $p$ -value<0.00001), and dataset 2 of mRNA microarray of liver biopsy specimens (36 PBC cases and 5 normal controls) which included 1574 genes with fold expression change>2 as compared to controls ( $p$ <0.05). Hierarchical cluster-analysis was also performed using dataset 2. There were 27 genes, 10 pathways, and 149 upstream regulators that were overlapped between the datasets 1 and 2. All of these 10 pathways were immune-related and the most significant common upstream regulators that are associated with disease-susceptibility to PBC were identified as IFNG and CD40L followed by B2M, SPIB, and FASL. The hierarchical cluster-analysis of dataset 2 revealed the presence of two distinct groups of PBC patients depending on their disease-activity, and the most significant upstream regulators that are associated with disease-activity were identified as IFNG and CD40L followed by IL6 and CC chemokine ligands and receptors. Our integrated analysis using GWAS and microarray datasets predicted that IFNG and CD40L are the most significant upstream-regulators in PBC. Further studies focusing on IFNG – and/or CD40L – pathways are needed to identify the new molecular targets for PBC patients.

## PgmNr 2915: Can you clap to the beat? Findings from the first genome-wide association study of rhythm in 606,825 individuals.

### Authors:

M. Niarchou <sup>1</sup>; J. Fah Sathirapongsasuti <sup>2</sup>; N. Jacoby <sup>3,4</sup>; J. Devin McAuley <sup>5</sup>; E. Bell <sup>4</sup>; M. Mosing <sup>6</sup>; P. Straub <sup>1</sup>; N. Creanza <sup>7</sup>; F. Ullén <sup>6</sup>; N. Cox <sup>1</sup>; D. Hinds <sup>2</sup>; L. Davis <sup>1</sup>; R. Gordon <sup>1,8</sup>; 23andMe Research Team

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt University Medical Center, Nashville, Tennessee.; 2) 23andMe, Inc; 3) Max Planck Institute for Empirical Aesthetics, Frankfurt; 4) Center for Science and Society, Columbia University; 5) Department of Psychology, Michigan State University; 6) Department of Neuroscience, Karolinska Institute, Sweden; 7) Department of Biological Sciences, Vanderbilt University; 8) Department of Otolaryngology, Vanderbilt University Medical Center

**Background:** Over the last decade moderate heritability estimates for various measures of musical ability have been established, but the human genome variations underlying such traits are not yet well-defined. Obtaining a better understanding of the biological underpinnings forming causal influence from genes to brain function to musical behavior may result in significant clinical-translational applications, in light of recently amplified interest in music and health.

**Methods:** We conducted the first genome-wide association study (GWAS) to identify common genetic variants associated with a musical rhythm phenotype, i.e. “Can you clap in time with a musical beat?”, collected from 606,825 individuals participating in research with 23andMe, Inc. We also examined whether there was a shared genetic susceptibility of rhythm with other traits with publicly available GWAS as well as whether these associations were causal.

**Results:** 68 independent Single Nucleotide Polymorphisms (SNPs) reached genome-wide significance ( $p < 5 \times 10^{-8}$ ) and the observed SNP heritability was 0.05 (s.e.=0.0002). Integration of association results with GTEx gene expression across 30 tissue types showed that SNP associations were significantly enriched in the brain tissue and pituitary gland tissue gene set. There were 31 statistically significant genetic correlations between rhythm and other traits estimated from publicly available GWAS summary statistics. For example, increased genetic predisposition towards rhythm was correlated with greater hand grip strength ( $rg(\text{left})=0.18$ ,  $se=0.02$ ,  $p=3.6 \times 10^{-16}$ ,  $rg(\text{right})=0.16$ ,  $se=0.02$ ,  $p=6.91 \times 10^{-15}$ ), faster processing speed ( $rg=-0.16$ ,  $p=3.22 \times 10^{-13}$ ) and evening chronotype ( $rg=0.09$ ,  $p=3.8 \times 10^{-5}$ ). We partitioned heritability by functional annotation and found enrichment in many functional categories including regions conserved in humans and also in functional categories involved in acetylation of histone H3 at lysine 9 (H3K9ac), monomethylation of histone H3 at lysine 4 (H3K4me3) as well as intron annotations from Refseq gene models obtained from the UCSC Genome Browser

**Discussion:** Our study is the first to examine the genetic architecture of rhythm ability. Taking into account that the biological foundations of rhythm are deeply ingrained in humans, as is language, our

study paves the way for future research into the neurobiological pathways of this universal behavioral trait.

## PgmNr 2916: Does the size matter? Telomere length distribution in Italian isolated populations.

### Authors:

M. Cocca <sup>1</sup>; M. Mezzavilla <sup>1</sup>; A. Robino <sup>1</sup>; M.P. Concas <sup>1</sup>; E. Campana <sup>2</sup>; F. Faletra <sup>1</sup>; P. Gasparini <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Maternal and Child Health, IRCCS "Burlo Garofolo", Trieste, Trieste, Italy; 2) Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy

---

Whole genome sequence (WGS) data represent a valuable source of information in terms of genetic characterisation of an individual: it can be reliably used to identify both SNP variations and Structural Variations (SV), and also provide better resolution to study Copy Number Variations (CNV). Recently, computational methods have been developed to estimate the mean telomeric length (MTL) from WGS data. MTL is a highly heritable phenotype that has been associated with a wide range of phenotypes (from lifespan to age-related diseases and cancer) and with population ancestry.

This study aims to exploit high coverage WGS data to provide information on MTL distribution in different isolated populations from the North-East Italy, where several phenotypes, including age-related ones, are available.

The 'computel' software was used to generate preliminary data on a subset of 400 samples on genome-wide MTL. MTL measurements ranges from 1,186 kb to 6,410 kb (mean=2,563 kb; sd=0,808 kb), with a negative correlation (Pearson correlation coefficient  $R=-0.31$ , p-value=  $2 \times 10^{-11}$ ) with samples' age, as expected. Moreover, on a subset of 131 samples, where sensory phenotypes are available, logistic or linear regressions (sex and age adjusted) were applied to verify the association between MTL and age-related phenotypes. The MTL resulted a significant predictor for the capacity to discriminate following odours: peppermint (p-value=0.041), lemon (p-value=0.034), rose (p-value=0.0036) and fish (p-value=0.0040). In addition, MTL showed a suggestive p-value (=0.063) when the association with smell impairment (defined as the total number of errors in odours discrimination) was tested.

Since the ageing process is strongly related to sensory decay, the study on MTL can contribute to a better understanding of this phenotype as well as relationships with other age-related traits and diseases. This approach could also help to define an "optimal" MTL length, which strikes a balance between the advantages and disadvantages of having long vs short telomeres.

## PgmNr 2917: Back to the future: Advancing disease mapping through efficient use of founder effects.

### Authors:

A.D. Paterson<sup>1,2,3</sup>; A. Dimitromanolakis<sup>1</sup>; B. Xiao<sup>1</sup>; L.J. Strug<sup>1,2</sup>; L. Sun<sup>4,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Genetics & Genomic Biology, Hosp for Sick Children, Toronto, Ontario, Canada; 2) Divisions of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Canada; 3) Divisions of Epidemiology, Dalla Lana School of Public Health, University of Toronto, Canada; 4) Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto, Toronto, Canada

---

Large biobanks coupled with technology advances allow for a new era in disease mapping through implicit shared haplotypes. Founder effects are well-known, but have been little-used for disease gene mapping. They can be used within the framework of population-based linkage analysis to identify shared genomic segments between apparently unrelated individuals. However, its broader use requires crucial progress in analytical efficiency. We have recently developed TRUFFLE (Dimitromanolakis et al., AJHG 2019) for fast and accurate identification of shared genomic segments, suitable for studies with hundreds of thousands of individuals. TRUFFLE does not need phased data, and a built-in error model corrects for segment break-ups that occur because of genotyping errors.

Rarely was information from haplotype sharing across families used to map genes in which variants cause simple Mendelian diseases, due to the low resolution of genotyping and low frequency of founder variants. As proof-of-principle for disease mapping we revisited the extent of haplotype sharing for a classic autosomal recessive Mendelian disease, cystic fibrosis, which is caused by mutations in the Cystic Fibrosis (CF) transmembrane conductance regulator (*CFTR*) and was originally fine-mapped using haplotypes. We applied TRUFFLE to Illumina 660 GWAS data from 2,436 individuals with CF from the Canadian CF Gene Modifier study where CF-causing variants were not genotyped. We varied the length of segments (from 700 kb to 5 Mb) to detect IBD segments around *CFTR* to determine the likely optimal size of shared haplotypes. At lengths of <2 Mb we see evidence for excess sharing around *CFTR*, and this becomes stronger and narrower as we reduce to 700 kb. Even at this length, we do not see excess sharing at other regions in the genome, except at HLA. Since a single common variant (rs113993960, deltaF508) accounts for the majority of disease causing variants at *CFTR*, we also examined sharing patterns in the subset of individuals with this variant. We see a stronger and narrower signal at *CFTR* when the analysis is restricted to individuals carrying the variant allele at rs113993960. As expected, other *CFTR* disease-causing variants show different lengths of shared segments due to their different ages and recombination histories. Founder effects are typically only discovered after the disease gene have been identified. The proof of principle that we present here suggests that it could enable disease gene mapping.

## PgmNr 2918: Understanding the role of gene expression in multiple tissues in stratifying blood protein levels.

**Authors:**

S. May-Wilson; E. Macdonald-Dunlop; N. Pirastu; J. Wilson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

---

More and more studies are relying on integrating transcriptomics data, in the form of eQTLs, in order to provide more power for the discovery of causal loci and provide evidence of biological function. However, despite their use in examining differential expression there is increasing evidence that eQTLs do not predict translated protein levels effectively. This is believed to be due to the many layers of control that exist between transcription and protein translation. This includes multiple splicing variants, the existence of *trans*-eQTLs, mRNA half-life, and the rate of translation which can be influenced by a number of factors. Another possibility is that the protein levels in specific tissues may be modulated by exportation of proteins from one tissue to another. To test this hypothesis we first created *cis*-expression polygenic scores (*cis*-PGS) for each gene in each tissue utilising GTEx eQTL data and verified their ability to predict circulating protein levels for approximately 1000 proteins measured in ORCADES. In order to maximise our prediction ability we first compared two methods for creating *cis*-PGS: PrediXcan and LDpred. This analysis had mixed results and shows that the two methods perform equally well. Then when comparing the predictive scores for specific tissues, the whole blood eQTLs only best predicted the blood proteins in 6% of genes suggesting that the source of most circulating proteins needs to be sought elsewhere. We then created a “multi-tissue” model by fitting a multivariable model which included all *cis*-PGS for each gene using the blood protein as the outcome and evaluating prediction through cross validation. The multi-tissue model had a higher predictive power in 54% of the 711 tested genes suggesting that in many cases the circulating protein levels are the result of the production of several different tissues. In conclusion we have shown that LDpred and PrediXcan have similar performances in creating *cis*-PGS suggesting that LDpred can be used without requiring raw data. Moreover, we have shown that in many cases using a multi-tissue approach seems to improve prediction of circulating protein levels and may at least partly explain the relatively low predictive power of *cis*-eQTLs.



## PgmNr 2919: Methods for analysis of inheritance of multiple complex diseases in families.

### Authors:

R.A. Myers<sup>1</sup>; L. Orlando<sup>1</sup>; R.R. Wu<sup>1,2</sup>; G.S. Ginsburg<sup>1</sup>; E. Hauser<sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics & Precision Medicine, Duke University, Durham, NC.; 2) Cooperative Studies Program Epidemiology Center, Durham VA Medical Center, Durham, NC; 3) Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC, USA

---

Families often suffer from a collection of common conditions, some of which may be due to shared environments and some of which due to shared genetics, i.e. pleiotropic genetic effects. In contrast to studies that seek to identify within individual pleiotropy, we are interested in identifying pleiotropic relationships within families using family health histories (FHH). Standardized collection of FHH in primary care represents a source of pedigrees well suited for exploring clusters of common complex diseases.

The web-based FHH collection tool, MeTree®, was used to collect FHH for two patient cohorts: Dataset 1 (DS1) with 1608 primary care patients for 44 conditions and Dataset 2 (DS2) with 1889 primary care patients for 91 conditions. In DS1 and DS2 respectively, patients averaged 25 and 13 relatives per family, 2.4 and 2.1 conditions for themselves, and 0.78 and 1.12 conditions per relative. We estimated the risk to first-degree relatives (FDR) for pairs of conditions by modeling disease A in the FDRs (within family counts/proportions) with disease B in the proband using nonparametric rank correlation (NP) and two parametric models: Bayesian generalized linear models with logit (BLGLM) and Poisson (BPGLM) links, adjusted for family size, age, sex, and race. The BPGLM and BLGLM results were concordant in DS1 but not in DS2. When disease A (FDRs) and disease B (proband) are the same, we expect significant positive associations in disease presence, due to heritability. In the DS1 and DS2 BPGLM results, respectively, 40 of 44 and 84 of 91 diseases were significantly, positively associated. In the DS1 and DS2 NP results, respectively, 43 of 44 and 90 of 91 diseases were significant. In the disease combinations for which disease A and B were different, there were 59 of 2068 and 194 of 8645 significant disease pairs in the BPGLM results. The NP findings had a number of significant discordant disease pairs (55 of 2068 and 272 of 8645, DS1, DS2, respectively). Within disease associations were detected by all methods, however the higher number of diseases detected by the NP approach may be due confounding (e.g. age, race); the two parametric approaches yielded consistent results in DS1 but not DS2. These suggest that FHH clustering analysis is robust to method selection for high heritability conditions, is sensitive to method selection for lower heritability conditions and/or noise, and can be used to study familial pleiotropy in high heritability conditions.

## PgmNr 2920: Modifying sample splitting in ComPaSS-GWAS for improved performance on rare SNP applications.

**Authors:**

J.A. Sabourin; A.F. Wilson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Genometrics Section, Computational and Statistical Genomics Branch, NHGRI, NIH, Baltimore, MD

---

ComPaSS-GWAS [Sabourin et. al., 2019] is a method that uses repeated sample splitting to identify SNPs that are corroborated across random splits as an approximation to replication when appropriate replication data are not available. However, for rare or very rare SNPs, sample splitting can by chance, leave too few (or no) subjects with the minor allele in one of the sample splits to allow for corroboration, reducing the SNPs score. This reduces both the type I error rate and power to detect rare SNPs with ComPaSS-GWAS. In this study we present a modification to ComPaSS-GWAS, ComPaSS-rare, that replaces sample splitting with the use of complementary random weights, motivated by the weighted likelihood bootstrap. In ComPaSS-rare, the two halves determined by complementary weights are not independent and the theoretical squaring of  $\alpha$  can't be applied, so the parameter value ( $\alpha$ ) for ComPaSS-rare must be appropriately adjusted.

Two simulation studies were performed. In the first, the rare variant modification (ComPaSS-rare) was applied to the common variant simulation model from the previous study [Sabourin et al 2019] which was based on a typical GWAS on SNPs with  $MAF \geq 0.01$ ; causal SNPs had MAFs of 0.05. In the simulations for common variants, there was consistent performance for type I error and power for traditional GWAS, ComPaSS-GWAS and ComPaSS-rare, suggesting that the modification for rare SNPs does not significantly change the performance of the method for common SNPs. The second simulation was based on the GAW19 whole exome sequencing data that focused on rare SNPs with MAF between 0.0025 and 0.05 with causal SNPs having MAFs of 0.005. Under this model, the traditional GWAS with critical value of  $5 \times 10^{-8}$  showed inflated type I error rates on SNPs with  $MAF < 0.01$  and ComPaSS-GWAS had reduced power and type I error rates for rarer SNPs due to sample splitting issues. However, ComPaSS-rare had type I error and power consistent with the traditional GWAS over all MAF ranges. Based on these results we believe the modified procedure, ComPaSS-rare, is appropriate for analysis of both common and rare SNPs, and has improved performance for analysis of rare SNPs that are more prevalent in association studies with sequencing data.

## PgmNr 2921: Mediation analysis of high-dimensional phenotypes.

### Authors:

C.R. Solis-Lemus <sup>1</sup>; A.M. Holleman <sup>2,3</sup>; A. Todor <sup>1,3</sup>; L.M. Almli <sup>4</sup>; B. Bradley <sup>4,5</sup>; K.J. Ressler <sup>6</sup>; D. Ghosh <sup>7</sup>; M.P. Epstein <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University, Atlanta, GA; 2) Department of Epidemiology, Emory University, Atlanta, GA; 3) Center for Computational and Quantitative Genetics, Emory University, Atlanta, GA; 4) Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA; 5) Mental Health Service Line, Department of Veterans Affairs Medical Center, Atlanta, GA; 6) Department of Psychiatry, McLean Hospital, Harvard University, Belmont, MA; 7) Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO

---

A variety of statistical methods exist for testing global association between a genetic variant (or set of variants) and a collection of correlated phenotypes, in order to leverage pleiotropy for more powerful gene mapping. However, if such a test is significant, there is subsequent interest in assessing whether the cross-phenotype effect is due to biological pleiotropy or mediation pleiotropy. Given a significant cross-phenotype association between a SNP and two phenotypes (P1 and P2), biological pleiotropy implies that the SNP has direct effects on both P1 and P2. Mediation, on the other hand, considers the possibility that P1 lies in the pathway between the SNP and P2 so that part of the association between SNP and P2 can be indirectly explained by the direct relationship with P1. Here, we propose novel kernel-based methodology to identify whether an observed association between a variant (or a group of variants within a gene) and possibly high-dimensional phenotype data is completely or partially mediated by a different set of high-dimensional phenotypes. Our mediation-based approach is novel in that it can handle high-dimensional genetic data, mediators, and outcome variables. The method further adjusts for covariates. We show that our method controls for type I error and is powerful to detect a variety of mediation scenarios in simulated data. We further apply our mediation technique to multivariate symptom data from the Beck Depression Inventory (BDI) that were collected as part of the Grady Trauma Project GWAS. We show that initial cross-phenotype associations between a candidate risk locus and the 21-item multivariate BDI outcome can be partitioned into 3 symptoms directly associated with the locus while the remaining 18 symptoms are only indirectly associated.

## PgmNr 2922: A cross-trait genetic risk prediction framework leveraging biobank-scale GWAS summary statistics.

### Authors:

Z. Zhao <sup>1</sup>; Y. Hu <sup>4</sup>; Q. Lu <sup>1,3</sup>; H. Zhao <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison School of Medicine and Public Health, Madison, WI; 2) Department of Biostatistics, Yale School of Public Health, New Haven, CT; 3) Department of Statistics, University of Wisconsin-Madison, Madison, WI; 4) The Voleon Group

---

Polygenic risk score (PRS) methods have quickly gained popularity in human genetics research due to its simplicity and robustness. However, despite recent advances in the estimation of variant effects and variant prioritization using biological annotation information, PRSs for most complex traits remain only moderately predictive. It has been demonstrated that jointly modeling multiple traits could effectively improve the predictive power of PRS. However, existing methods can be applied to only a small number of traits. Therefore, summary statistics from large-scale biobanks cannot be effectively utilized. Here, we introduce a principled framework to jointly model thousands of GWASs from large biobanks to improve the prediction accuracy of PRS. We adaptively prioritize thousands of traits in the UK biobank based on their genetic correlations with the phenotype of interest, jointly model each selected trait and the main phenotype to re-estimate variant effects, and compute a re-weighted PRS that combines information across many genetically correlated traits. Through extensive simulations, we demonstrate that our method provides significantly increased prediction accuracy compared to state-of-the-art single- and multi-trait PRS models, especially under sparse genetic architecture. We applied this framework to breast cancer (BC), celiac disease (CEL), Crohn's disease (CD), rheumatoid arthritis (RA), and type-II diabetes (T2D), and evaluated the predictive performance of our method using external cohorts independent from the input GWAS. For each disease, we integrate the GWAS summary statistics with more than 4,000 traits from the UK biobank. We show that our method substantially outperforms single-trait methods including standard PRS with fine-tuned p-value cutoffs and LDpred, and multi-trait methods such as PleioPred. Furthermore, we illustrate that weighing PRSs based on the level of genetic similarity could further strengthen predictive power. In particular, as we adopt a more stringent trait-selection criterion based on genetic correlation, our method produces consistently more predictive risk scores for multiple diseases. Finally, our method does not include tuning parameters and is computationally efficient. We believe this method could have wide applications in genetic prediction applications.

## PgmNr 2923: Genome-wide mapping for palpebral fissure inclination.

### Authors:

T. Zou <sup>1</sup>; M. Lee <sup>2</sup>; L. Souza <sup>1</sup>; C. Skinner <sup>1</sup>; A. Azevedo <sup>1</sup>; M. Marazita <sup>1,2</sup>; S. Weinberg <sup>1,2,3</sup>; J. Shaffer <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human genetics, University of Pittsburgh, Pittsburgh, PA.; 2) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA; 3) Department of Anthropology, University of Pittsburgh, Pittsburgh, PA

---

Palpebral fissure inclination (PFI; i.e., eye inclination), defined as the angle between the line connecting the medial and lateral canthi and the horizontal plane, varies remarkably in humans based on age, gender, and ethnicity. Additionally, genetic disorders often present with extreme PFI. However, there is limited information about the genetic basis of normal PFI variation. We aimed to identify genetic variants for three PFI measures (right eye angle, left eye angle, and combined angle) in a sample of 2210 healthy individuals of European ancestry using three complementary genetic approaches: (1) Genome-wide association study (GWAS), testing common variants (MAF > 5%); (2) Genome-wide gene-based association analysis of low-frequency variants (MAF < 5%) in aggregate; and (3) *In silico* candidate gene study based on the published literature of genetic disorders that often present with extreme PFI (e.g., CHARGE syndrome, Glass syndrome) and prioritizing putatively functional variants. Participants were genotyped on the Illumina (San Diego, CA) HumanOmniExpress+Exsome v1.2 array. No genome-wide significant associations were observed ( $p < 5E-8$ ). However, several suggestive signals ( $p < 1E-5$ ) were observed near genes with potential biological roles in facial morphology, such as *TENM2*, *CHD9*, *SLC23A2*, *CNTN1*. Results for low-frequency variants and candidate genes will be presented. Understanding the genetic architecture of PFI may yield insights into the mechanisms impacting facial morphology and aid in the assessment of facial dysmorphology in the clinical setting.

## PgmNr 2924: A SVM-based ancestry inference method for large-scale genetic data.

**Authors:**

Z. Zhu; A. Manichaikul; C.C. Robertson; S. Onengut-Gumuscu; S.S. Rich; W.-M. Chen

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia.

---

Ancestry inference in genetic studies has been routinely performed for the purpose of quality control and reduced population stratification effects in association analyses. We present our integrative implementation of a support-vector-machine (SVM)-based method to identify the most likely ancestral group(s) for an individual by leveraging known ancestry in a reference dataset (e.g., the 1000 Genomes Project data). Our method involves first projecting each study sample to the principal component (PC) space of a reference dataset, followed by training and classifying the ancestry of each study sample using an SVM algorithm. This algorithm has been well integrated in the computationally efficient tool, KING. To evaluate the performance of our ancestry inference method, we conducted a simulation study using simuPOP, a forward-time simulation tool. For 9 simulated HapMap populations, our SVM-based tool had 100% accuracy in classifying ancestry. To further investigate the performance of our SVM-based ancestry inference method as implemented in KING, compared to peddy, another SVM-based method, we analyzed 13,181 participants who were genotyped with the Illumina HumanCoreExome array (with GWAS backbone) as well as the Illumina ImmunoChip (a custom fine-mapping array covering 186 loci). The inferred ancestry is identical in 95.4% (N=12,574) of subjects across 3 different analyses: 1) HumanCoreExome array data with KING; 2) ImmunoChip array data with KING; and 3) HumanCoreExome data with peddy. The consistency in our ancestry classification suggests that the example data consists of EUR (78.6%), AFR (9.3%), AMR (6.1%), SAS (3.4%), and EAS (2.6%) ancestry samples. In comparison, the ancestry inference from ImmunoChip array data using peddy is substantially less reliable than the three other inference approaches: 15.3% of these 12,574 subjects are inferred as unknown, and another 16.4% are inferred as AMR, instead of EUR as consistently inferred by all three other inferences. This analysis demonstrates that the KING implementation of the SVM-based ancestry inference method can provide accurate and consistent inference across different genotyping platforms. Further, the implementation of the approach in KING is orders of magnitude faster than existing methods and is scalable to large datasets with sample size beyond one million subjects (17 minutes to classify ancestry for all subjects in a dataset consisting of 1 million samples on our server with 96 CPU cores).

# PgmNr 2925: The Polygenic Score (PGS) Catalog: a database of published PGS to enable reproducibility and uniform evaluation.

## Authors:

S.A. Lambert<sup>1,2,3,4</sup>; G. Abraham<sup>1,5</sup>; H. Parkinson<sup>3,4</sup>; J. Danesh<sup>2,3,6,7,8,9</sup>; J.A.L. MacArthur<sup>4</sup>; M. Inouye<sup>1,2,3,5,10,11</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Cambridge Baker Systems Genomics Initiative, Cambridge, UK, and Melbourne, Victoria, Australia; 2) MRC/BHF Cardiovascular Epidemiology Unit, Dept of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 3) Cambridge Substantive Site, Health Data Research UK, Wellcome Genome Campus, Hinxton, UK; 4) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK; 5) Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia; 6) NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Dept of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 7) National Institute for Health Research Cambridge Biomedical Research Centre at the Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; 8) British Heart Foundation Cambridge Centre of Excellence, Dept of Clinical Medicine, University of Cambridge, Cambridge, UK; 9) Wellcome Sanger Institute, Hinxton, UK; 10) The Alan Turing Institute, London, UK; 11) Dept of Clinical Pathology and School of BioSciences, University of Melbourne, Parkville, Victoria, Australia

---

**Introduction:** Polygenic scores (PGS) aggregate the effects of hundreds-to-millions of variants across the genome into a single score and have been shown to have predictive value for many common diseases and traits. Studies have shown that PGS (e.g. for Coronary Artery Disease; CAD) can provide independent risk stratification when included into models with traditional risk factors (age, smoking, cholesterol, etc.) and family history. PGS are actively being translated into the clinic; however, adoption has been hindered by inadequate annotation and availability, variable or incomplete benchmarking of multiple PGS for the same disease, and variable performance in diverse populations.

**Objective:** A complete understanding of each PGS and its predictive value is necessary for its application; however, no central resource of consistently annotated PGS exists. In collaboration with the NHGRI-EBI GWAS Catalog, we are developing the PGS Catalog to increase the transparency and reproducibility of PGS by distributing the information necessary for both research and translational uses of PGS for a wide range of diseases and traits.

**Methods:** Initially, we have focused on cataloging PGS that have been developed for five well-studied traits (CAD, diabetes (types 1 and 2), obesity/BMI, breast cancer, and Alzheimer's) based on their potential clinical utility and public health burden of the disease.

To populate the PGS Catalog we reviewed the literature for purpose-built PGS developed after 2010. We have defined and extracted metadata describing each PGS (e.g. SNP rsID, effect allele, effect size), as well as how it was developed and applied. This includes direct links to the GWAS Catalog study (sample size, ancestry, cohorts) used to develop the score, characteristics of the training and

validation cohorts, and reported accuracy of the PGS.

**Results & Future Directions:** We have developed a framework to consistently summarise and represent PGS, to be housed in a publicly accessible database. The pilot PGS Catalog currently contains over 100 consistently extracted PGS for our five disease areas. We seek community feedback and are inviting PGS researchers to submit their PGS to increase coverage and maximize the utility of their scores. With the PGS Catalog, we intend to provide a complete picture of the predictive accuracy of all available PGS in external population cohorts and distribute directly comparable performance metrics between scores as part of the complete Catalog.



## PgmNr 2926: Whole exome sequences reveal rare and common variants associated with 1102 plasma proteins.

### Authors:

L. Klaric <sup>1</sup>; T.S. Boutin <sup>1</sup>; C.V. Van Hout <sup>2</sup>; N. Gosalia <sup>2</sup>; A. Shuldiner <sup>2</sup>; C. Haley <sup>1</sup>; J.F. Wilson <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom; 2) Regeneron Genetics Center, Tarrytown, NY, USA; 3) Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, UK

---

Proteins are major drivers of all biological processes. Blood plasma proteins reflect the proteomes of a wide variety of cells, providing an insight into physiological processes occurring in diverse biological systems. Therefore, they are potentially useful biomarkers of various diseases. Combining proteomics with protein-altering variation captured by whole exome sequencing provides an opportunity for a more direct insight into mechanisms behind complex traits and diseases.

We performed exome-wide association analysis of 1,102 proteins measured using OLINK proximity extension assays in plasma from 1059 individuals from the genetically isolated ORCADES cohort (Scotland) and whole exome sequencing (WES) generated by the Regeneron Genetics Center. We found 3,545 significant ( $p \leq 4.5 \times 10^{-11}$ ) associations between 374 proteins and 968 genes mapping to 377 genomic regions, where a region was defined based on all significant associations within 1Mb from the strongest association in the region. 5% of associations were with low frequency variants (MAF<0.05). Of the 374 proteins, 237 (63%) had *cis* associations only, 88 (24%) *trans* only and 49 were associated with two loci - 26 (7%) *cis* and *trans*, 23 (6%) *trans*. Overall, 171 sentinel variants are missense, with 5 resulting in gained or lost stop or start codons. 136 of our sentinel pQTL variants were in high LD ( $r^2 \geq 0.8$ ) with disease-associated variants. Association of very rare variants was assessed using aggregate tests and revealed further loci influencing the proteome.

Given the direct role of proteins in various biological processes and more interpretable functional consequence of WES variants, genetic associations of a large number of proteins and protein-altering variants provide not only a basis for understanding mechanisms behind complex traits and diseases, but also provide a framework for assessment of causality of individual biomarkers.

# PgmNr 2927: A method for rare variant burden analysis using case-only exome data.

## Authors:

Y. Na <sup>1</sup>; H. Rasouly <sup>1</sup>; A. Gharavi <sup>1</sup>; I. Ionita-Laza <sup>2</sup>

[View Session](#)
[Add to Schedule](#)

## Affiliations:

1) Medicine, Columbia University, New York, New York.; 2) Biostatistics, Columbia University, New York, New York.

---

## Background:

Rare variants have been shown to contribute to many sporadic genetic disorders, motivating systematic genome-wide sequencing studies for complex traits. The standard designs include family-based trios for the analysis of *de novo* mutations (DNMs), or case-control (CC) cohorts. A main challenge in DNM and CC analysis is the necessity for availability of family members or controls matched for ethnicity and sequence coverage respectively.

## Hypothesis:

Based on the assumption that ultrarare variants are likely to be recent DNMs, we developed a statistical framework for performing a burden analysis of rare variants with case-only dataset.

## Methods:

Exome sequencing: Genomic DNA was captured using Roche or IDT kit, sequenced on Illumina platforms. Variants were called using GATK.

Identification of rare variants: We identified very rare qualifying variants (gnomAD MAF < 10<sup>-5</sup> and internal cohort MAF < 10<sup>-3</sup>).

Mutational model: We calculated mutation rates based on both sequence context and genomic features according to mutation type considering 7-mer sequence context.

Adjustments to the mutational model: We calculated mutation rates per gene using the estimated mutation rates while adjusting for the observed coverage, deleteriousness of missense variants and regional intolerance to mutation.

Modelling rare variant events: We modelled the observed number of variants per gene using a Poisson distribution, with the expected number of variants per gene calculated based on the fitted values from a polynomial regression model.

## Results:

As a proof of concept, we applied this statistical framework to detect rare mutations in a published cohort of 553 patients with chronic kidney disease that had been analyzed under a CC design. We recovered all the top genes from the CC analysis such as *PKD1*, *PKD2* and *COL4A5* analysis. In addition, we detected several suggestive candidate genes not previously identified in the CC collapsing analysis - *PKD1L2* (P=2.19E-03; Deleterious Missense), *COL27A1* (P=2.84E-03; Non-benign Missense) and *COL6A3* (P=3.60E-03; Non-benign Missense). We are now extending this analysis to other published datasets to validate the approach.

## Conclusions:

The current statistical model based on the probabilities of mutation of rare variants extends the conceptual framework of DNM and CC analysis and provides a robust and cost-effective approach for rare variant burden analysis by utilizing case-only exome data when family member or control samples are not available.

## PgmNr 2928: Genetic variants and candidate genes for nonsyndromic cleft palate detected by whole-exome sequencing.

### Authors:

Y. Pan <sup>1,2</sup>; S. Lou <sup>1,2</sup>; L. Ma <sup>1</sup>; X. Yu <sup>1,2</sup>; D. Li <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Nanjing Medical University, Nanjing, China; 2) Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, China

---

### Objective:

Genetic factors modifying the nonsyndromic cleft palate only (NSCPO) have been investigated through genome-wide association studies (GWAS) of common genetic variants. However, these variants explain only a fraction of NSCPO heritability.

### Methods:

We conducted a whole-exome sequencing study of common, low-frequency and rare variants to associate genetic variations with NSCPO in 126 cases and 2,827 controls. Single-variant association analysis were performed using a logistic regression model. Burn tests were carried out for coding variants with a minor allele frequency less than 0.01. Pathway enrichment analysis were based on KEGG database. Associated genes were tested for pathway enrichment using Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

### Results:

Twenty-five SNPs reached the genome-wide significance ( $P < 5E-08$ ) and exhibited  $P < 0.05$  after Bonferroni correction. Gene-based analysis revealed that *APOE?KLHDC9?KIF1A* might account for the association with NSCPO risk ( $P < 1E-04$ ). Pathway enrichment analysis indicates that protein digestion and absorption pathway ( $P = 2.31E-03$ ), dilated cardiomyopathy ( $P = 6.03E-03$ ), arrhythmogenic right ventricular cardiomyopathy (ARVC) ( $P = 6.28E-03$ ), and hypertrophic cardiomyopathy (HCM) ( $P = 4.21E-02$ ) reached the thresholds of adjusted  $P$  value  $< 0.05$ .

### Conclusion:

Our results highlight the important role of common and low-frequency variants play in NSCPO susceptibility and indicate that candidate genes and pathways are potentially biologically relevant to NSCPO.

## PgmNr 2929: Integrating eQTLs to identify SNP associations in whole-genome sequencing association studies: A genetic study of platelet aggregation.

### Authors:

J.S. Ngwa <sup>1</sup>; L.R. Yanek <sup>2</sup>; K. Kammers <sup>3</sup>; M.A. Taub <sup>1</sup>; K. Kanchan <sup>2</sup>; R.B. Scharpf <sup>3</sup>; N. Faraday <sup>4</sup>; L.C. Becker <sup>2</sup>; R.A. Mathias <sup>2</sup>; I. Ruczinski <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health; 2) Department of Medicine, Johns Hopkins School of Medicine; 3) Department of Oncology, Johns Hopkins School of Medicine; 4) Department of Anesthesiology and Critical Care Medicine, Johns Hopkins School of Medicine.

---

GWAS studies have successfully identified thousands of SNPs associated with complex traits, however identifying the functional elements through which these genetic variants exert their effects remains a critical challenge. Recently, there is increasing evidence that SNPs associated with complex traits are more likely to be expression quantitative trait loci (eQTLs). Thus, incorporating eQTL information can potentially improve power in highlighting causal genes. Our goal was to investigate the potential to detect novel risk loci among eQTLs only. Our data was comprised of nine platelet aggregation traits from the GeneSTAR family-based study, with whole genome sequencing (WGS) data in European Americans (EA) and African Americans (AA). RNA-seq data were generated from extracted non-ribosomal RNA from 185 (84 AA, 101 EA) iPSC derived megakaryocyte cell lines and 290 (110 AA, 180 EA) platelet samples. We fit a linear mixed model for genetic association within each racial group, including age, sex, principal components for ancestry and random effects from genetic relatedness. We conducted fixed effects meta-analysis using summary statistics in the EA and AA groups. Since eQTLs typically exhibit very strong patterns of linkage disequilibrium, we performed permutation analysis using 1000 permutations for all nine traits to obtain family-wise error rates for eQTL SNPs, substantially lowering the genome-wide significance threshold compared to standard Bonferroni threshold. Our analyses confirmed known platelet aggregation loci such as PEAR1, ADRA2A and ARHGEF3. A number of novel genetic loci were also identified: ADAM22, APIP, ARAP2, BANF2, C6orf195, CBLN2, CEP68, CTNNA1, GPR98, GTF2IRD1, HIVEP2, IMPG2, LOC642236, MACROD2, NT5C1B-RDH14, PI4KAP1, RAB1A, RPTOR, SENP7, SLC1A4 and TMEM120B. Our future work entails exploring Bayesian GWAS by incorporating eQTL information to detect novel risk loci. Our study has the potential of identifying associated SNPs that underlie biological control of gene expression and genes involved in platelet aggregation.

# PgmNr 2930: Multivariate generalized Levene's test for detecting latent gene-gene or gene-environment interactions.

## Authors:

T. Zhang <sup>1,2</sup>; C. Greenwood <sup>2,1,3</sup>; L. Sun <sup>4,5</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Epidemiology, Biostatistics and Occupational Health, McGill University; 2) Lady Davis Institute for Medical Research, Jewish General Hospital; 3) Gerald Bronfman Department of Oncology, McGill University; 4) Statistical Sciences, Faculty of Art of Science, University of Toronto; 5) Biostatistics, Dalla Lana School of Public Health, University of Toronto

---

**Introduction:** Complex traits are influenced by a combination of multiple genes and environmental factors, including gene-gene (GxG) or gene-environment (GxE) interactions. Current multi-SNP GxE tests (e.g. iSKAT of Lin et al. 2016) consider a single environmental factor and assume that the exposure data have been collected. However, relevant exposures may be unmeasured.

**Methods:** Here, we extend the generalized Levene's scale test (Soave and Sun 2017) to jointly study the potential latent interaction effects between multiple SNPs and one (or multiple) environmental factor(s) using a two-stage regression approach. The first stage models the main effects of a set of SNPs on a complex trait and obtains the residuals. The expected values of the absolute residuals from the first-stage are related to phenotypic variances, while in the presence of un-modeled, latent multivariate GxE interactions, the phenotypic variances differ between genotype groups. Thus, the second stage regresses the absolute residuals on this set of SNPs again and tests for the association, for example, using the classical F-test or variance-component type of tests.

**Results:** Simulation studies confirm that our method, the multivariate generalized scale test (MGS), has good power when the environmental factor is not available; and it outperforms direct GxE interaction tests when there are measurement errors in the environmental factor. We also identified type 1 error issue in several direct GxG or GxE interaction tests under the empirical null simulation design as opposed to the theoretical null design (Zhang and Sun 2018), while MGS maintains a correct type 1 error control.

**Conclusion:** The proposed multivariate generalized scale test can potentially serve as a useful tool to identify if any GxE or GxG interaction effects exist, and to identify phenotype-associated genetic variants that might have been missed if the interacting exposure variable(s) were not collected or measured with errors.

## PgmNr 2931: A genome-wide meta-analysis identifies novel loci associated with NSCL/P in Asian population.

### Authors:

Yu <sup>1,2</sup>; S. Lou <sup>1,2</sup>; D. Li <sup>1,2</sup>; L. Ma <sup>1</sup>; Y. Pan <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Nanjing Medical University, Nanjing, China; 2) Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, China

---

Nonsyndromic cleft lip with or without cleft palate (NSCL/P), which is one of the most common human birth defects, involves complex etiologies including genetic and environmental factors. So far, previous genome-wide association study (GWAS) research has found considerable loci with significant genetic significance. However, known variants account for a small percentage of the estimated heritability in risk to these craniofacial birth defects. We performed meta-analysis using four GWAS research databases about Asian population, including two case-control studies from our previous research and two case-parent trio subgroups analyzed by transmission disequilibrium test (TDT) from dbGaP. Fourteen loci reached genome-wide significance, of which six loci have been reported previously, while eight loci (rs6016400, rs9891446, rs6686599, rs1209515, rs9530361, rs1685144, rs596731, rs4751616) were novel. To evaluate the genetic effect of eight loci in NSCL/P risk prediction, we constructed the weighted genetic risk score (wGRS) model, which showed higher wGRS in cases than that in controls among Hua Xi case-control study ( $P=1.56 \times 10^{-13}$ ). It was also confirmed that these eight loci would improve the ability of risk prediction ( $AUC_{6\text{loci}}=0.630$ ,  $AUC_{14\text{loci}}=0.652$ ,  $P=0.011$ ) in the same cohort. In conclusion, This study indicated eight novel loci for NSCL/P susceptibility in Asian population, which can serve as a general and useful reference for further genetic research.

## PgmNr 2932: Ancestry mapping across known primary open-angle glaucoma loci.

### Authors:

T.G. Kinzy <sup>1</sup>; R.P. Igo, Jr. <sup>1</sup>; K.L. Miskimen <sup>1</sup>; G.J.T. Pauer <sup>2</sup>; E. Bala <sup>3</sup>; S.A. Hagstrom <sup>2</sup>; N.S. Peachey <sup>4</sup>; J.N. Cooke Bailey <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH.; 2) Department of Ophthalmic Research, Cole Eye Institute, Cleveland Clinic, Cleveland, OH; 3) Metrohealth Medical Center, Cleveland, OH; 4) Louis Stokes Cleveland VA Medical Center, Cleveland, OH; 5) Cleveland Institute for Computational Biology, Cleveland, OH

---

Primary open-angle glaucoma (POAG) is a leading cause of blindness worldwide; it is a complex disease with known risk factors including age, race/ethnicity, and genetics. Of >100 risk loci for POAG and/or related endophenotypes, most were found in European-descent samples, despite the fact that POAG risk is >2x higher in those of African descent. To increase diversity of POAG-focused genetic studies, we evaluated ancestry differences in known POAG loci in a small sample of African Americans. DNA samples were obtained and extracted by standard methods from 236 African Americans previously recruited in Cleveland, Ohio with glaucoma evaluation. Samples were genotyped on the Illumina Infinium® Expanded Multi-Ethnic Genotyping Array. Excluding individuals with genotyping rates <98%, sex discrepancies, estimated African ancestry <20%, and related individuals, 190 remained for analysis. Variants with a minor allele frequency (MAF) >0.05, Hardy-Weinberg Equilibrium p-value >0.0001, and missingness <1% were retained, resulting in 689,084 variants with genotyping rate =0.9995. Phasing was performed with Eagle 2.3 (1000 Genomes Phase 3 reference [1000G3]) and merged with 1000G3 CEU and YRI subpopulations. Across 267,317 variants with reference MAF >0.05 and  $F_{ST}$  >0.1 between CEU and YRI, RFMix was used to estimate local ancestry inference (3 EM iterations, node size 2, window size 0.2cM). The expected number of YRI alleles were estimated from RFMix forward-backward probability estimates. Global ancestry was estimated by summing the expected number of YRI alleles weighted by distance between markers divided by 2. Admixture mapping tests were limited to known POAG loci  $\pm$ 50kb. Associations were estimated with logistic regression models adjusted for global African ancestry proportion. *COL8A2*, *AXTN2*, and *FERMT2* had p-values <0.05. Within *COL8A2*, the strongest positive association between YRI ancestry and affected status had  $\log(OR)=0.832$ ,  $SE=0.410$ ,  $p=0.0423$ . *AXTN2* and *FERMT2* were negatively associated with YRI ancestry with  $\log(OR)=-0.635$ ,  $SE=0.317$ ,  $p=0.0450$  and  $\log(OR)=-1.141$ ,  $SE=0.529$ ,  $p=0.0309$ , respectively. A negative association was also detected in an intergenic region near *SPTSSA* at chr14:34774848,  $\log(OR)=-1.073$ ,  $SE=0.508$ ,  $p=0.0348$ . We aim to use this project as a catalyst for understanding the genetic difference that specifically increases POAG risk in individuals of African ancestry.



## PgmNr 2933: Genome- and environment-wide study of red meat consumption in 23andMe research participants.

### Authors:

B. Cameron; Y. Huang; K. Fletez-Brant; R. Gentleman; S. Aslibekyan; 23andMe Research Team

[View Session](#) [Add to Schedule](#)

**Affiliation:** 23andMe, Mountain View, CA.

A large body of evidence links red meat consumption to increased risk of chronic diseases and all-cause mortality. Dietary habits cluster in households, but known sociodemographic factors explain only a modest fraction of red meat intake variation. Notably, the genetic basis of red meat intake has never been comprehensively assessed. The 23andMe research database provides a rich opportunity to characterize both genetic and environmental contributions to red meat intake.

In the present study, we conducted the first genome-wide association study of red meat intake in 23andMe research participants of European descent. Red meat intake was separately assessed by the 23andMe Health Profile survey (N=2,135,523) and the NHANES dietary screener (N=420,037, non-overlapping). Linear regression models adjusted for age, sex, age\*sex interaction, ancestry principal components, and genotyping chip. We identified 126 independent hits ( $P < 5 \times 10^{-8}$ ) for red meat consumption in the Health Profile data. Reproducibility was defined as  $P < 0.05/126 = 0.0004$  with consistent direction of effect across data sets for the same variant. Under that definition, we replicated 38 of 126 hits in the smaller NHANES sample. Furthermore, in the UK Biobank data (N~360,000) we replicated 15 hits for processed meat, 14 for beef, 6 for lamb, and 6 for pork intake. Top replicated hits included rs35866622, a coding variant in MAMSTR ( $\beta = -0.017$ ,  $P = 2.9 \times 10^{-33}$ ), rs276454 in CCDC171 ( $\beta = -0.017$ ,  $P = 1.5 \times 10^{-31}$ ), rs429358, a coding variant in APOE ( $\beta = 0.02$ ,  $P = 1.6 \times 10^{-23}$ ), and rs537249739 in HIST1H2BD ( $\beta = -0.015$ ,  $P = 7.8 \times 10^{-21}$ ); all four peaks listed above were robust to additional adjustments for BMI.

We also tested associations between 685 lifestyle, clinical, and socioeconomic phenotypes and red meat intake in 23andMe research participants. Environment-wide models adjusted for age, sex, age\*sex, ancestry principal components, and cohort entry time. The strongest associations (all  $P < 10^{-300}$ ) were identified between red meat intake and other dietary variables (e.g.  $\beta$  per SD of fast food exposure=0.25), health habits (e.g.  $\beta$  per SD of packs smoked/day=0.09), fertility (e.g.  $\beta$  per SD of number of children=0.07), and obesity (BMI > 30;  $\beta$  per SD=0.13). Our findings provide novel insights into dietary behavior of US consumers, informing future causal (e.g. instrumental variable) studies of red meat in disease etiology and highlighting potential intervention targets.

## **PgmNr 2934: Complex traits: Assessment of shared environmental factors using survey responses and genetic information of AncestryDNA customers.**

### **Authors:**

M.V. Coignet <sup>1</sup>; N. Telis <sup>1</sup>; L. Nguyen <sup>1</sup>; K.A. Rand <sup>1</sup>; J.M. Granka <sup>1</sup>; E.L. Hong <sup>1</sup>; K.G. Chahine <sup>2</sup>; C.A. Ball <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Ancestry.com DNA, LLC, San Francisco, CA; 2) Ancestry.com DNA, LLC, Lehi, UT

---

Recent work has highlighted the need to evaluate the impact of environmental factors in complex traits with uncertain genetic aetiology. AncestryDNA has the largest genomic database in the world, with more than 15 million customers genotyped, many of these individuals participating in our Personal Discoveries Project® (PDP), a survey platform. The survey data collected from customers who have consented to participate in scientific research include responses about traits with low heritability (i.e. liking to sing and dance) and more heritable traits, such as eye color or hair color, for which environmental influences may outweigh genetics. These environmental influences among individuals can be characterized by demographic factors such as age, birth, cultural factors, and current geographic locations. Genetic ancestry and genetic communities, or groups of individuals identified to be more genetically similar than others in our database and representing a more recent fine-scale population structure, may also have an impact on these complex traits. Collaborative filtering methods were used to investigate whether the similarities between individuals' genetic information and shared environmental factors are related to the similarities of their survey responses. Here we present a novel approach to assessing the impact of shared environment in complex traits. Results suggest similarities across consistencies between survey responses and genetic information for these individuals, but the levels of consistency vary between traits with low heritability and more heritable traits.

## PgmNr 2935: Fetal and maternal haplotypes effects on spontaneous preterm birth occurrence in an Argentine population at risk.

### Authors:

D.E. Elias<sup>1,2</sup>; L.G. Gimenez<sup>1,2,3</sup>; H.B. Krupitzki<sup>1</sup>; F.A. Poletta<sup>1,2,3</sup>; C. Saleme<sup>4,5</sup>; E.C. Gadow<sup>1</sup>; J.S. Lopez Camelo<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Research Unit, Centro de Educación Médica e Investigaciones Clínicas-Consejo Nacional de Investigaciones Científicas y Técnicas (CEMIC-CONICET), Buenos Aires, Argentina; 2) Estudio Colaborativo Latino Americano de Malformaciones Congénitas (ECLAMC), CEMIC-CONICET, Buenos Aires, Argentina; 3) Instituto Nacional de Genética Médica Populacional (INAGEMP), CEMIC-CONICET, Buenos Aires, Argentina; 4) Department of Neonatology, Instituto de Maternidad y Ginecología Nuestra Señora de las Mercedes (IMGNSM), Tucumán, Argentina; 5) ECLAMC, IMGNSM, Tucumán, Argentina

Preterm birth (PTB) is defined as birth before 37 gestation weeks. It is the leading cause of perinatal morbimortality worldwide. It is estimated that 9,60% of all births in the world are preterm. Due to spontaneous PTB etiological heterogeneity, previous epidemiological genetic studies suggest their discrimination by clinical subtypes: preterm labor leading to PTB (PTB-I) and premature rupture of membranes (PTB-PPROM). The aims of this study is to evaluate the fetal and maternal haplotypes effects on spontaneous PTB occurrence by clinical subtype.

This study included families of neonates born between 2005 and 2010 in the Instituto de Maternidad y Ginecología Nuestra Señora de las Mercedes (Tucumán, Argentina). Inclusion criteria were singleton neonates delivered at less than 37 weeks of gestational age. Neonates were excluded for: congenital anomalies, multiple gestation, or maternal age less than 14 years. Of the 1290 families recruited, 436 (210 PTB-I and 226 PTB-PPROM) families was selected and 24 SNPs from 18 candidate genes of mother, father and child were genotyped. Genotyping quality controls were performed. Haplotypes were defined from protein interactions with a score greater than 0,7 in String database. Eighteen fully connected groups of 2 and 3 proteins were used. To estimate maternal and fetal haplotypes effects, and interaction effects between haplotypes and clinical subtypes, a log-linear additive model of Haplin R package was used.

In the PTB-I subtype, a fetal haplotype generated with the *IL1B-FN1-IGF1* genes presented a double dose Relative Risk (RR) (respect to reciprocal reference) greater than 3. On the other hand, in the PTB-PPROM subtype, fetal haplotype generated with *IGF1R-FN1* genes showed a RR greater than 4. The 95% confidence interval lower end of these RRs was greater than 1 (p-value < 0,01 and q-value < 0,1); the haplotype population frequency was greater than 0,2.

The interaction between clinical subtypes and fetal haplotypes of *IGF1R-FN1* and *TIMP2-FN1* genes presented a RR greater than 3 (calculated as RR PTB-PPROM over RR PTB-I). While the interaction between clinical subtypes and maternal haplotype of *IL1B-FN1* genes showed a RR greater than 2. The p-values of these RR were less than 0,06 and q-values were less than 0,15.

Results suggest that fetal and maternal haplotypes effects would contribute differently to spontaneous PTB risk according to clinical subtype.

## PgmNr 2936: A genome-wide association study of the age-of-onset of idiopathic pulmonary fibrosis.

### Authors:

L. Kraven <sup>1</sup>; R. Allen <sup>1</sup>; A. Taylor <sup>2</sup>; E. Oballa <sup>2</sup>; N. Kaminski <sup>3</sup>; M. Whyte <sup>4</sup>; A. Kang'ombe <sup>2</sup>; M. Tobin <sup>1,5</sup>; I. Sayers <sup>6,7</sup>; R. Hubbard <sup>7</sup>; W. Fahy <sup>2</sup>; T. Maher <sup>8,9</sup>; P. Molyneaux <sup>8,9</sup>; V. Navaratnam <sup>7</sup>; A. Yeo <sup>2</sup>; G. Jenkins <sup>6,7</sup>; L. Wain <sup>1,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Health Sciences, University of Leicester, Leicester, UK; 2) GlaxoSmithKline, Stevenage, UK; 3) Section of Pulmonary, Critical Care, and Sleep Medicine, Department of Medicine, Yale School of Medicine, Yale University; 4) Medical Research Council Centre for Inflammation Research at the University of Edinburgh, Edinburgh, UK; 5) National Institute for Health Research, Leicester Respiratory Biomedical Research Centre, Leicester, UK; 6) National Institute for Health Research, Nottingham University Hospitals, Nottingham, UK; 7) Division of Respiratory Medicine, University of Nottingham, Nottingham, UK; 8) National Institute for Health Research Respiratory Biomedical Research Unit, Royal Brompton Hospital, London, UK; 9) National Heart and Lung Institute, Imperial College, London, UK

**Rationale:** Idiopathic pulmonary fibrosis (IPF) is characterised by progressive scarring of the lungs, predominantly affects people above the age of sixty and has a poor prognosis. The precise age-of-onset of IPF is difficult to determine as early disease is likely to be asymptomatic or may present with similar symptoms to other more common respiratory diseases. Previous genome-wide association studies (GWAS) have identified 17 genetic signals associated with IPF susceptibility but ours is the first to investigate the genetic determinants of the age-of-onset. Identifying genetic determinants of disease can improve our understanding of underlying mechanisms and highlight important pathways for future drug targets.

**Objectives:** To investigate whether there is a genetic component to the age-of-onset of IPF by performing the first GWAS for age-of-diagnosis as a proxy for age-of-onset in IPF.

**Methods:** We performed a discovery GWAS in 465 subjects with IPF from the PROFILE study, assuming an additive genetic model and adjusting for sex, smoking history and the first 10 genetic principal components. Genetic variants which showed suggestive statistical significance ( $P < 1 \times 10^{-5}$ ) in the discovery were followed-up in two additional independent cohorts; the Trent Lung Fibrosis study ( $n=210$ ) and UK Biobank ( $n=98$ ). The results from all three cohorts were then meta-analysed. Genetic signals were defined as genome-wide significant if they reached  $P < 5 \times 10^{-8}$  in the meta-analysis.

**Results:** The genome-wide analysis was performed on 10,858,143 genetic variants. There were 14 independent genetic signals that showed suggestive statistical significance in the discovery analysis. In the meta-analysis there were no genetic variants which reached genome-wide significance but four variants did display suggestive statistical significance ( $P < 1 \times 10^{-5}$ ), with three of these variants displaying a consistent direction of effects across the three cohorts. The most significantly associated variant was rs75681116, which is located on chromosome 8 between the genes *LZTS1* and *RNU3P2*, where each copy of the risk allele was associated with a 9 month earlier age-of-onset of IPF (95% CI [6 months, 12 months],  $P = 2.2 \times 10^{-7}$ ).

**Conclusion:** Our results suggest that there could be a genetic component to the age-of-onset of IPF and highlight four signals of potential interest that require further investigation. We will seek

additional support for the four novel signals in additional independent data sets.

## PgmNr 2937: The association of *APOL1* genotype and pathological factors with preeclampsia risk in Black women.

### Authors:

A. Miller <sup>1</sup>; J. O'Toole <sup>2</sup>; J. Sedor <sup>2,3</sup>; S. Williams <sup>1</sup>; R. Redline <sup>4</sup>; L. Bruggeman <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetic and Genomic Sciences and Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, Ohio.; 2) Departments of Inflammation & Immunity and Nephrology, Cleveland Clinic and the Division of Nephrology, MetroHealth, Medical Center; 3) Department of Physiology and Biophysics, Case Western Reserve University, Cleveland, Ohio.; 4) Departments of Pathology and Reproductive Biology, University Hospitals, Case Western Reserve University School of Medicine, Cleveland, OH 44106

---

Black women in the United States and Africa are at an increased risk for preeclampsia. Previous research has shown that genetic variants in apolipoprotein LI (APOL1, SNPs rs717185313, rs73885319, rs60910145) increase the risk for kidney disease in black populations. Human placentas express APOL1, and although APOL1 is uniquely found in primates, transgenic mice expressing human APOL1 develop preeclampsia. Additionally, previous studies report that the fetal- not the maternal- APOL1 mutation increases the mother's risk of preeclampsia in a recessive inheritance pattern. By further understanding the association of APOL1 variants and preeclampsia, targeted interventions may be developed to improve pregnancy outcomes. The association of fetal APOL1 variants with preeclampsia was evaluated in a case-control study of deliveries from black women at the University Hospitals Cleveland Medical Center (282 controls and 395 cases). We used logistic regression to assess the association between APOL1 variants and preeclampsia and also evaluated associations between APOL1 and several pathological features within several case definitions based on prematurity and severity of preeclampsia using uncomplicated term pregnancies as controls. In logistic regression models adjusted for maternal age, gravidity, and overall villous architecture maturity, a significant association was found between APOL1 genotype and preeclampsia. The association varied by assumed mode of inheritance in all cases: additive 1.22 (95% CI 0.97, 1.54 p = 0.0973), recessive 1.17 (95% CI 0.74, 1.88 p = 0.5071), and dominant 1.40 (95% CI 1.01, 1.94 p = 0.0422). APOL1 genotype was not associated with any pathological feature e.g. villous infarcts, placental weight, and maternal vascular malperfusion. Further studies are needed to increase confidence in the mode of inheritance as our results indicate a different mode of inheritance compared to prior work and to assess whether APOL1 genetic testing can predict risk for preeclampsia.

## PgmNr 2938: Genetic association testing with multivariate outcomes: Comparing statistical methods with application to cognition and eye disease.

### Authors:

Z. Montazeri <sup>1</sup>; F. Moradi <sup>1</sup>; J.E. Ewusie <sup>1</sup>; K. Burkett <sup>2</sup>; E. Freeman <sup>1,3,4</sup>; M.H. Roy-Gagnon<sup>1</sup> <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology and Public Health, University of Ottawa, Ottawa, Ontario, Canada; 2) Mathematica and Statistics, University of Ottawa, Ottawa, Ontario, Canada; 3) Maisonneuve-Rosemont Hospital, Montreal, Canada; 4) Department of Ophthalmology, University of Ottawa, Ottawa, Canada

---

There is growing evidence that some complex traits are caused by overlapping biological mechanisms and thus under common genetic influences. In addition, complex traits are often measured by multiple correlated phenotypes. Analyzing the different related traits in multivariate analyses could provide increased power to detect genetic associations, but is especially challenging with both discrete and continuous phenotypes. Retinal diseases (like age-related macular degeneration (AMD) and glaucoma) and cognitive decline are thought to share biological mechanisms. In addition, cognitive function can be measured by several related cognitive tests. To study the effect of different genetic variants on retinal diseases and cognition, we analyze data from 312 Caucasian participants recruited from the Ophthalmology clinics of the Maisonneuve-Rosemont Hospital (Montreal, Canada). Variables collected included 6 continuous cognitive phenotypes, 2 discrete eye-disease phenotypes (AMD or Glaucoma vs. normal vision), 39 candidate genetic variants and covariates like age and sex. We first used univariate and multivariate regression to assess genetic associations within subsets of multivariate outcomes. SNPs rs1061170 and rs10808746 were found to be associated with glaucoma based on the univariate logistic regression analyses. These two SNPs also showed nominal significance with cognitive phenotypes in multivariate linear regression analyses. Using simulations based on our data, we assess the performance of different methods that can accommodate both discrete and continuous phenotypes in multivariate analyses, including MARV (Multi-phenotype Analysis of Rare Variants; Kaakinen et al., 2017) and GAMuT (Gene Association with Multiple Traits; Broadaway et al., 2016). We apply these methods to our data.

## **PgmNr 2939: An empirical comparison of multivariate and univariate methods for detection of cross-phenotype associations: An example using asthma and body mass index.**

### **Authors:**

Y.D. Salinas; A.T. DeWan

[View Session](#)[Add to Schedule](#)

**Affiliation:** Yale School of Public Health, New Haven, Connecticut.

---

Myriad univariate and multivariate approaches can be used to screen for cross-phenotype associations in genetic epidemiologic studies. Although multivariate methods provide distinct statistical power advantages, they remain underutilized in genome-wide association (GWA) studies. However, this will likely change as multivariate datasets and population-wide biobank data become more readily available for research use. To inform future investigations, we empirically compared the two classes of methods by applying them to a search for cross-phenotype associations and pleiotropy for two positively-correlated phenotypes, asthma and body mass index (BMI). We conducted GWA analyses among 305,945 UK Biobank subjects using univariate regression, MV-PLINK, and MultiPhen, and searched for signals elevated to genome-wide significance ( $p < 5 \times 10^{-8}$ ) in the multivariate analyses, relative to the univariate analyses. Overall, the multivariate approaches detected an additional ten independent genetic loci displaying cross-phenotype associations with asthma and BMI that were not detected by the univariate approach. Among these loci was *RERE*, which harbors variants with the same direction of effect on both phenotypes. To provide more robust statistical evidence of pleiotropy at *RERE*, we dissected the cross-phenotype association using mediation analysis, and show that it remains after accounting for the asthma-BMI relationship in our study population. Our results suggest that *RERE* has pleiotropic effects on asthma and BMI and demonstrate that multivariate methods can improve discovery of cross-phenotype associations and pleiotropy, even though they are not specifically optimized to detect these phenomena.



## PgmNr 2940: Testing 15 billion cis-eQTL associations: Fine-mapping and multivariate methods.

### Authors:

G. Wang <sup>1</sup>; Y. Zou <sup>2</sup>; M. Stephens <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, The University of Chicago, Chicago, IL; 2) Department of Statistics, The University of Chicago, Chicago, IL

---

Contemporary genetic association analyses are typically performed on millions of variables (SNPs) over thousands of phenotypes, for tens of thousands of samples. Phenotypes of interest range from complex traits to molecular features such as gene and protein expression in different tissues or cell types. However, due to the presence of Linkage Disequilibrium (LD), localizing non-zero effect SNPs for these phenotypes from many other correlated neighboring SNPs is a hard problem. In particular, when SNPs are observed to have non-zero effects across phenotypes, it is both interesting and challenging to distinguish true pleiotropic effects from LD induced confounding. It is therefore desirable to perform multi-phenotype fine-mapping in genetic association studies. However, currently few tools are available for the task. This is due in large part to the computational challenge in simultaneously evaluating association between large number of SNPs and phenotypes.

We have previously developed a Sum of Single Effects (SuSiE) model and an efficient Iterative Bayesian Stepwise Selection algorithm for variable selection in regression (Wang et al 2018). The method has several attractive properties for single phenotype genetic fine-mapping. We have recently extended it to fine-map multiple phenotypes. The new method exploits our earlier work on an empirical Bayes approach to multivariate association testing, which models arbitrary correlations of multi-phenotype genetic effects to increase power (Urbut et al 2019). Patterns of effect correlations are first identified across genome via Multivariate Adaptive Shrinkage (MASH), then used as prior covariances to fine-map across phenotypes (MV-SuSiE). This method, called M&M, is therefore capable of rapidly performing multivariate fine-mapping while revealing effect size heterogeneity across phenotypes. We illustrate features of M&M through extensive simulations using GTEx genotypes, under different multi-phenotype effects models. M&M outperforms MV-SuSiE when naive prior effects are used. Compared to the state-of-the-art multivariate fine-mapping method MT-HESS (Lewin et al 2016), M&M is magnitude faster and significantly more powerful. When applied to detecting cis-eQTL across 49 GTEx tissues, M&M is capable of jointly analyzing ~10,000 SNPs under 5 minutes on a desktop computer. We show that M&M elucidates genetic architecture of eQTL in human tissues, and has great potential in large-scale multi-trait GWAS analyses.

## PgmNr 2941: A scalable method for genetic fine-mapping using summary statistics.

### Authors:

Y. Zou <sup>1</sup>; G. Wang <sup>2</sup>; P. Carbonetto <sup>2</sup>; M. Stephens <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistics, University of Chicago, Chicago, Illinois.; 2) Department of Human Genetics, University of Chicago, Chicago, Illinois.

---

Genome-wide association studies (GWAS) have successfully identified many genomic regions associated with complex diseases and traits. To gain more precise biological understanding of these associations, researchers have turned to fine-mapping methods to help pinpoint the causal variants contributing to these diseases and traits. The most successful approaches to fine-mapping are based on Bayesian variable selection methods, which quantify uncertainty in the causal genetic variables by taking into account patterns of Linkage Disequilibrium (LD). However, current Bayesian variable selection methods rely on computationally intensive procedures to make posterior inferences (e.g., exhaustive enumeration, stochastic search). We have recently developed a method, “Sum of Single Effect Regression” (SuSiE), for Bayesian variable selection. SuSiE is computationally efficient and provides a simpler way to summarize fine-mapping results. However, SuSiE requires individual-level genotype and phenotype data, which prevents this approach from being applied to data sets where only summary-level data are available. Here, we extend the SuSiE approach so that it requires only summary statistics (z scores and an LD reference panel). This approach enjoys all of SuSiE’s advantages, including computational complexity that is lower than comparable methods. Additionally, in numerical experiments we show that misspecification of the LD panel can lead to unreliable inferences, and we describe an approach for correcting inconsistencies to improve performance. In simulation studies using data from the Genotype-Tissue Expression (GTEx) project, we show that our method works as well as currently available fine-mapping methods, and provides well-calibrated posterior probabilities. Further, we show that our method has the potential to be robust to misspecification of the LD matrix. Finally, we apply our method to fine-map associations for height and bone mineral density in the UK Biobank data with about 500,000 individuals.

## PgmNr 2942: Improving efficiency of polygenic risk score by leveraging graphical structures among genetic variants and functional annotations.

### Authors:

Y. Zhao <sup>1</sup>; L. Sun <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada.

---

Polygenic risk scores (PRS) was proposed to aggregate genetic effects across a large number of variants that do not individually achieve statistical significance. The original approach to calculating risk scores demands linkage disequilibrium (LD) -based pruning of SNPs and further thresholding based on marginal p-values of individual SNPs. An increasing number of literatures have demonstrated that modeling the inherent correlations between genetic variants can improve power of PRS. However, optimal approaches to exploit the (diverse) correlation structures have not been sufficiently studied. In addition, performance of PRS can be further improved by incorporating external variant-specific functional information. In this work, we propose a summary-statistic based linear mixed effect model to leverage the graphical structures among genetic variants and functional information simultaneously. Benefiting from the sparse graphical structures described by the conditional correlations among genetic variants, the proposed method is capable of estimating the structure via a tuning-robust penalized procedure designed for graphical model from a reference panel. Without access to individual-level data, the proposed framework can integrate variant-specific functional information with ease. A test statistic is established by jointly testing the fixed effect and one-sided random effect (variance component) based on the proposed mixed-effect regression model. The one-sided test for the random effect is proposed to address the boundary issue in terms of type I error control for variance component test. The asymptotical null distribution of the proposed test statistic is investigated, and an efficient p-value calculation approach is developed for finite samples. Empirical evidences from simulation studies and applications will be provided.

## PgmNr 2943: For what sets of variants can array genotyping and imputation with large reference panels replace deep genome sequencing?

### Authors:

S. Hanks <sup>1</sup>; C. Pato <sup>2</sup>; M. Boehnke <sup>1</sup>; L.J. Scott <sup>1</sup>; C. Fuchsberger <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI; 2) State University of New York Downstate Medical Center College of Medicine, Brooklyn, NY; 3) Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Medical University of Innsbruck, Innsbruck, Austria; 4) Institute for Biomedicine, Eurac Research, Affiliated Institute of the University of Lübeck, Bolzano, Italy

Genome-wide association studies (GWAS) have identified thousands of genetic variants associated with hundreds of human traits, with recent studies beginning to leverage sequencing data to detect associations across the full allele frequency spectrum. Deep whole-genome sequencing (WGS) is the current gold standard for capturing genetic variation, particularly rare and structural variants. However, array genotyping followed by imputation is currently ~20 times less expensive than WGS and, as reference panel size and diversity increase, allows for analysis of increasingly lower minor allele frequency variants in samples with diverse ancestry. Our goal is to determine the extent to which variants seen in European and African ancestry WGS studies can be accurately imputed using large haplotype reference panels.

We performed WGS of 3,074 Finns and 7,718 African Americans from the METSIM and InPSYght studies at average depth of 24X and 27X, respectively. For each study, we subsetting the markers from WGS to those included on the Illumina Infinium OmniExpress array. We imputed each array with 1000 Genomes, HRC, and TOPMed reference panels. We calculated the observed Pearson correlation ( $r^2$ ) between the imputed gene dosages and the WGS variant calls for all biallelic SNVs for each imputed dataset. We assigned  $r^2=0$  for any variants not present in a reference panel. We evaluated each WGS study and imputation panel based on the percentage of study-specific WGS SNVs that were well imputed ( $r^2>0.8$ ) within different allele frequency bins.

For METSIM, we found that 90-95% of 5.9M common ( $MAF>5\%$ ), 68-87% of 3.7M low frequency ( $0.5\%<MAF<5\%$ ), and 9-24% of 16.9M rare ( $MAF<0.5\%$ ) variants were well imputed depending on reference panel choice. Only 3-11% of 8.4M singletons were well imputed, and between 86-89% of poorly imputed variants ( $r^2<0.3$ ) had  $MAF<0.1\%$ . The TOPMed and HRC reference panels gave the greatest percentage of well imputed variants for this study.

Our results from METSIM indicate that while imputation with large reference panels is sufficient to capture the large majority of common and low frequency SNVs, deep WGS is still needed to accurately capture most rare variants. We are in the process of assessing imputation quality in InPSYght and considering structural variants and the effect of genomic position on imputation performance in both studies.

## PgmNr 2944: Identifying and exploiting disease-relevant tissues using penalized eQTL enrichment analysis.

### Authors:

Y. Gautam; T.B. Mersha

[View Session](#) [Add to Schedule](#)

**Affiliation:** Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati, OH, 45229, USA

---

Genetic variants discovered from the genome-wide studies are independent of tissue types. However, the biological role of such variants have tissue specific effects. The studies of expression quantitative trait loci (eQTLs) link the disease associated variants with the gene expression in the cell-type specific and thus providing the basis of understanding the biological effects of the genetic variants. Existing functional annotation tools provide tissue-enrichment analyses for a given set of genes without accounting the underlying eQTL associations. Identifying tissues or cell types enriched for the disease associated eQTLs may be helpful to prioritize disease-relevant tissues and develop useful tissue-specific hypotheses of functional mechanisms. In this study, we present a novel approach of *cis*-eQTL-specific tissue scoring analysis using a functionally prioritized set of genome-wide variants. Our approach is carried out in two steps. First, we identified prioritized set of disease associated variants from NHGRI GWAS Catalog including variants in linkage disequilibrium using a composite functional annotation score (cFAS). cFAS integrates multiple functional and regulatory annotations from the transcription start site, regulatory elements in the intergenic regions of the genome. Second, eQTL linked with genetic variants were identified across 48 different tissues from the GTEx project. We developed a tissue scoring approach that penalize the distribution of eQTL signals across different tissues. The null distribution of the score was estimated using a bootstrapping approach. We applied our approaches to multiple allergic diseases - asthma, atopic dermatitis (AD), allergic rhinitis, food allergy, and Eosinophilic esophagitis. Our results unravel tissue enriched for eQTLs for the allergic diseases. Skin was identified as the most enriched tissue for the AD-associated variants, while lung and bloods were ranked among the top scored tissues for asthma. Our enrichment analysis approach could lead to discovery of tissue-specific gene regulations affected by the genome-wide variants and biological insights into various complex diseases including asthma and other allergic diseases.

## PgmNr 2945: Genome-and environment-wide study of sleep quality in 23andMe research participants of European descent.

### Authors:

Y. Huang; S. Aslibekyan; R. Gentleman; 23andMe Research Team

[View Session](#) [Add to Schedule](#)

**Affiliation:** Research, 23andMe. Inc, Mountain View, California.

Previous research has indicated that sleep-related traits, including disturbance and duration, are influenced by both genetic and environmental factors. The genetic basis of overall sleep quality and its non-genetic determinants are not well understood.

In this study, the Pittsburgh Sleep Quality Index (PSQI), a validated measure of overall sleep quality, was used to investigate the genetic and environmental contributions to sleep quality. It is created by summing values on 7 subscores that represent different aspects of sleep quality. A total of 340,195 23andMe research participants of European descent were included.

We first conducted 8 genome-wide association studies (GWAS) using linear regression separately with PSQI global score or its components as the dependent variable. Covariates included age, sex, age\*sex, 5 ancestry principal components, and genotyping chip. 13 loci surpassed the genome-wide significance threshold ( $P=5 \times 10^{-8}$ ) for the PSQI global score. Of those, 4 replicated in the UK Biobank data (<http://www.nealelab.is/uk-biobank/>), including *rs12468708* (the strongest hit, near *FANCL*,  $\beta = -0.012$ ,  $P=1.2 \times 10^{-13}$ ) and *rs9815484* (near *IGSF11*,  $\beta = 0.014$ ,  $P=1 \times 10^{-10}$ ) for “insomnia”, *rs34291892* (*FOXP2*,  $\beta = -0.010$ ,  $P=2.7 \times 10^{-10}$ ) for “sleep duration” and “being a morning person”, and *rs704633* (*NAV1*,  $\beta = 0.010$ ,  $P=4.7 \times 10^{-9}$ ) for “having sleep problems”. When considering components separately, 43 loci surpassed the genome-wide significance threshold, including 10 for duration, 2 for efficiency, 15 for onset latency, 4 for disturbance, 15 for sleep medication, 2 for self-rating sleep quality, and 4 for daytime dysfunction. Genetic correlations calculated using LD score regression between PSQI and its components ranged from 0.67 to 0.93 (all  $P < 4 \times 10^{-186}$ ).

We also systematically evaluated associations between 681 other lifestyle, clinical and socioeconomic phenotypes and PSQI using linear regression models adjusted for age, sex, age\*sex, cohort entry time and ancestry. Effects were expressed per standard deviation of exposure to facilitate comparisons across phenotypes. After adjusting for multiple comparisons, the strongest predictors (all  $P < 10^{-300}$ ) were mental health and stress (e.g.  $\beta$  per SD of perceived stress = 0.27) and health conditions (e.g.  $\beta$  per SD of back pain frequency = 0.15). Our findings advance the understanding of both genetic and environmental sleep quality determinants in the European population and lay the groundwork for future interventions.

## PgmNr 2946: First GWAS on transferrin N-glycans: One step closer to understanding the genetics of protein glycosylation.

### Authors:

A. Landini <sup>1</sup>; P. Navarro <sup>2</sup>; I. Trbojevic-Akmacic <sup>3</sup>; F. Vuckovic <sup>3</sup>; C. Hayward <sup>2</sup>; G. Lauc <sup>3,4</sup>; L. Klaric <sup>2</sup>; J.F. Wilson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom; 2) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Crewe Road South, Edinburgh, United Kingdom; 3) Genos Glycoscience Research Laboratory, Borongajska cesta 83h, Zagreb, Croatia; 4) Faculty of Pharmacy and Biochemistry, University of Zagreb, Ante Kovacica 1, Zagreb, Croatia

---

Glycomics, the study of the collection of glycans in biological systems, is an emerging field among omics data. Despite glycans being involved in the aging process and in a wide variety of diseases (including cancer, immunological and autoimmune diseases, diabetes and congenital disorders), the genetic regulation of glycosylation is not fully understood.

To partly address the knowledge gap regarding genetic regulation of glycosylation, we focused on transferrins, blood plasma glycoproteins regulating iron's level in body fluids. This is the first study integrating genomics and transferrin glycomics, thus allowing us to investigate genetic regulation of transferrin N-glycosylation and whether the same or distinct genes are involved in the N-glycosylation of different proteins. Accordingly, we compared results obtained for transferrin N-glycans to state-of-the-art knowledge about genetic regulation of immunoglobulin G (IgG) N-glycosylation.

We performed genome-wide association study (GWAS) of 35 transferrin N-glycans, using data collected on 948 samples from the CROATIA-Korcula cohort, a genetic isolate characterized by high kinship. This dataset features UPLC-quantified N-glycan traits and genetic data imputed to the Haplotype Reference Consortium.

We identified 778 genome-wide significant loci (p-value < 5-e08/35 traits), mapping in genes encoding glycosyltransferases (*MGAT5*, *FUT6*, *FUT8*, *ST3GAL4*, *B3GAT1*), genes potentially related to the glycosylation process (*MSR1*, *TUSC3*) and transferrin (*TF*), or genes previously associated with N-glycosylation of IgG (*NXPE1*, *NXPE4*). Some of these genes (*TF*, *MSR1*, *TUSC3*) have not previously been associated with glycosylation.

We thus suggest that while some genes are responsible for regulating the N-glycosylation of both transferrin and IgG proteins, others instead appear to control N-glycosylation of only one of these proteins. Our findings shed light on the complex mechanisms regulating protein N-glycosylation and also corroborate the notion that while some genes are specific for a restricted number of proteins, others instead affect the N-glycosylation of multiple proteins.



## **PgmNr 2947: Model selection and permutation testing for association studies within a large direct-to-consumer genetic cohort.**

### **Authors:**

E.P. Sorokin <sup>1</sup>; D.S. Park <sup>1</sup>; M.V. Coignet <sup>1</sup>; K.A. Rand <sup>1</sup>; J.M. Granka <sup>1</sup>; E.L. Hong <sup>1</sup>; K.L. Chahine <sup>2</sup>; C.A. Ball <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) AncestryDNA, San Francisco, California; 2) AncestryDNA, Lehi, Utah

---

When conducting genome-wide association studies (GWAS), model selection is critical to identify and adjust for confounding factors, while permutation testing enables an empirical assessment of statistical significance. In conducting GWAS within a large cohort comprised of AncestryDNA customers, who have consented to participate in scientific research, we performed logistic regression on binary outcomes collected from self-reported survey data. We regressed on predictors including genotype, age, biological sex and principal components within a cohort comprised of European-ancestry individuals. Using this approach, we were able to replicate previously identified loci from GWAS literature. To conduct these analyses on well-studied traits with known genetic architectures and also to extend this work to previously unpublished traits, it was essential to assess whether we had sufficiently controlled for confounding factors, especially European population substructure. In an example GWAS on a previously unpublished trait, we observed substantial genomic inflation. We looked at top signals for evidence of population stratification, examining allelic differentiation across Northern and Southern Europeans, both in frequencies from 1000 Genomes, and also within our European ancestry cohort. Furthermore, we observed differences in prevalence of traits between largely unadmixed European-ancestry groups in the AncestryDNA cohort. These findings led to evaluating various approaches to account for population substructure. We also used permutation testing in order to empirically set a threshold of significance for these GWAS. We tested setting significance thresholds by permuting naively, permuting by conditioning on binary covariates, and permuting by conditioning on all covariates jointly. We found increased value in the conditional methods compared to the naive approach. In conducting a GWAS in previously unexplored traits, we show that model selection and empirical significance thresholds are study and trait-specific.



## PgmNr 2948: Genome-wide association study of human milk oligosaccharides in lactating mothers and interaction effects on lung health among infants in the CHILD Study.

### Authors:

A. Ambalavanan<sup>1</sup>; L. Chang<sup>1,2</sup>; J. Choi<sup>1</sup>; A. Lamri<sup>3</sup>; B. Robertson<sup>4</sup>; C. Yonemitsu<sup>4</sup>; S.E. Turvey<sup>5,6</sup>; P.J. Mandhane<sup>7</sup>; A.B. Becker<sup>8,9</sup>; T.J. Moraes<sup>10</sup>; S.S. Anand<sup>11</sup>; G. Pare<sup>12</sup>; D.L. Lefebvre<sup>11</sup>; M.R. Sears<sup>11</sup>; P. Subbarao<sup>10</sup>; L. Bode<sup>4</sup>; M.B. Azad<sup>8,9</sup>; Q.L. Duan<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical and Molecular Sciences, Queen's University, Kingston, Ontario, Canada; 2) School of Computing, Queen's University, Kingston, Ontario; 3) Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario; 4) Department of Pediatrics and Larsson-Rosenquist Foundation Mother-Milk-Infant Center of Research Excellence, University of California San Diego, La Jolla, California; 5) Division of Allergy and Immunology, Department of Pediatrics, University of British Columbia, Vancouver, British Columbia, Canada; 6) Department of Pediatrics, Child and Family Research Institute and British Columbia Children's Hospital, Vancouver, British Columbia, Canada; 7) Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada; 8) Manitoba Developmental Origins of Chronic Diseases in Children Network (DEVOTION), Children's Hospital Research Institute of Manitoba, Winnipeg, Manitoba, Canada; 9) Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada; 10) Department of Pediatrics, Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada; 11) Department of Medicine, McMaster University, Hamilton, Ontario, Canada; 12) Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada

---

Human milk oligosaccharides (HMOs) are complex carbohydrates found uniquely in human breast milk and absent in formulas<sup>1</sup>. HMOs shape the growth of infant gut microbiota and contribute to immune system development<sup>2</sup>. While genetic variants in *FUT* genes have been correlated with HMO secretion, it remains to be determined if other genetic factors modulate HMO concentrations in breast milk<sup>3</sup>. This is the first genome-wide association study (GWAS) of HMOs among lactating mothers. Moreover, we tested the interaction effects of HMO exposures with asthma risk variants on respiratory outcomes among infants of these mothers. Breast milk samples were collected 3-4 months postpartum from 980 lactating mothers recruited in the Canadian Healthy Infant Longitudinal Development (CHILD) birth cohort. A total of 19 HMOs were analyzed by high-performance liquid chromatography. Genotyping data from the mothers and their children were obtained using the Illumina HumanCoreExome BeadChip and imputations were performed using the Haplotype Research Consortium data on the Michigan Imputation server. In addition to GWAS of maternal HMO concentrations, we computed genetic risk scores correlated with recurrent wheeze, asthma diagnosis and lung function at age 5 in their children using summary statistics from the largest published GWAS study of asthma<sup>4</sup>. Our results indicate that multiple loci in chromosome 19 (*FUT2* & *FUT3*) were correlated with HMO concentrations. In addition, several novel loci on chromosome 3 were correlated with 6'-sialyllactose ( $P = 3.92 \times 10^{-9}$ ) and disialyllacto-N-hexose ( $P = 1.51 \times 10^{-8}$ ); the most significant variant located 3' of the *ST6GAL1* gene. Interestingly, children exposed to high concentrations of sialylated HMOs had reduced prevalence of recurrent wheeze between ages 2-5. These results were

significant among children with high genetic risk of developing wheeze and asthma. In conclusion, our study replicated previous genetic associations with HMO secretion but also identified several novel loci. Furthermore, our results demonstrated protective effects of specific HMOs on risk of respiratory outcomes including recurrent wheeze and asthma during early childhood.

**References:**

Bode et al., *Glycobio*, 2012

McGuire et al., *Curr Opin Biotech*, 2017

Azad et al., *J Nutr*, 2018

Demanais et al., *Nature Gen*, 2018

## PgmNr 2949: Genome-wide association studies of n-3 and n-6 polyunsaturated fatty acids in African Americans and Hispanic.

### Authors:

C. Yang <sup>1</sup>; J. Veenstra <sup>2</sup>; D. Siscovick <sup>3</sup>; T.M. Bartz <sup>4</sup>; D. Mozaffarian <sup>5</sup>; Y.-D.I. Chen <sup>6</sup>; D. Riadi <sup>7</sup>; L. Steffen <sup>8</sup>; M.Y. Tsai <sup>9</sup>; A.C. Wood <sup>10</sup>; F.H. Chilton <sup>11</sup>; S.S. Rich <sup>1</sup>; N.L. Tintle <sup>2</sup>; R.N. Lemaitre <sup>12</sup>; A. Manichaikul <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 2) Departments of Biology and Statistics, Dordt College, Sioux Center, IA; 3) New York Academy of Medicine, New York, NY; 4) Department of Biostatistics, University of Washington, Seattle, Washington; 5) Friedman School of Nutrition Science & Policy, Tufts University, Tufts School of Medicine and Division of Cardiology, Tufts Medical Center, MA; 6) Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA; 7) Department of Computer Science, Dordt College, Sioux Center, IA; 8) Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN; 9) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, USA Alexis C Wood, USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX; 10) USDA/ARS Children's Nutrition Research Center, Baylor College of Medicine, Houston, TX, USA Floyd H Chilton, Department of Nutritional Sciences and the BIO5 Institute, University of Arizona, Tucson, AZ; 11) Department of Nutritional Sciences and the BIO5 Institute, University of Arizona, Tucson, AZ; 12) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington

---

**Background.** n-3 and n-6 long chain polyunsaturated fatty acids (LC-PUFAs) play critical roles in many physiologic and pathophysiologic processes. We investigated the association of common genetic variation with eight major LC-PUFAs through genome-wide association studies (GWAS) in African American [AfAm] and Hispanic American [HisAm] cohorts from the CHARGE Consortium.

**Methods.** Circulating n-3 (eicosapentaenoic acid [EPA], docosapentaenoic acid [DPA], docosahexaenoic acid [DHA], alpha-linolenic acid [ALA]) and n-6 (arachidonic acid [AA], gamma-linolenic acid [GLA], dihomo-γ-linolenic [DGLA], linoleic acid [LA]) PUFAs were measured in participants from three population-based cohorts (the Multi-Ethnic Study of Atherosclerosis [MESA] (AfAm: n=1462 and HisAm: n=1218), the Cardiovascular Health Study [CHS] (AfAm: n=425) and the Framingham Heart Study [FHS] (AfAm: n=203 and HisAm: n=211)). Genome-wide analysis was carried out separately in each cohort with covariate adjustment for age, sex, study site and principal components of ancestry. For each genetic variant and fatty acid trait, cohort-specific GWAS results were filtered using EasyQC based on MAC > 6 and imputation R-squared > 0.3. Cohort-specific GWAS results were combined by inverse-variance weighted meta-analysis in METAL.

**Results.** After filter on effective heterozygosity > 60, we confirmed genetic loci in *FADS1/2* and *ELOVL2*, previously reported in European-ancestry based GWAS in CHARGE. We further identified novel genome-wide significant ( $P < 6.25 \times 10^{-9} = 5 \times 10^{-8} / 8$  traits) loci (4 in AfAm and 7 in HisAm) across

the LC-PUFA traits. In AfAm, significant associations were identified for DPA (*CASC15*, rs717894,  $P = 3.302 \times 10^{-9}$ ) and EPA (*MGAT5*, rs112877348,  $P = 4.624 \times 10^{-10}$ ; and *MRVI1*, rs16908026,  $P = 1.812 \times 10^{-9}$ ). In HisAm, we observed significant associations for AA (*TPCN2*, rs3750967,  $P = 6.436 \times 10^{-9}$ ; *PHF21A*, rs76595296,  $P = 3.026 \times 10^{-11}$ ; and *KDM2A*, rs77988791,  $P = 1.527 \times 10^{-11}$ ), DGLA (*PDXDC1*, rs4985128,  $P = 1.178 \times 10^{-9}$ ), EPA (*OR5B3*, rs17152659,  $P = 2.594 \times 10^{-9}$ ; *SLC22A25*, rs75269907,  $1.316 \times 10^{-11}$ ), GLA (*LOC107986082*, rs55918609,  $P = 2.788 \times 10^{-9}$ ) and LA (*NUDTB*, rs145834191,  $5.536 \times 10^{-11}$ ; and *PHF21A*, rs76595296,  $1.627 \times 10^{-9}$ ).

Conclusion. Next steps will include replication of the identified variant associations and eQTL colocalization analyses. The results of our study may be useful in defining future targeted recommendations for LC-PUFAs intake in the African American and Hispanic population.

## PgmNr 2950: Quantifying pairwise trait co-regulation through modeling individual genetic correlations.

### Authors:

S.KG Forsberg <sup>1</sup>; A. Ziyatdinov <sup>2</sup>; J.F. Ayroles <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Princeton University, Princeton NJ; 2) Harvard T.H. Chan School of Public Health, Boston Massachusetts

---

One major goal in quantitative and medical genetics is to determine how gene networks drive individual variation and establish when their dysregulation may lead to disease emergence. To that end, correlation patterns between genes or metabolites are frequently used to assemble co-expression networks, under the well-supported assumption that correlation often implies co-regulation. However, whether correlation networks vary across individuals is largely an open question. We still have a poor understanding of how genetic variation, environmental perturbations, or disease states can modify essential patterns of co-regulation.

Correlations are generally estimated as the property of a population, not an individual. This is a key limitation because it does not allow for the investigation of individual-specific factors that may influence variation in co-regulation. One solution is to gather multiple observations per individual (for example over time), thus allowing us to view correlation between traits as a property of an individual. This shift in focus from a population to an individual estimate, enable us to address new questions: 1) Does the correlation between genes or metabolites vary between individuals? 2) Are the patterns of correlation heritable? 3) Are certain disease states characterized by a change in correlation pattern?

In order to tackle these questions, we developed a mixed linear model that accommodates intra individual correlation through a random slope. When decomposing individual correlations into an environment and genetic components, a polygenic random slope quantifies the degree of correlation between traits. We show through extensive simulations, how this statistical framework can to be used to distinguish between three different phenomena that may lead to correlation between traits: i) additive genetic effects on each trait independently, ii) genetic effects on the correlation between traits (i.e. a correlation QTL), and iii) pleiotropic effects (i.e. a given variant affects multiple traits). We show that this approach can estimate the heritability of correlation in an unbiased fashion - provided repeated measures for a given individual. We apply this approach to longitudinal data from medical records with the goal of identifying the genetic basis of co-regulation and find cases where a change in the degree of correlation between traits (e.g. metabolites) may reveal the loss of homeostasis and potentially disease emergence.

# PgmNr 2951: An evaluation of the frequency of Intron 1 and 22 inversion and the clinical characterization of haemophilia A in Nigeria.

## Authors:

E. Olude <sup>1</sup>; T. Adeyemo <sup>1</sup>; A. Akanmu <sup>1</sup>; F. Bello <sup>1</sup>; O. Oyewunmi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Haematology and blood transfusion, Lagos university teaching hospital, Lagos, Nigeria; 2) Molecular Laboratory Department, Sickle Cell Centre, Idi-Araba, Lagos, Nigeria

---

**Background:** Haemophilia A(HA) is an X-linked recessive disorder. Its aetiology is founded on mutations within the *F8* gene which result in reduced or absent circulating levels of the F8 protein being produced. The mutational mechanisms include inversions, deletions, point mutations, missense etc. This study focussed on the detection of 2 of the most recurring mutations, intron 1 and 22 inversions of the *F8* gene. Intron 1 and 22 inversions are responsible for 1-7% and 40-50% of the mutations in Haemophilia A. The clinical characteristics and inhibitor screen were also assessed in this study.

**Materials and Methods:** 76 Haemophilia A patients registered with the Haemophilia Foundation of Nigeria were enrolled in the study. Their clinical characteristics, FVIII concentrations, inhibitor status and mutation types were assessed. Coagulometry involved clot-based assays via semi-automated coagulometer. The Inverse-Shifting PCR protocol was used for detection of the inversion mutations. IS-PCR protocol is designed to be able to detect inversions, duplications and deletions within intron 22.

**Results:** 75% (57/76) of the participants were found to have severe Haemophilia A( FVIII level < 1%) while 25% were of moderate severity with FVIII level 1-5%. The most frequent presentation was joint bleed with a frequency of 60% and 20% of patients had developed significant disuse atrophy. Of the 76 subjects, 27 had inversion mutations; 2 with intron 1 inversion and 25 with intron 22 inversions. Intron 1 and 22 inversions were documented in 2/57(3.5%) and 19/57(33.3%) of severe HA respectively. Twenty-nine of 76 subjects were positive for FVIII inhibitors. No significant associations were found when the mutations were matched against the clinical characteristics, FVIII levels and inhibitor status.

**Conclusion:** This study documented the prevalence of severe HA among Nigerian subjects with diagnosed haemophilia to be 75%. Inversions of introns 1 and 22 occurred in 3.5 and 33.3% of the severe haemophiliacs. FVIII inhibitors were found in 38.2% of Nigerian haemophiliacs.

## PgmNr 2952: Severity of TB associated with eQTL regulatory variants in two independent cohorts from Kampala, Uganda.

### Authors:

M.L. McHenry<sup>1,2</sup>; R. Sobota<sup>7</sup>; J. Bartlett<sup>1</sup>; R.P. Igo Jr.<sup>1</sup>; N.B. Hall<sup>1</sup>; H. Mayanja-Kizza<sup>3,4</sup>; C. Wejse<sup>5</sup>; M. Joloba<sup>3,4</sup>; W.H. Boom<sup>2,3</sup>; G. Sirugo<sup>6</sup>; S.M. Williams<sup>1</sup>; C.M. Stein<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH; 2) Department of Medicine, Case Western Reserve University School of Medicine, Cleveland, OH; 3) Uganda-CWRU Research Collaboration, Cleveland, OH; 4) School of Medicine, Makerere University, Kampala, Uganda; 5) Department of Infectious Diseases/Center for Global Health, Aarhus University, Aarhus, Denmark; 6) Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania, Philadelphia, PA; 7) Feinberg School of Medicine, Northwestern University, Evanston, IL

---

Several studies have demonstrated human genetic susceptibility for tuberculosis (TB), but few have examined TB severity. We assessed genetic association with TB severity in our household contact study in Kampala, Uganda. We analyzed genotype and severity in 148 HIV-infected TB cases (Cohort 1). We sought to replicate our findings in Cohort 2, 198 independently collected HIV-uninfected TB cases. We analyzed only SNPs available in both cohorts. The two cohorts were similar with respect to mean age, distribution of sex, and distribution of TB severity. We modeled TB severity using the TBscore by Wejse et al., which ranges from 1-13, incorporating clinical symptoms present. It has previously been associated with risk of mortality as well as important biomarkers in TB patients. We modeled severity as a function of host SNP genotype, analyzing 214,070 SNPs in both cohorts derived from an exome chip. We observed association between TBscore and numerous eQTL SNPs in Cohort 1 that replicated in Cohort 2, including multiple SNPs in linkage disequilibrium and regions with multiple significant SNPs in the same gene. In total, 67 SNPs were associated with a combined  $p\text{-value} < 10^{-4}$  and only 12% of these were exonic. The strongest association was found for rs10176270 in the combined analysis ( $\beta=0.82$  and  $p=1.17 \times 10^{-6}$ ), an eQTL that maps to the HEATR5B gene on chromosome 2. HEATR5B is ubiquitously expressed in human tissues and is involved in endocytosis, protein localization, and retrograde transport to the Golgi apparatus. Because of the large proportion of non-exonic localization of associated SNPs, we used FUMA GWAS and HaploReg v4.1 to annotate SNP functionality. We observed that many of these SNPs had potential regulatory influences on TB immune cells. For example, we determined that rs10176270 has been identified as an enhancer and promoter that is associated with differential expression of HEATR5B in human whole blood, lymphoblastic, and CD14+ monocyte cell lines. The importance of this SNP in monocytes is especially important, as these cells are precursors to tissue macrophages, which are intracellularly infected by *Mycobacterium tuberculosis* and thought to be the primary facilitators of the host immune response to TB infection. Further, rs10176270 is in LD ( $r^2 > 0.8$ ) with other SNPs that have previously identified in the GWAS catalog to be significantly associated with red cell distribution width, hematopoiesis, and regulation of T cells.

## PgmNr 2953: Evaluation of the host genetic effects of tuberculosis-associated variants among patients with type 1 and type 2 diabetes mellitus.

### Authors:

Y.V. Sun <sup>1</sup>; H. Zhong <sup>1</sup>; Y. Huang <sup>1</sup>; Q. Hui <sup>1</sup>; M. Gwinn <sup>1</sup>; M.J. Magee <sup>2</sup>; N.R. Gandhi <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA; 2) Division of Epidemiology and Biostatistics, School of Public Health, Georgia State University, Atlanta, GA; 3) Division of Infectious Diseases, School of Medicine, Emory University, Atlanta, GA; 4) Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA

---

**Background:**As the global burden of non-communicable diseases continues to expand, the public health importance of understanding the link between tuberculosis (TB) and diabetes is increasing. Genetic association studies have identified numerous host genetic variants linked to TB. However, to date, host genetic mechanisms linking TB and diabetes remain unknown.

**Methods:**We conducted a genetic association study to evaluate the effects of seven previously reported TB-related host genetic variants (genome-wide significant associations from published studies) in patients with diabetes using genetic and phenotypic data from the UK Biobank. The study included 409,692 adults of European ancestry including 13,976 type-2 diabetes (T2D) cases and 2,177 type-1 diabetes (T1D) cases defined by ICD-10 diagnosis codes.

**Results:**Out of seven TB-associated SNPs, four were associated ( $p < 0.05$ ) with T2D and one with T1D ( $p < 0.05$ ), after adjusting for age, sex, BMI, smoke, alcohol use and population structure. After correction for multiple testing, the C allele of SNP rs3135359 (*BTNL2* gene on HLA-A) was significantly associated with increased prevalence of T2D (OR 1.06, 95% CI 1.025-1.096), and increased prevalence of T1D (OR 1.72, 95% CI 1.57-1.88). The rs3135359 association with T2D diminished after further exclusion of probable T1D cases defined by ICD-10 codes. The SNP rs3135359 was also strongly associated with thyroid disease and additional autoimmune diseases including multiple sclerosis and rheumatoid arthritis. SNPs rs12680942 (OR 1.03, 95% CI 1.00-1.06) and rs4733781 (OR 1.03, 95% CI 1.00-1.06) on chromosome 8 (*ASAP1* gene) were associated with T2D after exclusion of T1D cases.

**Conclusion:**Our findings suggest that host genetic effects may partially explain observed associations between TB and T2D. Our analyses demonstrated common genetic factors underlying TB and non-communicable disease, particularly via immune functions of the HLA region.



## PgmNr 2954: Genotype-phenotype correlation in lysosomal storage diseases detected by newborn screening in Mexico.

### Authors:

J. Navarrete <sup>1,2</sup>; D. Cervantes <sup>1,2</sup>; A. Limon <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Gen, Hosp Sur PEMEX, Mexico City, Federal District, Mexico; 2) Facultad de Medicina universidad Nacional Autónoma de México

---

The goal of newborn screening is an early detection of inborn errors of metabolism diseases. In Mexico we began newborn screening since 1977 with very few inborn errors of metabolism such as phenylketonuria, galactosemia, congenital hypothyroidism, sickle cell anemia and cystic fibrosis. (1) Petróleos Mexicanos is a big governmental institution with approximately ten thousand workers and their families. Since 2005 a larger screening has been done to all newborns in this institution through all the country.

We test for most aminoacidopathies; since August 2012 we included primary immunodeficiencies, Gaucher disease, Niemann-Pick (A/B) disease, Pompe disease, Krabbe disease, Fabry disease, and MPS1.

We analyzed our results from August 2012 to December 2018 , we have found 5 newborns with Fabry disease confirmed with enzyme activity and molecular analysis; and 6 patients with Pompe disease, three were pseudodeficiencies. one was a late onset presentatio; one patient with MPS I homozygous and one Gaucher patient.

These findings give us the opportunity to make genotype-phenotype correlation for the early treatment of lysosomal storage diseases. This early detection will allow us to make a close follow up of these patients; and through a study of their genealogy detect other Family members with the disease.

## PgmNr 2955: Genetic risk factors for recombination-related maternal chromosome 21 nondisjunction.

### Authors:

J.M. Chernus<sup>1</sup>; E. Feingold<sup>1,2</sup>; S.L. Sherman<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania, United States of America; 3) Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, United States of America

---

Nondisjunction in oocytes is the leading cause of pregnancy loss, as well as intellectual disabilities and birth defects among pregnancies surviving to term. In addition to increasing maternal age, altered patterns of meiotic recombination are a known risk factor for nondisjunction. Specifically, errors in meiosis I (MI) and meiosis II (MII) are each associated with distinct “susceptible” patterns, suggesting their etiologies may differ. For MI nondisjunction of chromosome 21 these include a lack of recombination on the chromosome or a single, telomeric exchange; MII errors are associated with exchange near the centromere.

Previous genome-wide association studies have found associations between common variants and recombination phenotypes (e.g., global recombination rate, recombination outside of hotspots). Our own previous and ongoing work also suggests that common genetic variants may contribute to the risk of nondisjunction of chromosome 21, with some variants conferring risk for nondisjunction in just one stage and others conferring risk in either stage. Here we extend that work by analyzing the relationship between common variation and the occurrence of susceptible recombination patterns.

A set of 2,186 study participants was genotyped on the HumanOmniExpressExome-8v1-2 array. These included 749 live birth offspring with maternally-derived standard trisomy 21 and 1,437 of their parents. We developed methods for determining the stage of meiosis in which nondisjunction occurred and for locating recombination events on chromosome 21 in these parent-child trios and dyads. We then analyzed recombination patterns in two ways. In one group of analyses, susceptible recombination patterns were treated as phenotypes for GWAS and candidate-gene testing. In the other, we performed stratified GWAS of nondisjunction. Finally, we performed age-stratified analyses of nondisjunction.

# PgmNr 2956: Systematic Mendelian randomization and colocalization analyses of the plasma proteome and blood transcriptome to prioritize drug targets for complex disease.

## Authors:

J. Zheng <sup>1</sup>; B.M. Brumpton <sup>2</sup>; P.G. Bronson <sup>3</sup>; Y. Liu <sup>1</sup>; P.C. Haycock <sup>1</sup>; B. Elsworth <sup>1</sup>; V. Haberland <sup>1</sup>; D. Barid <sup>1</sup>; V. Walker <sup>1</sup>; J.W. Robinson <sup>1</sup>; S. John <sup>4</sup>; B. Prins <sup>5</sup>; H. Runz <sup>3</sup>; M.R. Nelson <sup>6</sup>; M.R. Hurle <sup>6</sup>; G. Hemani <sup>1</sup>; B.O. Åsvold <sup>2,7</sup>; A. Butterworth <sup>8</sup>; G. Davey Smith <sup>1,9</sup>; R.A. Scott <sup>10</sup>; T.R. Gaunt <sup>1,9</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) MRC Integrative Epidemiology Unit (IEU), Bristol Medical School, University of Bristol, Oakfield House, Oakfield Grove, Bristol, BS8 2BN, UK; 2) Department of Public Health and Nursing, K.G. Jebsen Center for Genetic Epidemiology, NTNU, Norwegian University of Science and Technology, Trondheim, Norway; 3) Human Target Validation Core, Translational Biology, Biogen, 250 Binney Street, Cambridge, MA 02142, USA; 4) Translational Biology, Biogen, 250 Binney Street, Cambridge, MA 02142, USA; 5) Human Genetics, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB10 1SA, United Kingdom; 6) Genetics, GlaxoSmithKline, Collegeville, PA, USA; 7) Department of Endocrinology, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway; 8) MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK; 9) NIHR Bristol Biomedical Research Centre, Bristol, UK; 10) GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

**Background:** The plasma proteome and blood transcriptome are potential sources of therapeutic targets. Genetic studies of these molecular traits enable systematic comparison of genetic architecture between protein and gene expression; and estimate causal associations of these traits on human diseases.

**Methods:** Here, we estimated the effects of 1,740 plasma proteins and 16,058 blood transcripts on 576 phenotypes in Europeans using two-sample Mendelian randomization (MR) followed by single and multi-trait colocalization (coloc/moloc). We report the findings of 9.46 million gene-phenotype associations in an accessible database: EpiGraphDB ([www.epigraphdb.org/xqtl/](http://www.epigraphdb.org/xqtl/)).

**Results:** We compared 4,094 cis/trans pQTLs and eQTLs associated with the same gene (1,194 genes). Only 317 (7.7%) of the pQTLs and eQTLs were in LD ( $r^2 > 0.1$ ), of which 314 were in the cis region. Although the pQTLs and eQTLs differed, their estimated causal effects on the disease were highly correlated ( $r = 0.98$ ) for the 29 associations with strongest MR and colocalization evidence from pQTL and eQTL ( $P_{MR} < 5.3 \times 10^{-9}$ , coloc probability  $> 0.8$ ). We found 6 associations with multi-trait coloc evidence, including an approved drug target (PLAU) with a novel indication (Crohn's disease). Mediation analyses of these 6 findings suggest that PLAU gene expression influenced PLAU protein expression, ultimately affecting susceptibility to Crohn's disease. Of 1,718 protein-phenotype associations with MR evidence, 1,215 (70.7%) had coloc evidence. Of 19,775 gene expression-phenotype associations with MR evidence, 12,787 (64.7%) had coloc evidence. This work improves on previous omics studies by making careful use of multiple cis and trans instruments and identifying additional gene-phenotype associations. Retrospective evaluation of 268 developed drugs (Pharmaprojects) showed that target-indication pairs with proteomic MR and

coloc support were more likely to succeed (OR=26.4; 95%CI=4-580). Further, we validated 8 marketed drugs, prioritised 8 targets in current drug trials, and identified 53 repurposing opportunities.

**Conclusions:** This is the first systematic MR and coloc analysis of the plasma proteome and blood transcriptome. Our results suggest that genetic confounding due to LD may be widespread in phenome-wide association studies of molecular traits. We identified novel gene-phenotype associations and provide evidence that proteomic MR/coloc support of drug targets increases probability of success for drug discovery.

## PgmNr 2957: Shrinkage methods utilising the estimation of error quantity to improve GWAS and polygenic risk score.

### Authors:

Y. Ruan<sup>1,2, 4</sup>; S. Choi<sup>2,3</sup>; P. O'Reilly<sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanley Center, Broad Institute, Cambridge, MA.; 2) IoPPN, King's College London, London, UK; 3) Icahn School of Medicine at Mount Sinai, New York; 4) Bio-X Institutes, Shanghai Jiao Tong University, Shanghai, China

---

GWAS have been used as a tool to discover the associated loci of polygenic traits and empowered downstream analysis such as Polygenic Risk Scores (PRS) to further investigate the aetiology. The effect size estimates of GWAS can be influenced by 'Winner's Curse' especially the top SNPs, which are more likely to be included in PRS calculation and other analyses. The widely-used methods utilise Bayesian approaches, LASSO etc. to generate a posterior distribution of effect sizes. Here, we developed a set of novel shrinkage methods that estimate error quantity in the observed effect size estimates from a null effect size distribution generated by permuting raw phenotype data ('Permutation Shrinkage') or order statistics ('Order Statistics Shrinkage') and subtracted the estimated null effects from the observed estimates. We tested our methods, 'Permutation Shrinkage' and 'Order Statistics Shrinkage', with quantitative traits in UK Biobank. Our methods lead to an average 35% relative increase of PRS  $R^2$ .

## **PgmNr 2958: 16S rRNA analysis of 17<sup>th</sup> and 18<sup>th</sup> century New York African burial ground grave soil samples to detect human microbiome signatures and infectious disease pathogen.**

### **Authors:**

C. Clinton <sup>1,2</sup>; C. Robinson <sup>1</sup>; F. Jackson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Biology, Howard University, Washington, District of Columbia.; 2) W. Montague Cobb Research Laboratory, Washington, District of Columbia.

---

In 1991, in Lower Manhattan, during the construction of a federal building the 400- year old New York African Burial Ground (NYABG) was unearthed. This cemetery is the largest and oldest site of free and enslaved African remains. Four hundred and four individuals were excavated and analyzed using archaeological, skeletal biological, and historical methods, then reburied at the historical site. At the time of excavation cadaver-associated soil samples were also retrieved creating the NYABG grave soil collection, currently maintained at Howard University. This investigation observes the bacterial community diversity of 74 unique burials from multiple body regions compared with 8 non-burial controls from Lower Manhattan. Comparing burial vs non-burial samples, we see increased detection of species abundance in the Firmicutes phylum indicating a pronounced signature for human microbiome. In addition, this investigation identifies bacterial disease pathogen associated with specific burials suggesting that the individual suffered from this infection in life and it may have contributed to their death. For example, bacteria genera associated with *Shigella* (dysentery), *Blautia* (present in the GI tract of cholera patients) and *Legionella* (Legionnaires disease) have been detected in burials 277, 245, and 230 respectively. These findings give profound insight to the quality of life and possible cause of death of NYABG inhabitants. Consequently, we are able to deduce information about living conditions of Africans in Lower Manhattan in the 17<sup>th</sup> and 18<sup>th</sup> centuries. This project helps us to learn more about the lives of free and enslaved Africans and the occurrence of pandemic outbreaks in the New York City area.

## **PgmNr 2959: A powerful microbial group association test based on the higher criticism analysis for sparse microbial association signals.**

**Authors:**

H. Koh; N. Zhao

[View Session](#) [Add to Schedule](#)

**Affiliation:** Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland.

---

The microbial group association analysis has been widely conducted to evaluate the association between microbial group (e.g., community or clade) composition and a host phenotype (or any other health/disease-related factor) of interest. In response, a variety of microbial group association tests have been proposed while incorporating the unique features of the microbiome data (e.g., high-dimensionality, compositionality, phylogenetic relationship). These tests generally fall in the class of aggregation tests which amplify the overall group association by combining all the underlying microbial association signals; as such, they are powerful when many microbial species are associated (i.e., low sparsity). However, in practice, the microbial association signals can be highly sparse, and this is especially the situation where we have a difficulty to discover the microbial group association. Hence, here we introduce a powerful microbial group association test for sparse microbial association signals, namely, microbiome higher criticism analysis (MiHC). MiHC is a data-driven optimal test taken in a search space spanned by tailoring the recent higher criticism test to incorporate phylogenetic information and/or modulate sparsity levels. Our simulations show that MiHC maintains a high power at different phylogenetic relevance and sparsity levels with correct type I error controls. We also demonstrate the use of MiHC with tree real data applications to test the disparity in respiratory-tract microbiome by smoking status, the disparity in infant's gut microbiome by delivery mode and the disparity in gut microbiome by type 1 diabetes status. In summary, MiHC is attractive because of its high adaptivity to diverse phylogenetic relevance and sparsity levels, which are usually unknown in practice.

## PgmNr 2960: Comparison of imputation quality for an Arab population using different references and methods.

**Authors:**

K. Kunji; M. Saad

[View Session](#) [Add to Schedule](#)**Affiliation:** Qatar Research Computing Institute, Hamad Bin Khalifa University, Doha, Qatar, Qatar

---

There has been a significant gap between research on European genetics and that of other ancestries. The ExAC dataset is 54.97% European and the Telenti et. al 10,000 deeply sequenced genomes are 78.55% European. There are efforts underway to address this data bias, e.g. the work of the Consortium on Asthma among African-ancestry Populations in the Americas, the Personal Genome Project's 10,000 Korean genomes goal, etc. Despite these, there has been a relative dearth of data and analysis of Middle Eastern populations. Here we attempt to identify which methods, reference panels and Genome-Wide Association Study (GWAS) SNP arrays that perform well for a Qatari cohort.

We used a public set of 108 Qatari Whole Genome Sequences (WGSs). We formed three GWAS arrays by keeping only the SNPs present in two GWAS panels from Illumina: (1) The Multi-Ethnic AMR/AFR-8 Kit (AMR\_AFR), and (2) Infinium Multi-Ethnic EUR/EAS/SAS-8 Kit (EUR\_EAS\_SAS) panels, and (3) we also considered selecting SNPs at random. Two imputation methods were performed on each GWAS data, Minimac and the recent Positional Burrows-Wheeler Transform (PBWT) using two reference populations, 1KGP3 (1000 Genomes Phase 3) and Haplotype Research Consortium (HRC). We find that our best results are achieved using the 1KGP3 reference with the EUR\_EAS\_SAS GWAS panel and Minimac. Minimac combined with the 1KGP3 reference performed particularly well, and Minimac was consistently the better imputation method. The EUR\_EAS\_SAS panel did slightly better than the AMR\_AFR panel, but the random SNP selection performed worse than either panel. Rare SNPs were imputed markedly well.



## PgmNr 2961: Quantifying the loss of accuracy in trans-ethnic genetic prediction.

### Authors:

Y. Wang<sup>1</sup>; P. Visscher<sup>1</sup>; L. Yengo<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia; 2) Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia

---

Polygenic risk scores (PRS) have been widely used to predict genetic component of complex traits or risk of diseases using variants identified from genome-wide association studies (GWASs). However, to date, the predictive performance of PRS has predominantly been evaluated in populations of European ancestry. The prediction accuracy of PRS has been found to be substantially reduced when the target population has a different genetic background from the GWAS discovery population. Here we derive a theoretical model to quantify such a loss of accuracy (LOA) of PRS and then use simulation studies and real data analyses to evaluate the model. We assume that: 1) the causal variants are shared, and 2) their allele substitution effects ( $\alpha$ ) are constant between populations. The model is a function of population genetic parameters including  $S$  measuring the relationship of  $\alpha$  and minor allele frequency (MAF), linkage disequilibrium (LD) correlations between the observed SNP markers and the causal variants, and MAF of the observed SNP markers and the causal variants between populations. It only requires summary statistics from GWAS and a reference panel to calculate the LD correlations and MAFs. All the analyses are conducted using the real genotypes of UK Biobank, including unrelated participants from four homogenous populations, namely, self-reported British ( $N = 323,284$ ), South Asian ( $N = 9,448$ ), East Asian ( $N = 2,257$ ) and African ( $N = 7,015$ ). A subset of 10,000 British individuals are randomly sampled as a target population. We perform simple linear association tests on  $\sim 0.98$  Million HapMap3 SNPs in the remaining 313,284 British and apply “LD clumping” algorithm to select approximately independent genome-wide significant SNPs, which are used to build the genetic predictors in all target populations. In simulations we vary trait heritability, the number of causal variants and  $S$ , and show that the LOA in non-European descents can be accurately predicted by our theory when ethnically-diverse populations differ in LD structure and MAF. In the analysis of height and BMI, we find that more than half of LOA (58% and 65%, respectively) in African can be explained by our theory. Our results highlight the necessity of collecting large scale of genome-wide data across different ancestries to achieve a complete understanding of such a LOA in transeethnic prediction; moreover, to fulfil the potential and equitable use of PRS in the precision medicine era.

## PgmNr 2962: Localization of rare variants in loci previously identified in aggregate rare variant testing.

### Authors:

R. Blumhagen <sup>1,2</sup>; D. Schwartz <sup>3</sup>; C. Langefeld <sup>4</sup>; T. Fingerlin <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genes, Environment and Health, National Jewish Health, 1400 Jackson St., Denver, CO, 80206, USA; 2) Department of Biostatistics and Informatics, Colorado School of Public Health, 13001 E. 17th Place, Aurora, CO, 80045, USA; 3) School of Medicine, University of Colorado, 13001 E. 17th Place, Aurora, CO, 80045, USA; 4) Wake Forest School of Medicine, Medical Center Boulevard, Winston-Salem, N.C., 27157, USA

---

Studies that examine the role of rare variants in both simple and complex disease are increasingly common. Though the usual approach of testing rare variants in aggregate sets is more powerful than testing individual variants, it is of interest to identify the variants that are plausible drivers of the association. We present a novel method for prioritization of rare variants after a significant aggregate test. Our approach leverages quantification of the influence of the variant on the aggregate test of association. In addition to providing a measure used to rank variants, we use outlier detection methods to present a computationally efficient approach to identify a subset of variants that is driving the disease association. We evaluated several outlier detection methods that vary based on the underlying variance measure: interquartile range (Tukey fences), median absolute deviation and standard deviation. We performed 1000 simulations for each region and compared the true and false positive rates. All outlier detection methods observed higher sensitivity to detect uncommon variants ( $0.001 < \text{MAF} < 0.03$ ) compared to very rare variants ( $\text{MAF} < 0.001$ ). For uncommon variants, the inner Tukey had the highest median true positive rate of 0.60 (IQR: 0.51, 0.87) compared to the standard deviation having a median of 0.30 (IQR: 0.10, 0.59). We compared our method using the inner Tukey to two existing methods: adaptive combination of p-values (ADA) and a Bayesian hierarchical model (BeviMed). For uncommon variants, the inner Tukey observed a lower median false positive rate of 0.10 (IQR: 0.06, 0.18) compared to the ADA, which observed a median false positive rate of 0.23 (IQR: 0.18, 0.36). ADA and the inner Tukey had significantly higher true positive rates than that observed for BeviMed. Finally, we applied this method to data from our targeted resequencing study in idiopathic pulmonary fibrosis (IPF). Within two aggregate sets including 100 rare variants, we identified six variants with the greatest evidence for influencing the association with IPF. In summary, our localization methods to classify risk variants obtain a high true positive rate while maintaining a low false positive rate. This work provides an approach to obtain greater resolution of the rare variant signals within significant aggregate sets; this information can provide an objective measure to prioritize variants for follow-up experimental studies and insight into the biological pathways involved.

## PgmNr 2963: Admixture mapping of lung function in Puerto-Rican Islanders.

### Authors:

E. Lee <sup>1</sup>; A. Mak <sup>2</sup>; D. Hu <sup>2</sup>; K. Keys <sup>2</sup>; M. White <sup>2</sup>; S. Huntsman <sup>2</sup>; C. Eng <sup>2</sup>; J. Rodriguez-Santana <sup>3</sup>; M. Seibold <sup>4</sup>; E. Burchard <sup>1,2</sup>; NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioengineering and therapeutic sciences, UCSF, San Francisco, California.; 2) Department of Medicine, UCSF, San Francisco, California.; 3) Veterans Caribbean Health Care System, San Juan, Puerto Rico.; 4) Center for Genes, Environment and Health, Department of Pediatrics, National Jewish Health, Denver, Colorado

---

In the U.S., asthma disproportionately affects urban minority populations, with Puerto Ricans showing the highest asthma prevalence 16.1%. However, what causes high asthma prevalence among the Puerto Ricans is not very well understood. The varying proportions of African, European, and Native American genetic ancestry in Puerto Ricans can be leveraged to identify genetic determinants of clinically significant measures and outcomes such as lung function and asthma severity. We have previously demonstrated that genetic ancestry is associated with lung function.

In this current study, we sought to discover potential causal genes that may be associated with lung function in Puerto Rican islanders by applying admixture and fine mapping methods. A genome-wide admixture mapping study of lung function was conducted in 843 Puerto Rican children with and without asthma, who were recruited from Puerto Rico. After mapping admixture segments, we fine-mapped the genomic region using whole genome sequencing data. Functional annotation was then used to elucidate underlying biologic links between genetic variants and lung function.

The association test between local genetic ancestry and forced expiratory volume in 1 second (FEV<sub>1</sub>) identified a strong admixture mapping peak on chromosome 1q32, indicating that a one percent increase in African ancestry was associated with 0.13 liters decrease in the volume of exhaled air [95%CI: -0.07 to -0.18, p-value=1.3×10<sup>-6</sup>]. Variants identified from the fine-mapping analysis were associated with the inflammation-related genes (*INAVA* and *NEK7*). The minor allele of the intron variant associated with *NEK7* gene (rs111428936) tracks with European ancestry and is associated with well-controlled asthma and better lung function, whereas the major allele had the opposite effect.

The results from admixture mapping and functional annotations point a potential enhancer variant for the *NEK7* gene. It is possible that *NEK7* may be involved with inflammasome activation in the lung, which may influence the lung function and asthma severity of Puerto Ricans.

## PgmNr 2964: Correcting gene expression batch effect in single-cell RNA sequencing.

**Authors:**Z. Zhao <sup>1</sup>; M. Li <sup>2</sup>[View Session](#) [Add to Schedule](#)**Affiliations:**

1) Texas Academy of Mathematics and Science, University of North Texas, Denton, Texas.; 2) Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine

---

Single-cell RNA-sequencing (scRNA-seq) enables unbiased measurement of gene expression in individual cells, and provides unprecedented opportunities to characterize cell types and cellular heterogeneity. The continuously decreasing cost of scRNA-seq has made it possible to profile the transcriptomes of hundreds to thousands of cells simultaneously. Unfortunately, scientific discoveries from these abundant data are hampered by technical artifacts and inherent biological heterogeneity in the data. The former is usually termed “batch effects,” referring to systematic differences in gene expression from one batch to the other, and the latter is often modeled by different cell types. Batch effects are inevitable, particularly in human tissues, because samples have to be processed quickly, making it impossible to randomize the samples and cells in the experiment. Additionally, different laboratories might use different cell dissociation protocols, library preparation kits, or sequencing platforms. All of these could lead to unwanted technical differences across batches. Failure to remove batch effect will obscure downstream analysis and lead to erroneous findings. To correct for batch effect, a common strategy is to merge all scRNA-seq data together first, and then use regression models to regress out batch effect. After batch effect is removed, further analyses, such as cell type clustering and differential expression, can be performed. However, this global batch effect correction procedure may not be ideal for scRNA-seq data, because some cells are more vulnerable to batch effect than others. The global batch effect correction approach may over correct for some cells but less effective for other cells. To address this limitation, we propose a new procedure that distinguishes the cell types first, and then performs batch effect correction for each type of cells. Our simulation results demonstrate that this new procedure outperforms the traditional global batch effect correction approach. We show that it has well controlled type I error rates and high sensitivity in downstream differential expression analysis under a wide range of scenarios. With the growing popularity of scRNA-seq, we believe our procedure will offer a valuable tool for proper analysis of scRNA-seq data and aid scientific discoveries.

## PgmNr 2965: Graphical analysis using genetic variables for phenome-level causal discovery in population-scale biobanks.

### Authors:

D. Amar <sup>1,2</sup>; N. Sinnott-Armstrong <sup>3</sup>; E. Ashley <sup>1</sup>; M. Rivas <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Inherited Cardiovascular Disease, Stanford University, Stanford, California, United States of America; 2) Department of Genetics, School of Medicine, Stanford University, Stanford, CA, USA; 3) Department of Biomedical Data Science, Stanford University

---

Availability of large genetic databases has led to the development of powerful causal inference methods that use genetic variables as instruments to estimate causal effects. However, such methods were developed for analysis of a selected exposure-outcome pair and are limited when performing a phenome-level analysis (e.g., sensitive to biases and assumptions). Here, we present cGAUGE, a new pipeline for causal Graphical Analysis Using GEnetics that analyzes the genetic variables and many traits jointly to find patterns that carry causal information. Specifically, cGAUGE utilizes conditional independence (CI) tests to either filter out associations that are not likely to represent direct causation or to detect unique combinatorial patterns that serve as evidence for causal links that are justifiable even if the underlying causal diagram has cycles. Using these analyses, cGAUGE finds valid genetic instruments, directly searches for cycles in the underlying causal diagram, and detects cases where Mendelian Randomization (MR) can be performed with minimal risk of horizontal pleiotropy bias. For the best of our knowledge, cGAUGE is the first algorithm to detect such patterns. Moreover, for MR analysis cGAUGE scores the reliability of the causal link using a simple Bayes false discovery rate statistic. We present theoretical justification, simulations, and apply our pipeline to 96 complex traits from 337,198 subjects from the UK Biobank. We illustrate how CI filters remove around 40% of the GWAS results, highlighting genetic variants that are more likely to be proper instruments for MR. Our MR-based results cover 121 causal links supported by at least 10 instruments and detected at 1% FDR. Many of the links expected causal links, some of which are missed by extant methods, but also new associations previously speculated to be causal such as detrimental effects of depression on kidney and liver function, and a protective effect of immune system activity in prostate cancer. Our results also recapitulate a refined structure of the underlying causal diagram and highlight central hubs, such as sleep duration, that indirectly affects many biomarkers, traits, and diseases.

## PgmNr 2966: Dealing with confounding in causal network inference for biological data.

### Authors:

M.B. Badsha; A.Q. Fu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Statistical Science, Center for Modeling Complex Interactions, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, Moscow, ID 83844-1104

It is a challenge to reliably infer causal relationships among multiple molecular phenotypes (such as gene expression) when many other variables (demographic and hidden variables) may bias the inference. When genotype data are available, we can employ the principle of Mendelian randomization (PMR) for causal inference. As a randomization principle, the PMR assumes that the alleles of a genetic variant (which is an instrumental variable) are randomly assigned to individuals in a population, analogous to a natural perturbation experiment, which helps establish causality.

We take a graphical model approach to causal inference. Our method, namely MRPC, learns a causal network from integrating genotype and molecular phenotype data at the individual level, where directed edges indicate causal directions. MRPC implements an online false discovery rate (FDR) control method to control the overall error rate and incorporates the PMR (Frontiers in Genetics, 2019; arXiv, 1806.01899). It can capture a wide variety of causal relationships, beyond the canonical causal model ( $V \rightarrow X \rightarrow Y$  in which  $V$  is a genetic variant,  $X$  a regulator, and  $Y$  a target).

With a graphical model, we can explicitly include confounding variables as additional nodes into the network, without making assumptions on how they affect the variables of interest. This implies that we can check whether the conditions in the PMR are satisfied while performing causal inference when other methods often assume these conditions to be true. If a large number of confounding variables are present, we derive lower-dimensional surrogate variables (e.g., top principal components of the entire gene expression matrix and demographic variables) as composite confounding variables.

We took this approach to dissect direct and indirect targets among multiple genes associated with the same eQTLs. The GEUVADIS consortium has identified 62 eQTLs in LCLs to be associated with more than one gene. We applied MRPC to each eQTL-gene set and incorporated the top principal components into the analysis. We identified 10 types of causal networks between the eQTL and its associated genes. The presence of confounding variables did not alter the inference. Among these eQTL-gene sets, 46 also had data in the GTEx LCL samples and the causal networks inferred for 30 of them were replicated in GTEx.

## **PgmNr 2967: Wilson-interval corrected population frequency estimates for reduced-coverage areas in population reference databases.**

### **Authors:**

A.S. Brown <sup>1</sup>; B.N. Pusey <sup>1</sup>; N.J. Balanda <sup>1</sup>; W.A. Gahl <sup>1,2</sup>; U.D.N. Undiagnosed Diseases Network <sup>3</sup>; D.R. Adams <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Undiagnosed Diseases Program, NIH, Bethesda, MD; 2) Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 3) Common Fund, Office of the Director, National Institutes of Health

---

**INTRODUCTION:** Population-frequency criteria are frequently used to target disease causing variants in clinical genome and exome datasets. A common approach removes variants above a defined cutoff, for example 1% frequency, based on a hypothesis that a severe, rare, highly-penetrant Mendelian disorder will be unlikely to be caused by more frequent variants derived from a generally healthy population. Variant frequencies vary by population. Evidence that a variant is overly frequent may be derived from a population with relatively few genotyped individuals. We hypothesized that the frequency of a minority of potentially clinically-relevant DNA variants is overestimated in available population databases, thereby causing them to be excluded from analysis. We further hypothesized that adjusting the frequency to include these variants would increase the sensitivity of potentially pathogenic variant identification in our genetic data analysis procedures.

**METHODS:** We obtained frequency data from the widely used gnomAD database (Lek, et al 2016, PMID: 27535533). A Wilson score interval was used to generate a frequency-estimate confidence interval for areas with a low number of genotyped alleles. The Wilson score interval (Wilson, 1927, DOI: 10.2307/2276774) is a statistical tool designed to find the interval that is likely to contain the true mean given number of observations and number of successes. It is particularly suited to cases with few observations (Newcombe, 1998, PMID: 9595616). Frequency estimates from gnomAD were compared with estimates based on the lower frequency bound from the Wilson calculation.

**RESULTS:** In regions with less than 300 observed alleles, there were 187,258 variants with a 1% allele frequency or less according to the Wilson interval. The gnomAD allele frequency at this level of observation shows 171,385 variants. The difference between these groups (17,313 variants) represent variants that would have been discarded using a 1% frequency threshold based on the reported gnomAD frequency.

**DISCUSSION:** Population frequencies are a powerful means of focusing attention on a subset of important variants during clinical genomic analysis. We present a procedure for revising frequency estimates that may be inappropriately high given the underlying number of successfully genotyped alleles. The presented data will further explore the overlap between the set of revised alleles and datasets identifying known pathogenic variants.



## PgmNr 2968: Visualization of relationship inference in large-scale genetic studies.

### Authors:

W.-M. Chen; Z. Zhu; C.C. Robertson; S. Onengut-Gumuscu; S.S. Rich; A. Manichaikul

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Public Health Sciences & Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia.

---

Relationship inference is fundamental to genetic data and analyses. Recent advances in computational algorithms have made it feasible to infer relatedness in > 1 million samples, *e.g.*, using our software tool KING. Here we present visualization techniques to facilitate and enhance relationship inference in large-scale genetic studies. The concept underlying the visualization of relationship inference is that the inferred relationships can be displayed in undirected graphs, with nodes representing samples, and colored edges representing different types of relationships (*e.g.*, using the R package igraph), with the resulting pedigree evaluation visualized in classical pedigree structure (*e.g.*, using the R package kinship2). Our visualization processes are as follows: 1) To visualize pedigree errors in a family, we display documented relationships in a classic pedigree structure, next to inferred relationships in an igraph plot, with inconsistent relationships highlighted; 2) To visualize a summary of all documented and cryptic relationships, we utilize the powerful igraph operations on subgraphs (*i.e.*, families) and visualize all unique family configurations (such as sib pairs, trios, *etc.*) in one graph, together with the counts of each configuration; 3) Reconstructed families are visualized in classic pedigree structures; 4) Clustered families are visualized in igraph plots; 5) Genomes that contain runs of homozygosity are visualized using the ggplot routines; 6) To visualize potential sample swaps for a set of samples that are genotyped or sequenced with multiple platforms/technologies, we use a directed igraph to connect samples with identical genomes but different names. The visualization of mismatch patterns in a directed igraph provides additional information regarding the possible error mechanisms (*e.g.*, sample switch in pairs, one sample shift, *etc.*). All visualization techniques are implemented in R and integrated in KING *via* the “rplot” option. Further, our efficient implementation in KING makes all of the described visualizations feasible for biobank-scale data. We demonstrate various visualization techniques in several large-scale datasets. Data visualization in KING is also detailed in the KING tutorial at <http://people.virginia.edu/~wc9c/KING/KINGvisualization.html>. These newly implemented visualization algorithms allow KING to communicate results more effectively with the software users, providing a better understanding of their genetic data.



## PgmNr 2969: Improving polygenic risk prediction by incorporating LD information from multi-ethnic Biobank data.

### Authors:

W. Chung<sup>1,2</sup>; L. Liang<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston, Massachusetts.; 2) Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts.; 3) Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts.

---

Although various polygenic risk prediction methods have been put forward in recent years, most available methods are inadequate for genetic risk prediction for complex traits or diseases in non-European populations because they utilize only a single population, mostly European ancestry, for both training and validation datasets. Here, we introduce a novel penalized regression based polygenic risk prediction method for cross-ethnicity studies. We introduce a new penalty function to incorporate the different LD structure across multiple ethnicities and evaluate the predictive performance using the Lasso and the minimax concave penalty (MCP) to induce a sparsity solution. These functions can improve the prediction accuracy for complex traits of non-European ancestry (primary ethnicity) using European ancestry (secondary ethnicity) in large sample size. Our novel implementation of a parallel computing algorithm makes it feasible to apply our method to biobank-scale multi-ethnic GWAS data. Furthermore, they can take advantage of the secondary ethnicities based on summary statistics and thus have the potential to utilize information from most published GWAS summary statistics. We illustrate the method using Chinese GWAS datasets with 1000 Genomes imputation through collaboration with Tongji Medical College, including: training (N=7,417) and testing (N=1,452). We also picked 7,417 individuals from UK Biobank as training set. LD score (LDSC) were estimated using 1000 Genomes EUR and EAS populations. LDD (LD difference) is defined as  $LDSC_{EUR} - LDSC_{EAS}$ . We first simulated SNPs with relatively small LD difference between the UK and Chinese populations. We sampled 5000 SNPs (2K with LDD >0.02, 2K with LDD < -0.03 and 1K in middle). 1,000 causal SNPs (20%) were randomly selected from the 5,000 SNPs and SNP-heritability was set to 50%. We then simulated SNPs with larger LD difference between the UK and Chinese populations. We sampled 5000 SNPs (2K with LDD >0.03, 2K with LDD < -0.05 and 1K in middle). The simulation results clearly shows the advantage of using a cross-population penalty over a single population method. When LD difference is relatively small, UK trained model + China summary statistics even outperform trained models based on Chinese individual data.

## PgmNr 2970: Methylome-wide association study identifies CpG sites associated with 30 complex traits.

### Authors:

J.J. Fryett <sup>1</sup>; A.P. Morris <sup>2</sup>; H.J. Cordell <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Genetic Medicine, Newcastle upon Tyne, United Kingdom; 2) School of Biological Sciences, University of Manchester, Manchester, United Kingdom

---

Transcriptome-wide association studies (TWAS), in which gene expression is imputed from SNP genotypes and then tested for association with a trait of interest, have become a popular post-GWAS approach for elucidating the role of gene expression in complex traits. This methodology can also be extended to other complex biological intermediates, such as CpG methylation. Here, we explored how well CpG methylation can be predicted from SNP genotypes, and conducted a methylome-wide association study (MWAS) of 30 complex traits.

Using data from the Avon Longitudinal Study of Parents and Children (ALSPAC), we investigated how well three penalised regression methods - ridge regression, elastic net and LASSO - predicted CpG methylation from local SNP genotypes. Methylation was poorly predicted at most CpG sites (150,661 of 173,769 CpGs showed prediction  $R^2 < 0.1$ ), although for a small subset, it was relatively well predicted (2486 CpGs showed  $R^2 > 0.5$ ). Elastic net and LASSO performed similarly, both outperforming ridge regression, indicating a potential sparse underlying genetic architecture of CpG methylation.

Following identification of CpGs that can be predicted well enough from SNP genotypes, we then constructed prediction models for 65,853 CpG sites (showing  $R^2 > 0.01$ ) and applied them to publicly available genome-wide association study (GWAS) summary data for 30 complex traits in an MWAS. We identified 7214 predicted methylation - trait associations at 4269 unique CpGs, including CpGs that tag known GWAS risk genes. To further elucidate the function of these CpGs, we tested association between imputed methylation levels and imputed expression of nearby genes, finding 101 associations between 64 CpGs and 34 genes, providing insight into the interplay between CpG methylation, gene expression and complex traits.

We conclude that MWAS represents a powerful method for investigating the role of CpG methylation in complex traits, and that further application and integration with other data (including gene expression) may help improve understanding of the biological mechanisms underlying GWAS signals.

## PgmNr 2971: Rare variant association analysis for studies with rich phenotyping subject to missingness.

### Authors:

S. Gaynor<sup>1</sup>; Z. McCaw<sup>1</sup>; X. Lin<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, Harvard University, Boston, Massachusetts.; 2) Department of Statistics, Harvard University, Cambridge, Massachusetts.

---

Many large-scale genome-wide association studies, such as TOPMed, pair rich phenotyping with whole genome sequencing; however, not all phenotypes are measured in all studies or for all individuals. Existing rare variant association methods only allow for the use of completely observed phenotyping, and therefore do not take full advantage of studies with broad but incomplete phenotyping. We propose kernel association tests based on a multivariate regression framework to test for multi-phenotype associations and phenotype-specific associations. We consider jointly modeling related phenotypes in a multivariate linear model adjusting for population structure for quantitative phenotypes, which is extended to a mixed model to account for relatedness using a sparse genetic relatedness matrix. We allow for missing data in phenotypes by considering the likelihood contribution of all observations, given that any of the modeled phenotypes are observed. We specifically allow for the use of all available data by performing estimation through the expectation conditional maximization algorithm. We demonstrate in simulation the Type 1 error rates and power of the proposed tests and compare them to existing rare variant association methods that only allow for complete phenotype data. We simulate settings with different proportions of missingness and correlation structures between phenotypes. We further apply our methodology to large-scale sequencing efforts to study related pulmonary phenotypes that have been incompletely observed. Pulmonary traits have been shown to have shared genetic architecture among common variants, high statistical correlation, and related biological mechanisms. We model pulmonary and related traits jointly and test for trait-specific and multi-trait associations. Our methods allow for the complete use of all available data to empower statistical tests for rare variants by better purposing the available data, rather than relying exclusively on growing whole genome sequencing study sample sizes to improve power.

## PgmNr 2972: Increased *SLC6A14* expression is associated with reduced cystic fibrosis lung function mediated by increased susceptibility to *P.aeruginosa* infection.

### Authors:

G. He <sup>1,2</sup>; M. Esmaili <sup>1</sup>; C. Wang <sup>1</sup>; N. Panjwani <sup>1</sup>; F. Wang <sup>1,3</sup>; J. Gong <sup>1</sup>; K. Keenan <sup>1</sup>; J. Avolio <sup>4</sup>; F. Lin <sup>1</sup>; L. Sun <sup>2,3</sup>; J.M. Rommens <sup>1,5</sup>; L.J. Strug <sup>1,2,3,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and genome biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 4) Program of Translational Medicine, Hospital for Sick Children, Toronto, Ontario, Canada; 5) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 6) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada

Progressive lung disease, the major concern for morbidity and mortality in cystic fibrosis (CF), results from cycles of infection and inflammation and is modulated by genes in addition to the CF transmembrane conductance regulator (*CFTR*). A genome-wide association study (GWAS) by the International CF Gene Modifier Consortium (ICFGMC) identified a CF lung disease locus on chromosome X between *SLC6A14* and *AGTR2* (Corvol, *et al*, 2015). We sought to understand how the associated non-coding SNPs contribute to CF lung disease.

First, we established that human primary nasal cells (HNE) are a suitable airway model to study lung disease modifiers in CF by comparing RNA-seq of HNE to the gold standard human primary bronchial (HBE) model from 13-paired CF samples. We observed a strong correlation in gene expression between the two airway models. Notably, the correlation patterns of *SLC6A14* expression with apical constituents (including *CFTR*) that were considered relevant epithelial factors were consistent between HNE and HBE ( $p=2.4E-5$ ). *AGTR2* showed no expression in either cell type.

We then conducted expression quantitative trait loci (eQTL) analysis using RNA-seq of HNE from 70 Canadian individuals with CF. The *SLC6A14* eQTL pattern co-localized with the ICFGMC GWAS lung disease results using the *Simple Sum* colocalization method ( $p\text{-value}=9.5E-9$ ; Gong *et al*, 2019). The finding supports that the associated SNPs impact lung disease through the expression of *SLC6A14* with increased *SLC6A14* expression corresponding to reduced lung function.

The expression of *SLC6A14* and *CD45* from the HNE were correlated ( $\text{corr}=0.68$ ,  $p=7.2E-10$ ). *CD45* is a marker of immune cell composition in RNA-seq of HNE (Polineni *et al*, 2018) and thus its strong correlation with *SLC6A14* suggests increased *SLC6A14* expression may be associated with susceptibility to bacterial infections. We carried out an association analysis between age of first *P. aeruginosa* (PsA) infection and variants at the ICFGMC lung disease GWAS locus in 1,522 CF Canadians and observed an association pattern similar to the lung disease GWAS. A mediation analysis demonstrated that the age of first PsA is a significant mediator ( $p=0.007$ ) between the top *SLC6A14* eQTL and lung disease, corresponding to 34% of the total effect.

Together, this study suggests that increased *SLC6A14* expression results in reduced lung function in CF patients that manifests with earlier onset of PsA infection.

## PgmNr 2973: Leveraging genetic ancestry for new insights into complex traits in admixed populations.

### Authors:

A.R.V.R. Horimoto <sup>1</sup>; N. Franceschini <sup>2</sup>; T.A. Thornton <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 3) Department of Statistics, University of Washington, Seattle, WA

---

Genetic studies in multi-ethnic cohorts offer great potential to elucidate the genetic factors influencing complex traits. A variety of statistical methods have recently been developed to overcome special challenges for whole genome association analysis of complex traits in large-scale multi-ethnic cohorts. With existing methodology, however, genetic ancestry differences among sampled individuals are often treated as a confounder to be adjusted for in an analysis to protect against spurious association. Leveraging genetic ancestry can provide improved complex trait mapping in multi-ethnic populations, such as African Americans and Hispanic/Latino populations, who have admixed ancestry derived from multiple continents. We have developed mixed effects models for admixture mapping that incorporate both local and global ancestry for the identification of genetic loci influencing complex traits. The proposed mixed model admixture mapping methods have been developed for continuous and dichotomous traits, and are completely applicable to large-scale whole genome studies with multi-ethnic samples from a variety of study designs. We demonstrate the utility of leveraging genetic ancestry for complex trait mapping in applications to the Hispanic Community Health Study / Study of Latinos (HCHS/SOL) to elucidate the genetic determinants of glomerular filtration rate (GFR), a trait associated to chronic kidney disease. The genome-wide admixture mapping was performed on 12,702 samples using our proposed linear mixed model with a joint test for the local ancestry count from all European, African and Native American ancestries. We identified a significant signal on chromosome 15 ( $p\text{-value} < 5.7 \times 10^{-5}$ ) in an admixture-linkage disequilibrium block of 37 variants in the *RYR3* gene, a gene implicated in cellular calcium signaling. Single tests into each ancestral group separately showed that this signal is driven by the African ancestry. The genome-wide association mapping on about 2,5 million SNPs using a similar linear mixed model was able to capture significant association signals in three chromosomes, but not in chromosome 15. These results emphasize the great potential of admixture mapping as an alternative approach to the traditional genome-wide association studies for the genetic mapping of complex traits in multi-ethnic population.

# PgmNr 2974: Robust genome-wide association tests in the presence of latent genetic interactions: Application to cystic fibrosis lung disease.

## Authors:

S.O. Kim <sup>1,2</sup>; B. Xiao <sup>2</sup>; N. Panjwani <sup>2</sup>; K. Keenan <sup>2</sup>; J.M. Rommens <sup>2,3</sup>; L.J. Strug <sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, ON; 2) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON; 3) Department of Molecular Genetics, University of Toronto, ON; 4) Department of Statistical Sciences, University of Toronto, ON

---

**Introduction:** For complex traits such as lung disease in Cystic Fibrosis (CF), Gene x Gene or Gene x Environment interactions can impact disease severity but these remain largely unknown. Unaccounted-for genetic interactions introduce heterogeneity in the variance of the quantitative trait across the genotypic groups (Pare et al., 2010); thus, a GWAS using variance testing can identify variants putatively involved in genetic interactions. Joint tests of both mean and variance (or full distributional differences) across genotype groups can account for unknown genetic interactions and increase power for gene identification. Aschard et al. (2013) proposed a variant of the Kolmogorov-Smirnov test but its computational requirements limit implementation genome-wide. Parametric joint tests that assume a Gaussian trait under general models (joint location-scale test (JLS); Soave et al., 2015) or under additive assumptions (heteroskedastic linear model (HLM); Yonge et al., 2018) are accessible for GWAS, but under departures from normality, these can suffer from type I error inflation or be inefficient, even after normalization.

**Method:** Here, we overcome previous limitations by developing a quantile regression-based JLS method (qJLS) without any distributional assumptions, is robust to outliers, detects linear and non-linear changes in mean and variance and in the shape of the distribution and is computationally efficient for GWAS. A simulation study compares the statistical power of qJLS to the JLS and HLM methods, investigating the effect of varying interaction magnitudes and phenotypic distributions including skewed, heavy-tailed and count data. We apply the qJLS in a GWAS of CF lung disease in the Canadian CF Gene Modifier consortium (n=2,835).

**Results:** The simulation study shows that qJLS is more powerful than JLS and HLM with skewed or heavy-tailed distributions, even with a quantile normalization. Under the Gaussian trait scenario for which the JLS and HLM were designed, JLS and HLM are more powerful. Type I error was well-controlled for all methods; without normalization, JLS and HLM showed inflated type I error under non-normality.

**Discussion:** The qJLS statistic for GWAS is an alternative to conventional mean-based tests in the presence of unknown or unmeasured genetic or environmental interactions, is applicable and more powerful for non-normal phenotypic distributions (e.g. those that are skewed or heavy-tailed) and in the presence of outliers.

## **PgmNr 2975: Meta-prediction of gene expression levels from genotypes across multiple tissues and datasets using summary-level data.**

### **Authors:**

A. Liu; H.M. Kang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Biostatistics, Univ MICHIGAN, Ann Arbor, Michigan.

---

Transcriptome wide association studies (TWAS) is a powerful method to interpret the underlying biological mechanisms underlying GWAS loci by mapping gene expression levels with phenotypes. In TWAS, gene expression is often imputed from individual-level data trained from external resources, such as GTEx, and several approaches shared the prediction models as a summary-level data (e.g. eQTLs and effect sizes). Because the eQTL discovery can be limited by the sample size, which in turn, limits the power of the downstream TWAS, we developed an integrative approach to leverage the abundance of prediction models trained from multiple tissues and datasets. Here, we propose SWAM (Smartly Weighted Averaging across Multiple datasets), a meta-prediction method to improve prediction accuracy by integrating relevant models across tissues and datasets. Instead of training directly from a single reference data, SWAM aggregates existing prediction models, calculating expression using an optimal linear combination of tissues for each gene. As a result, tissue-specific expression calculated from SWAM is not limited by the sample size of the target tissue but leverages summary-level information from multiple datasets without requiring raw data.

To demonstrate the improvement in prediction accuracy, we applied our method to predictDB's single-tissue lymphoblastoid cell line (LCL) prediction model to obtain meta-LCL model. We then used our meta-LCL model to impute tissue-specific expression for 344 European samples from the GEUVADIS consortium. Overall, we were able to predict 2,948 genes at a false discovery rate (FDR) < 0.05, a significant improvement over the 1,809 genes from the prediXcan single-tissue model. SWAM also outperformed another cross-tissue model UTMOST, which requires raw data, and found 2,373 predictable genes at the threshold. We also extend our method to TWAS, using metaXcan and summary GWAS statistics for many traits, including HDL, LDL from GLGC. We found that for most tissues, there was a 1.5~4 fold increase in the number of significant signals using SWAM compared to single-tissue models, even after accounting for stricter transcriptomic-wide threshold. For example, the liver tissue from GTEx which has small sample size, SWAM detected 71 genes compared to 16 using single-tissue liver. Our results demonstrate that meta-prediction from summary-level data can substantially improve the power of TWAS compared to using a single trained model.



## PgmNr 2976: A Bayesian graphical model approach to inferring gene regulatory networks.

### Authors:

E.A. Martin <sup>1</sup>; A.Q. Fu <sup>1,2</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Bioinformatics and Computational Biology, University of Idaho, Moscow, Idaho.; 2) Statistical Science, University of Idaho, Moscow, Idaho

---

**Introduction:** Gene regulatory networks have been studied extensively in genomics to understand the relationship among genes. In a gene regulatory network, a directed edge between two gene nodes may be interpreted as a causal (i.e., regulatory) relationship. Many methods have been developed to infer regulatory networks directly from genomic data. However, due to noise in the data not all edges in the inferred network have the same level of confidence and none of the existing methods quantify such uncertainty. Additionally, existing methods often ignore potential confounding variables. **Methods:** We have developed a novel Bayesian approach for gene regulatory network inference. Our method, baycn, estimates the posterior probability of the state (presence/absence and direction) for each edge and provides a coherent framework to account for confounding variables. We develop a Markov Chain Monte Carlo algorithm for inference that accepts a candidate graph as input, which helps reduce the search space to a more manageable size. The candidate graph may be inferred by another, much faster, network inference method and may contain false edges or undirected edges. We incorporate into the algorithm the Principle of Mendelian Randomization (PMR), which is a randomization principle that helps establish causality from observational genomic data. To do so, we impose the constraint that genetic variant nodes are always parents of gene expression nodes. We further include confounding variables as nodes in the network to examine whether the conditions of the PMR are satisfied. **Results:** We performed extensive simulations under multiple networks with varying sizes and structures of the network, variable types (e.g., genotype and gene expression) and types of confounding variables. Posterior probabilities for these networks are close to what we would expect accounting for Markov equivalence, which means that multiple graphs of the same likelihood may exist. Our method can also identify false positive edges and performs well in the presence of multiple confounding variables.



## PgmNr 2977: Effects of reference panel composition on genotype imputation quality in founder samples.

### Authors:

K. Ryan<sup>1</sup>; C. Van Hout<sup>2</sup>; J. O'Connell<sup>1</sup>; J. Perry<sup>1</sup>; B. Gaynor<sup>1</sup>; H. Xu<sup>1</sup>; M. Montasser<sup>1</sup>; T. O'Connor<sup>1,3</sup>; N. Gosalia<sup>2</sup>; A. Shuldiner<sup>2</sup>; B. Mitchell<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program for Personalized and Genomic Medicine, Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland; 2) Regeneron Genetic Center, Regeneron Pharmaceuticals, Inc., Tarrytown, New York; 3) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland

---

Genotype imputation, the process of leveraging haplotypes observed in a reference panel to infer unobserved genotypes in other samples, is widely used in genetic studies. The performance of imputation depends on the representativeness of haplotypes present in the reference panel, which is a function of the number and diversity of haplotypes in the reference panel. In this study we evaluate imputation quality for DNA samples from a founder population (Amish) as a function of the relative mix of founder and nonfounder haplotypes in the reference panel.

We evaluated imputation performance using 3 different imputation reference panels for 6,153 Amish individuals genotyped on the Illumina Global Screening Array. Genotypes were imputed using the TOPMed reference panel that included 1,025 Amish whole genome sequences (n=54,035 subjs; TOPMed-global), or the one excluding the Amish (n=53,010 subjs; TOPMed-noAmish), or the one with Amish only (n=1,025; Amish-only). We evaluated imputation performance and quality across the 3 reference panels by comparing the number of imputed SNPs across panels and assessing concordance of imputed genotypes with genotypes obtained using whole exome sequencing (n=5,317).

All 3 panels imputed approximately the same number of common (MAF>5%) SNPs. Far more rare SNPs (MAF<0.5%) were imputed using the TOPMed-noAmish reference panel (n=1,929,020) than the Amish-only (n=126,510) and TOPMed-global (n=392,323) panels. Imputed genotypes from the TOPMed-noAmish panel also had much lower non-reference allele concordance rates compared to Amish-only and TOPMed-global (88.6% vs. 95.7% and 94.9% respectively for MAF 0.5 to 5%). These results suggest that the use of a cosmopolitan panel without a representation of the founder population could overcall false-positive rare variants (i.e., impute genotypes that may not exist) and support the further development of whole genome sequence reference panels for founder populations.

## PgmNr 2978: A unified framework for gene-based and multi-trait colocalization testing: Application to cystic fibrosis.

### Authors:

F. Wang<sup>1,2</sup>; C. Wang<sup>2</sup>; J. Gong<sup>2</sup>; N. Panjwan<sup>2</sup>; G. He<sup>2,3</sup>; S. Mastromatte<sup>2</sup>; L. Sun<sup>1,3</sup>; L. Strug<sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

To understand the mechanism by which genome-wide association study (GWAS) loci contribute to disease and to assess their contribution across traits demands integration of omics data from several sources. For example, integration of GWAS summary statistics with expression quantitative trait loci (eQTL) can inform whether a GWAS locus is driven by variation in gene expression and can pinpoint the responsible gene and tissue.

We previously developed a colocalization test, the *Simple Sum* (SS; Gong et al, 2019), that assesses pattern similarity between GWAS and eQTL statistics. The SS is of the form  $Z'AZ$ , with Z containing the GWAS statistics and matrix A incorporating eQTL or other omics data; p-values of SS can be computed analytically. When A is the identity matrix, the test statistic simplifies to a gene/set-based test. When A represents GWAS statistics for a second phenotype, the SS measures genetic correlation between phenotypes.

However, in a given GWAS region with significant association but, for example, no eQTLs, the SS could have type I error inflation, thus we previously restricted SS analyses to regions with sufficient eQTL evidence. To facilitate genome-wide implementation of SS as a general integration tool, here we propose an adaptive shrinkage method in the SS (*SS-shrink*) that is similar to the weighted lasso (Bergersen et al. 2011).

Simulation shows that *SS-shrink* has correct type I error control for all settings and improved power in the presence of linkage disequilibrium (LD) and allelic heterogeneity when compared to *COLOC* (Giambartolomei et al. 2014) and *eCAVIAR* (Hormozdiari et al. 2016). Several analyses in the International Cystic Fibrosis Gene modifier cohort (ICFGMC; n>6,500) demonstrate the effectiveness of *SS-shrink*. For example, three intestinal obstruction associated GWAS loci previously colocalizing with eQTLs in the pancreas using the SS (Gong et al. 2019) continue to do so with *SS-shrink* (p=3.10E-8, p=9.89E-10 and p=6.33E-5 at *ATP12A*, *SLC6A14* and *SLC26A9* respectively). While integration of lung function GWAS from the ICFGMC at the *SLC9A3* modifier locus (min. GWAS p=4.62E-12; set-based SS p=7.12E-8) with eQTLs from nasal epithelia of 81 ICFGMC Canadian participants (min. eQTL p=9.6E-3; set-based SS p=0.47) shows no colocalization evidence (*SS-shrink* p=1.00).

*SS-shrink* is a flexible, unifying framework for data integration and improves over traditional methods in the presence of weak signals, LD or allelic heterogeneity.

## PgmNr 2979: Identifying COPD subtypes using multi-trait genetics.

### Authors:

A. Ziyatdinov<sup>1,2</sup>; B.D. Hobbs<sup>2,3</sup>; S. Kanaan-Izquierdo<sup>5,6,7</sup>; P. Kraft<sup>1</sup>; E.K. Silverman<sup>2,3</sup>; H. Aschard<sup>1,4</sup>; M.H. Cho<sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA, USA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA, USA; 4) Department of Computational Biology - USR 3756 CNRS, Institut Pasteur, Paris, France; 5) Centre de Recerca en Enginyeria Biomèdica, Universitat Politècnica de Catalunya, Barcelona 08028, Spain; 6) CIBER of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Barcelona, Catalonia, Spain; 7) Institut de Recerca Sant Joan de Deu, Esplugues de Llobregat, Spain.

---

Chronic Obstructive Pulmonary Disease (COPD) is a phenotypically heterogeneous disease. Identifying subtypes would be of high value from a clinical and public health perspective. Extensive prior work has attempted to identify subtypes by phenotype clustering, but these methods often do not account for differences in severity, disease course, or underlying biologic mechanisms. Here, we used genetic association results to infer potential COPD subtypes and assessed the biological relevance of these subtypes using individual level data from the COPDGene study. We used a set of COPD and related phenotype-associated SNPs derived from recent GWAS, and decomposed the Z-score associations at those SNPs across hundreds of traits from the UK Biobank. Decomposition was performed using a recently developed Bayesian approach, and results were used to build SNP clusters with similar cross-trait characteristics. Finally, we derived cluster-specific genetic risk scores (GRS) and assessed their association with clinical features available within COPDGene.

We selected variants from three GWAS studies of obstructive lung disease: 164 from COPD, 279 from lung function, and 45 from asthma. Our clustering revealed three clusters consistent across different settings; the top trait (signed) weights were (1) +eosinophils, +wheeze, +neutrophils; (2) +obesity, -height; (3) +height, -neutrophils. Only 25% SNPs contributed with normalized weights >0.5, highlighting a sparse decomposition. By building cluster-specific GRSs in the COPDGene study and associating them to clinical measures relevant to COPD, we found cluster-specific differences in BMI, lymphocyte counts, and chronic bronchitis. Further, we found no significant overlap between our genetic and phenotypic clusters in COPDGene, potentially suggesting a complementary role of genetic stratification. We also conducted a consensus version of clustering for SNP sets from three GWAS studies: trait weights are shared, but SNP weights are set-specific. Contrasting SNP sets by weights confirm that COPD and lung function SNPs mostly overlap, while asthma SNPs are separated (SVM classifier: 82% precision, 60% recall).

Phenotypic clustering of obstructive lung disease related risk SNPs can identify genetically driven phenotypic patterns in COPD. Selection of SNPs and traits for the clustering analysis as well as further exploration of the characteristics of subjects with differential on these GRSs are open questions for future research.

## PgmNr 2980: Association study of genetic variants in autophagy pathway genes and NSCL/P risk.

### Authors:

S. Lou<sup>1,2</sup>; L. Ma<sup>1</sup>; X. Yu<sup>1,2</sup>; D. Li<sup>1,2</sup>; Y. Pan<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Nanjing Medical University, Nanjing, China; 2) Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, China

---

### Objective:

Although genetic variants in autophagy genes were associated with the risk of oral cancers and early development, their associations with nonsyndromic cleft lip with or without cleft palate (NSCL/P) risk are not understood.

### Methods:

We systematically selected single-nucleotide polymorphisms (SNPs) in 23 autophagy genes and investigate this issue in a case-control study. The logistic regression model was used to calculate the effects of SNPs on NSCL/P susceptibility. In addition, we performed a gene-based analysis via sequence kernel association test (SKAT). Expression quantitative trait loci (eQTL) analysis was conducted using the Genotype-Tissue Expression (GTEx) dataset.

### Results:

Two SNPs in *KRAS* were significantly associated with NSCL/P susceptibility in the case-control study (rs12813551: OR=0.75, 95% CI=0.61-0.93,  $P=7.95E-03$ ; rs147020828: OR=0.67, 95% CI=0.50-0.90,  $P=7.67E-03$ ). Rs2301104 in *HIF1A* was also associated with NSCL/P susceptibility (OR=1.39, 95% CI=1.00-1.92,  $P=4.55E-02$ ). Furthermore, rs12813551 and rs147920828 exhibited eQTL hits, which showed a significant association with expressions of *KRAS* ( $P=4.49E-04$  in fibroblast cells, and  $P=2.4E-02$  in lymphocyte cells, respectively). Moreover, the expression level of *KRAS* is higher in dental pulp stem cells (DPSC) of NSCL/P cases than those in controls, while *HIF1A* exhibits the opposite.

### Conclusion:

Our results indicated that SNPs in autophagy pathway genes *KRAS* and *HIF1A* played important roles in NSCL/P susceptibility.

## PgmNr 2981: Identification of new susceptibility loci for nonsyndromic cleft lip with or without cleft palate.

### Authors:

L. Ma<sup>1</sup>; S. Lou<sup>1,2</sup>; X. Yu<sup>1,2</sup>; D. Li<sup>1,2</sup>; Y. Pan<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Nanjing Medical University, Nanjing, China; 2) Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, China

### Objective:

Although several genome-wide association studies (GWAS) of non-syndromic cleft lip with or without cleft palate (NSCL/P) have been reported, more novel association signals could be exploited by combining individual studies together, which will further elucidate the genetic susceptibility of NSCL/P.

### Methods:

We performed a meta-analysis of two published Chinese GWAS cohort studies (858 NSCL/P cases and 1,248 controls) and followed by replication in two stages. In addition, we calculated the weighted genetic risk score (wGRS) of the susceptibility loci based on odds ratio of each variant from the replication cohort.

### Results:

In the combined analysis of GWAS meta-analysis and replication stages, we identified five novel association signals: rs11119445 (OR=0.73, 95% CI: 0.67 to 0.79;  $P=6.44E-14$ ), rs227227 (OR=1.34, 95% CI: 1.24 to 1.45;  $P=5.02E-13$ ), rs12561877 (OR=0.72, 95% CI: 0.65 to 0.79;  $P=2.80E-11$ ), rs643118 (OR=1.25, 95% CI: 1.13 to 1.37;  $P=4.45E-06$ ), and rs2095293 (OR=1.22, 95% CI: 1.11 to 1.34;  $P=2.98E-05$ ), which were independent of the previously reported SNP at the same locus. The nucleotide polymorphism (SNP) marker rs11119445 maps 7.4kb 3' of *SERTAD4-AS1*, a non-coding RNA on 1p32.2. The SNP markers rs12561877 and rs227227 locate in the intron of *SYT14*. The SNP marker rs643118 lies in the intron of *TRAF3IP3*. The SNP marker rs2095293 resides in an intron of *NR6A1*. The mean (standard deviation) of the wGRS was 1.83 (0.64) for NSCL/P cases and 1.59 (0.68) for controls respectively, which showed a clear separation between them ( $P=2.67E-16$ ). Based on the ToppFun online tool, the most enriched pathway is *POU5F1 (OCT4)*, *SOX2*, *NANOG* activate genes related to proliferation ( $P=1.04E-03$ ).

### Conclusion:

Our studies have improved our understanding of the genetic susceptibility to NSCL/P and provided further clues to its etiology in the Chinese population.

## PgmNr 2982: Transcriptome-wide association studies (TWAS) for regional brain volumes and brain white matter microstructure in >20,000 individuals.

### Authors:

Y. Shan <sup>1</sup>; B. Zhao <sup>1</sup>; Y. Yang <sup>1</sup>; Y. Li <sup>1,2,3</sup>; H. Zhu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of North Carolina at Chapel Hill; 2) Department of Genetics, University of North Carolina at Chapel Hill; 3) Department of Computer Science, University of North Carolina at Chapel Hill

---

Brain imaging features are promising endophenotypes for psychiatric disorders, cognitive function, and psychosocial factors. Understanding how genetic variants impact variations in brain structure and function help us understand how the brain works. Recent studies have investigated the heritability of brain imaging features and performed corresponding genome-wide association studies (GWAS), in particular for regional brain volumes (quantified by regions of interest [ROIs]) and for brain white matter microstructure (quantified by diffusion tensor imaging [DTI] features). To further reveal genetics underlying brain imaging phenotypes, we performed transcriptome-wide analysis (TWAS) on 101 ROIs (sample size  $n=22,748$ ) and 110 DTIs ( $n=20,115$ ) features derived from brain MRI from UK Biobank (UKB), Alzheimer's Disease Neuroimaging Initiative (ADNI), Pediatric Imaging, Neurocognition, and Genetics (PING) Study, the Philadelphia Neurodevelopmental Cohort (PNC), and the Human Connectome Project (HCP). Specifically, we first trained gene expression prediction models from the CommonMind eQTL data containing 467 samples with both genotype and dorsolateral prefrontal cortex gene expression data, then we performed transcriptome-wide ( $\sim 12,000$  genes) and phenome-wide (101 ROI and 110 DTI features) association analysis adjusting for kinship coefficients in a linear mixed model framework. Our TWAS identified 83 signals (49 for ROI and 34 for DTI; adjusting for transcriptome-wide and phenome-wide multiple testing), among which 45 (32 for ROI and 13 for DTI) are novel when compared to the corresponding GWAS results. We additionally applied Unified Test for Molecular Signatures (UTMOST) to perform TWAS across multiple tissues using UKB GWAS summary statistics, which confirmed 9 ROI and 13 DTI signals out of the 83 individual-level discoveries. Our individual-level TWAS signals include several genes with supporting evidences from the literature. For example, *EEF1AKMT2* and *LEMD3* (also discovered by our GWAS) were reported to be associated with hippocampal volume; *KTN1-AS1* (not by our GWAS) was reported to be associated with putamen volume. Among the identified genes, many are annotated with cognitive functions or psychiatric disorders. Our analyses demonstrate the value of TWAS for uncovering genetic mechanisms underlying regional brain volumes and brain white matter microstructure, which will shed light on genetics behind many brain related traits and disorders.

## PgmNr 2983: Analysis of whole exome sequencing of the *ClinSeq*<sup>®</sup> data with four regression-based tests of association.

### Authors:

H. Sung <sup>1</sup>; J.A. Sabourin <sup>1</sup>; K. Lewis <sup>2</sup>; J.C. Mullikin <sup>3,4</sup>; A.T. Remaley <sup>5</sup>; L.G. Biesecker <sup>2</sup>; A.F. Wilson <sup>1</sup>; NISC Comparative Sequencing Program

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genometrics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, NIH, Baltimore, MD; 2) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 3) Comparative Genomics Analysis Unit, Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 4) National Institutes of Health Intramural Sequencing Center (NISC), National Human Genome Research Branch, NIH, Bethesda, MD; 5) Lipoprotein Metabolism Laboratory, Cardiovascular and Pulmonary Branch, National Heart Lung and Blood Institute, NIH, Bethesda, MD

---

*ClinSeq*<sup>®</sup> is a medical sequencing study designed to investigate associations of sequence variants with traits related to Coronary Artery Disease (CAD). The study currently includes more than 1,000 participants between the ages of 45 and 65 with normal to severe coronary artery calcification scores. About 200 CAD-related traits were measured at the NIH Clinical Research Center in Bethesda, Maryland. Exome sequencing was performed with four distinct capture kits at the NIH Intramural Sequencing Center (NISC): Agilent SureSelect with 38 Mb and 50 Mb (Index and ICGC) capture kits for 396 and 320 individuals, respectively, TruSeq (V1 and V2 combined) for 285 individuals, and SeqCap EX Exom+UTR for 18 individuals respectively. Data cleaning was performed within data from each capture kit separately and then in the combined data after merging. Single nucleotide variants (SNVs) common to all capture kits were used yielding 589,727 SNVs in 886 unrelated European Americans for analysis. Tests of association were performed for each SNV and each of the 130 CAD-related untransformed and rank-inverse transformed traits pre-adjusted for age, sex, BMI, the use of cholesterol lowering medication and capture kit. Traits were selected if any SNV was significant at the indicated level for both the untransformed and rank-based inverse normal transformed trait. Tests included simple linear regression (SLR), ComPaSS-GWAS, ComPaSS-rare and tiled regression (TR). Fifteen traits were significant at a critical value of  $10^{-7}$  for SLR, 5 and 9 traits for a corroboration score of 0.6 for ComPaSS-GWAS and ComPaSS-rare, respectively, and 15 traits for TR with selection criteria of  $10^{-4}/10^{-5}/10^{-6}$  for tile, chromosome and genome levels. There were 5 traits with the same significant SNV found across all 4 tests of association. Compared to traditional SLR, ComPaSS-GWAS found fewer significant SNVs because the type I error rate was reduced through the use of repeated random splits. ComPaSS-rare using complementary random weights on each sample instead of binomial weights in splitting in half found more significant rare SNVs than ComPaSS-GWAS due to its increasing power to detect rare SNVs. TR found fewer significant SNVs than SLR by identifying independently and jointly significant SNVs across the genome. Tests of association performing ComPaSS-GWAS/ComPaSS-rare and TR in exome-sequencing data may help to reduce type I errors rates and prioritize SNVs for additional follow-up.



## PgmNr 2984: Design and prototype of the web implementation for the KING software package for relationship inference in genetic studies.

### Authors:

Chen; Z.N. Zhu; A.W. Manichaikul; W.M. Chen

[View Session](#) [Add to Schedule](#)

**Affiliation:** Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA

---

**Introduction:** We have developed an integrated software toolset (KING, <http://people.virginia.edu/~wc9c/KING/>) for relatedness inference in large-scale genetic studies. KING has been widely used for quality control (QC) and genetic analysis in large-scale studies including all Biobank data, and many genomics and transcriptomics sequencing projects. While KING as a command line tool is sufficient for many genetic studies, some research team may not have the computational resources to utilize the full functionality offered by KING, for example, some visualization of relationship inference may require proper installation of several R packages.

**Objective:** To increase KING's usability and its feasibility of community free distribution to avoid downloading of the binary C++ KING software and its supportive compiling libraries for various computer platforms, we have developed a web-based KING system with graphical user interface to allow community researchers to apply our latest developed algorithms directly for analysis of their sequencing and other high-throughput array data.

**Methods:** The web frontend of our KING web system is implemented using HTML and JavaScript programming languages. The Common Gateway Interface (CGI) protocol is used to get frontend data and web user inputs for backend server processing via Perl CGI and Java Servlet technologies. Apache web server is utilized for web hosting, and the system is implemented in Linux operating system using a powerful backend server with 96 cores for parallel computing.

**Progress and Further Work:** We have prototyped the web interface of KING software and have implemented a CGI program to get the frontend user inputs for backend data processing. Users' data are encrypted and secured. The CGI program returns the results on the web or via email with downloading URLs if the data/results are too big for web quick display and response. We have established a development instance and also a test instance for project development, and are in the process of setting up a mature production instance for final public release. The web-based KING will implement the full functionality of our published and public downloadable KING software, including data visualization and IBD segment analysis, and among others. We will follow the standard software engineering practice to go through: design and prototype, development and test, and production phases to put out a mature KING software web system for the community to use.



## PgmNr 2985: HPrep: A model-based method to quantify reproducibility of HiChIP and PLAC-seq data.

### Authors:

J.D. Rosen <sup>1</sup>; A. Abnoui <sup>4</sup>; I. Juric <sup>4</sup>; Y. Yang <sup>2</sup>; Y. Shen <sup>5</sup>; M. Hu <sup>4</sup>; Y. Li <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina, Chapel Hill, NC; 3) Department of Computer Science, University of North Carolina, Chapel Hill, NC; 4) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH; 5) Department of Neurology, Institute for Human Genetics, University of California, San Francisco, San Francisco, CA

---

High-throughput chromatin conformation capture (Hi-C) technology has been widely used to measure genome-wide chromatin spatial organization. Various statistical methods and computational tools have been developed for measuring the reproducibility of Hi-C data, a necessity for evaluating the validity of conclusions drawn from Hi-C data. Recently, HiChIP and PLAC-seq (HP) technologies have been developed for studying long-range chromatin interactions mediated by protein of interest, facilitating high resolution mapping of chromatin interactions at a much reduced cost. However, computational methods tailored for HP data are still lacking. In particular, existing methods for Hi-C reproducibility are not optimal for HP data, since none of them takes into consideration the ChIP enrichment level, a systematic biased unique to HP data.

To fill in this gap, we propose a novel model-based method, HPrep, to measure reproducibility in HP data. HPrep consists of four major steps: (1) preprocessing of the raw contact frequency matrix, (2) fitting positive Poisson regression model to normalize ChIP enrichment level and other systematic biases in HP data, (3) performing 1-D local smoothing on the normalized contact frequency, and (4) applying a stratified correlation algorithm to the smoothed contact frequency to quantify the reproducibility between two HP datasets.

We applied HPrep to H3K4me3 PLAC-seq data on both mouse embryonic stem cells (mESCs) and mouse anterior dorsal cerebellar vermis (ADCV), each consisting of two biological replicates. In addition, we created a set of pseudo-replicates by merging biological replicates followed by random sampling. We obtained measures of 0.94, 0.91, 0.84, and 0.52 for pseudo-replicates, mESC biological replicates, mouse ADCV replicates, and non-replicates, respectively. In addition, we applied HPrep to H3K4me3 PLAC-seq data on four human fetal brain cell types (radial glial cells, intermediate progenitor cells, neurons, and interneurons), each consists of two or three biological replicates, and found that biological replicates from the same cell type show higher reproducibility than biological replicates from different cell types. In comparison, results obtained by naively applying a standard Hi-C reproducibility methodology failed to show differentiation (for example, 0.09 for biological replicates and 0.05 for non-replicates). We believe that HPrep fills in an important gap to quantify reproducibility in HP data.

## PgmNr 2986: Regulatory-coding variant interactions shape genome structure and can be leveraged for gene mapping GWAS associations.

### Authors:

R. Brown <sup>1</sup>; A. Durvasula <sup>2</sup>; S. Sankararaman <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computer Science, UCLA, Los Angeles, California.; 2) Human Genetics, UCLA, Los Angeles, California.; 3) Computational Medicine, UCLA, Los Angeles, California.

---

While GWAS have discovered variants associated with numerous traits, the molecular mechanisms through which these variants impact phenotypes remain poorly understood. Recent studies have shown that cis-regulatory variants modulate the effects of coding variants in their target genes such that pathogenic coding alleles on under-expressed haplotypes have reduced penetrance. Here we show how the long-term effect of these interactions shapes the patterns of variation across the genome and how ascertainment bias distorts these patterns in a way that can be used for gene mapping GWAS hits. We analyze eQTLs from the Genotype-Tissue Expression Project (GTEx) with genetic variation from 1000 Genomes and find that haplotypes with down-regulating eQTL alleles contain 14% (p-value  $6.7 \times 10^{-6}$ , paired t-test) more derived coding alleles on average than haplotypes with up-regulating alleles. This indicates that haplotypes with down-regulating eQTL alleles appear to be more tolerant of derived coding alleles over time and compliments previous work that shows missense variants that are predicted to be strongly deleterious tend to be under-expressed when allele-specific expression data is available. We hypothesize that the distribution in a gene of the number of derived alleles on up- and down-regulated haplotypes will deviate from what is observed across a population sample if haplotypes are ascertained based on the the presence or absence of a GWAS top hit risk allele that is impacting the gene in question. We propose a test that maps top GWAS risk loci to pertinent genes without needing in-sample data or summary statistics. Across six traits with a large number of intergenic GWAS hits, we find genes prioritized by our method are 7.4 times (p-value  $2.5 \times 10^{-7}$ , Fisher's exact) more likely to be reported in the original GWAS as pertinent for the traits than the genes not prioritized. This demonstrates that our test provides a novel source of information for gene prioritization.

## PgmNr 2987: Estimating correlations between genetic effects at interacting regions.

### Authors:

Y. Ding<sup>1</sup>; K. Burch<sup>1</sup>; B. Pasaniuc<sup>2,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics Interdepartmental Program, University of California Los Angeles, Los Angeles, CA, USA; 2) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 3) Department of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 4) Department of Computational Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA

---

In the past few years, technological advances in chromosome conformation capture methods have identified numerous regions across the genome that physically interact. Interactions between regions such as enhancers and promoters are crucial to regulation of gene expression. For example, if a specific enhancer-promoter interaction regulates expression of a gene that is involved in some trait-influencing pathway, we would expect the effect sizes of variants in those regions to be correlated. However, the vast majority of methods for post-GWAS analysis assume that there is zero covariance between causal effect sizes. Thus, existing methods likely cannot reveal insights into how interacting regions influence complex traits.

Here, we propose an approach to model and estimate the correlation between effect sizes at two non-overlapping regions while explicitly modeling linkage disequilibrium (LD). We model the dependence between genetic effects in pairs of regions with a constant correlation coefficient between all contact pairs, and we estimate this correlation coefficient from GWAS data with a method-of-moments procedure. Through extensive simulations starting from real genotypes of the UK Biobank and 1000 Genomes Phase 3 reference panel, we show that given a known set of interactions and per-SNP variances, the correlation coefficient can be consistently estimated and the standard error decreases as the LD block size decreases.

## PgmNr 2988: Using information across tissues and genes to predict gene expression in transcriptome-wide association studies.

### Authors:

F. Morgante <sup>1</sup>; G. Wang <sup>2</sup>; M. Stephens <sup>2,3</sup>; Y.I. Li <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Section of Genetic Medicine, Department of Medicine, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Department of Statistics, University of Chicago, Chicago, IL

---

Recent years have seen the development of several methods that combine GWAS data with expression reference panels (e.g., GTEx) to discover genes associated with complex traits. One class of such methods is Transcriptome-wide Association Study (TWAS). These methods involve three steps. First, predictors of gene expression for each gene based on genotypes are constructed using reference panels such as GTEx. Second, the genetically regulated portion of gene expression is imputed in GWAS samples based on their genotypes and the predictors previously built. Finally, the imputed expression is correlated with the phenotype of interest.

While the TWAS methods initially developed build gene expression predictors for one tissue at a time, the increasing availability of measurements from multiple tissues opens to the possibility of exploiting patterns of expression Quantitative Trait Locus (eQTL) sharing across tissues to improve imputation accuracy. Recently, a couple of multivariate regression methods have been developed that use information from multiple tissues concurrently. Extensive analysis of both simulated and real data showed a greater prediction accuracy of multivariate methods compared to univariate methods. However, currently available multivariate prediction methods analyze only one gene at a time, therefore not exploiting the increased amount of information about eQTL sharing that multiple genes can provide. Motivated by this, we sought to evaluate the accuracy of gene expression prediction of a Bayesian multivariate sparse regression method – initially designed for fine-mapping – with a prior that captures patterns of eQTL sharing across tissues and borrows strength across genes. Preliminary results from simulations show that our method can predict as well as currently available multi-tissue methods, while enhancing interpretability due to concurrent fine-mapping of causal variants.

## PgmNr 2989: Linkage disequilibrium pruning improves gene expression prediction across diverse populations.

### Authors:

R. Schubert<sup>1,2,3</sup>; P. Okoro<sup>2</sup>; A. Luke<sup>4</sup>; L. Dugas<sup>4</sup>; H. Wheeler<sup>1,2,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Biology, Loyola University Chicago, Chicago, IL.; 2) Dept. of Mathematics and Statistics, Loyola University Chicago, Chicago, IL; 3) Program in Bioinformatics Loyola University Chicago, Chicago, IL; 4) Dept. of Public Health Sciences, Loyola University Chicago, Stritch School of Medicine, Maywood, IL; 5) Dept. of Computer Science, Loyola University Chicago, Chicago, IL

Transcriptome-wide association studies are widely used for identifying potential regulatory mechanisms underlying complex traits. However, the portability of gene expression prediction models varies between populations likely due to different linkage disequilibrium (LD) patterns and allele frequencies. We sought to determine whether LD pruning prior to elastic net model building improves or maintains cross-population predictive performance while lowering runtimes.

We used transcriptome data from two cohorts, MESA (Multiethnic Study of Atherosclerosis, monocytes) and METS (Modeling the Epidemiological Transition Study, whole blood), as our primary training and test datasets. We pruned linked SNPs with a window size of 50kb, variant SNP shift size of 5, and  $r^2$  threshold of 0.8 within each of the 3 self-identified populations that comprise MESA: African American (AFA=233), Hispanic (HIS=352), and European (CAU=578). Then, in each population, we built pruned and unpruned models using elastic net with nested cross-validation to assess the predictive performance ( $\rho$ ) of local SNPs (within 1Mb) surrounding each gene. Genes with average predicted vs observed  $\rho > 0.1$  were considered significant. More genes were significant in pruned (AFA=3449, HIS=4048, CAU=4534) than unpruned models (AFA=3096, HIS=3856, CAU=4419). We predicted gene expression in METS (n=77 Ghanaians and African Americans) using the pruned and unpruned models built in MESA and compared predicted to observed METS expression. Spearman correlations were not significantly different when comparing pruned (median=0.089) and unpruned AFA (median=0.093) in METS ( $p=0.44$ ). However, Spearman correlations were higher in the pruned models for both HIS (pruned median=0.070, unpruned median=0.058,  $p=3.6e-4$ ) and CAU (pruned median=0.048, unpruned median=0.042,  $p=0.016$ ). The average runtime for building pruned models was 55 s per gene (95% CI [45.8, 64.2]), compared to 205 s per gene for unpruned models (95% CI [192.8, 218.1]).

Our results indicate that by LD pruning we decrease runtime and predict more genes both within and between populations without sacrificing accuracy. While prediction is best between similar ancestry populations, LD pruning prior to elastic net modeling prefilters out highly correlated SNPs within a population and thus may reduce potential sources of noise in the elastic net selection, which leads to improved predictive performance across populations compared to unpruned models.

## PgmNr 2990: Accounting for covariates in tiled regression analysis of complex traits.

**Authors:**

A.J.M. Sorant; J.A. Sabourin; A.F. Wilson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Computational and Statistical Genomics Branch, NIH/NHGRI, Baltimore, Maryland.

---

Complex traits typically have both genetic and non-genetic components which should be considered in forming a model explaining trait variation. Tiled regression, which determines an additive model with stepwise regression applied in stages, can incorporate covariates with several methods. The traditional method is to adjust a trait for covariates of presumed importance by regressing the trait on the covariates and then to use the residuals from the fitted model as a new trait for the usual tiled regression considering genetic predictors. However, it is also possible to incorporate covariates directly into tiled regression either by forcing them into every considered model or by testing them along with genotypes as potential predictors. Forcing inclusion of all covariates produces a predictive model similar to that determined by pre-adjustment, except that the covariate coefficients are not fixed, but determined in combination with genetic components under consideration at each stage. When covariates are instead tested as potential predictors for inclusion in the model, they may or may not be retained. For this study, four ways of handling covariates were considered: (1) ignoring them, (2) pre-adjusting the trait for them, (3) forcing them to be included in every model considered, (4) testing them for inclusion in the final model. Tiled regression was performed with each method, using software TRQUANT v 1.0, for several previously studied metabolic phenotypes from the Trinity Student Study (TSS) with 670,081 imputed genotypes on 2,232 individuals and also for several simulated traits using the same genotypes. Analysis of TSS phenotypes methylmalonic acid, serum vitamin B12 and holo-transcobalamin concentrations (log-transformed), considering covariates age, sex, body mass index and vitamin B6 supplementary intake, produced similar predictive models with all analysis methods. With simulated phenotypes based on five independent SNPs and two covariates, preliminary results (with 200 replicates) show that all analysis methods have similar ability to detect causal SNPs, with pre-adjustment or keeping all considered covariates usually having slightly higher power and type I error rate.

## PgmNr 2991: Computational gene prioritization from GWAS using local and polygenic signal.

### Authors:

E. Weeks<sup>1</sup>; J. Ulirsch<sup>1,2</sup>; R. Fine<sup>1,2,3,4</sup>; K. Tashman<sup>1,2,3,4</sup>; C. Fulco<sup>1,5,6</sup>; J. Nasser<sup>1,7</sup>; T. Li<sup>1,6,7,8</sup>; C. Smillie<sup>9</sup>; M. Biton<sup>9,10</sup>; J. Ordoñas-Montañes<sup>7,9,11,12,13,14</sup>; A. Ananthakrishnan<sup>15</sup>; A. Shalek<sup>7,9,11,12,13</sup>; R. Xavier<sup>1,10,15,16,17,18</sup>; A. Regev<sup>9,19</sup>; R. Gupta<sup>1,20,21,22</sup>; K. Lage<sup>1,6,16</sup>; J. Engreitz<sup>1,23</sup>; E. Lander<sup>1,5,24</sup>; H. Finucane<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Institute of MIT and Harvard, Cambridge, MA; 2) Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA; 4) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA; 5) Department of Systems Biology, Harvard Medical School, Boston, MA; 6) Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA; 7) Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA; 8) Johns Hopkins School of Medicine, Baltimore, MD; 9) Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA; 10) Department of Molecular Biology, Massachusetts General Hospital, Boston, MA; 11) Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA; 12) Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA; 13) Ragon Institute of MGH, MIT and Harvard, Cambridge, MA; 14) Division of Infectious Diseases and Division of Gastroenterology, Boston Children's Hospital, Boston, MA; 15) Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease, Massachusetts General Hospital, Boston, MA; 16) Harvard Medical School, Boston MA; 17) Center for Microbiome Informatics and Therapeutics, Massachusetts Institute of Technology, Cambridge, MA; 18) Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA; 19) Howard Hughes Medical Institute and Koch Institute for Integrative Cancer Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA; 20) Division of Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 21) Center for Genomic Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA; 22) Division of Genetics, Brigham and Women's Hospital, Boston MA; 23) Harvard Society of Fellows, Harvard University, Cambridge, MA; 24) Department of Biology, Massachusetts Institute of Technology, Cambridge, MA

---

For the vast majority of trait-associated loci from genome-wide association studies (GWAS), the identity of the causal gene(s) underlying the association is unknown. Although previous work has shown that genome-wide patterns of enrichment can be informative about disease biology, no method we know of takes advantage of the full set of genome-wide signals to prioritize disease genes. Moreover, we are not aware of any statistical framework for combining results from independent methods to identify high-confidence genes supported by multiple lines of evidence.

Here, we first propose a new method that leverages genome-wide enrichments of gene features to compute a polygenic priority score (PoPS) and a p-value for each gene based on its features. We applied our method to summary statistics for 40 complex traits and diseases using 18,920 gene features derived from 12 gene expression datasets, 5 pathway databases, and a protein-protein interaction network. We validated our method on 14 sets of gold standard genes, where we predicted significantly higher scores for all 14 sets of gold standard genes tested, compared to all protein-coding genes. To further validate, we used Benchmarker (Fine et al. 2019 *AJHG*) to show that our

prioritized genes are significantly enriched for heritability for 38 of 40 traits, and we found that they are significantly enriched for their overlap with GWAS loci for 19 of 40 traits. We also used Benchmarker to show that we outperform DEPICT (Pers et al. 2015 *Nat Comm*) across 100% (13 of 13) traits tested.

Second, we propose a new framework for combining information from multiple independent gene prioritization methods. Based on the estimate that 30-70% of causal genes in GWAS loci are the closest gene to the index SNP, we estimated the probability that our prioritized genes are causal. By combining the results of our method with results of independent methods, we nominated 461 genes across 29 traits with probability of being causal greater than 0.8. These genes included well-established disease genes such as ACE for cardiovascular disease and MYOCD for atrial fibrillation. They also included genes recently nominated in other work as causal, such as IRF1 for IBD, and novel disease genes, such as FCER1G for Asthma. Overall, we present a framework to harness local and polygenic signal to prioritize causal genes from GWAS data and provide a promising list of candidate genes for targeted and cost-efficient follow up experiments in ~30 traits.



## PgmNr 2992: Incorporating single-cell RNA-seq data to infer allele-specific expression.

### Authors:

J. Fan; R. Xiao; M. Li

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA

---

Allele-specific expression (ASE) can be quantified by the relative expression of two alleles in a diploid individual, and such expression imbalance may explain phenotypic variation and disease pathophysiology. Traditional methods detect ASE using easily obtainable bulk RNA-seq data, a data type that averages out the possible heterogeneity in a mixture of different cell types. Since patterns of ASE may vary across cell types, it is desirable to infer ASE in a cell-type-specific manner. This can be achieved by analyzing single-cell RNA-seq (scRNA-seq) data across many individuals. However, such analysis requires full-length transcript sequencing in single cells, which are costly to generate. To circumvent this limitation, we propose to incorporate cell-type composition information inferred from a small set of scRNA-seq data, possibly obtained from a different set of individuals, to characterize cell-type-specific ASE in a larger set of bulk RNA-seq data. To do so, we first employ MuSiC, a deconvolution algorithm, to infer cell-type compositions in bulk RNA-seq data using an external scRNA-seq dataset as the reference. To infer cell-type-specific ASE, we regress the bulk level allele-specific read counts over the estimated cell-type proportions through a linear mixed-effects model, and test for the presence of cell-type-specific ASE effect. Extensive simulations shown that our method performed consistently well under a wide range of scenarios. The type I error rate was under control and it was powerful in detecting cell-type-specific ASE effect even for rare cell types. We further evaluated our method on a benchmark dataset generated from human pancreatic islets where the true cell-type-specific ASE effect is known. Among 11,290 genes analyzed, the traditional bulk ASE analysis method, only detected three genes showing ASE. However, our method not only detected these three genes, but also detected an additional 2,400 genes showing cell-type-specific ASE. We further applied our method to another pancreatic islet bulk RNA-seq dataset for 89 individuals. Controlling for cell-type composition differences across individuals, our preliminary analysis recovered genes showing cell-type-specific ASE that are related to T2D. As bulk tissue data are more easily accessible than scRNA-seq, our method allows the utilization of the vast amounts of disease relevant bulk tissue RNA-seq data for elucidating cell type contributions in disease.

## PgmNr 2993: Characterizing the block-jackknife based significance testing used in stratified LD-score regression.

### Authors:

K. Tashman <sup>1,2</sup>; B. Neale <sup>1,2</sup>; H. Finucane <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanley Center, Broad Institute, 75 Ames Street, Cambridge, MA 02142; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Richard B. Simches Research Center, 185 Cambridge Street, CPZN-6818, Boston, MA 02114

Stratified LD-Score Regression (S-LDSC; Finucane et al. 2015 Nat Genet) is a widely used method to detect whether trait heritability is enriched in a functional annotation. S-LDSC has previously been used to report the heritability enrichment of an annotation, the regression coefficient corresponding to the annotation (i.e., the contribution of the annotation to per-SNP heritability in a joint-fit model), and their corresponding p-values for a wide range of functional annotations. However, there are limitations of the block jackknife-based significance testing used in S-LDSC to control type 1 error for small annotations.

Here, we investigated type 1 error control in S-LDSC for small annotations using simulations. We implemented a simulation framework using 50,000 white-British UK Biobank samples and 9.3 million imputed variants to produce three sets of 1,250 phenotypes with 200, 1,000, and 10,000 causal SNPs and no functional enrichments. We tested 94 random annotations that varied in size, the number of jackknife blocks they overlapped, and average segment length of the annotation, 100 gene sets sampled from MSigDB, as well as the baseline\_v1.1 SNP annotations by running S-LDSC with one annotation at a time plus baseline\_v1.1 for each of the 3,750 phenotypes to estimate the per-SNP heritability and enrichment, and perform significance testing.

We observed inflation for both the one-sided and two-sided tests of the per-SNP heritability at low numbers of independent SNPs and overlapping jackknife blocks. Null calibration could be recovered by restricting to annotations overlapping more than 100 jackknife blocks and with more than 7,500 independent SNPs for the one-sided test of the per-SNP heritability. The two-sided test required restricting to 12,000 and 15,000 independent SNPs to recover null calibration for the most and least polygenic phenotypes, respectively. The enrichment estimates required a weaker threshold of 4,500 independent SNPs. The baseline\_v1.1 annotations had enrichment p-values that were well calibrated across all levels of polygenicity except for conserved regions, which had very mild inflation for the most polygenic set of phenotypes. We have implemented a test in the S-LDSC software to warn the user when their annotation does not meet the criteria required to produce statistically valid p-values and plan to continue to investigate analytical solutions to estimate standard errors in the S-LDSC framework.

## PgmNr 2994: Imputation of variable number tandem repeat variants reveals VNTRs driving GWAS loci.

### Authors:

R.E. Mukamel<sup>1,2</sup>; R.E. Handsaker<sup>2,3,4</sup>; S.A. McCarroll<sup>2,3,4</sup>; P. Loh<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 4) Department of Genetics, Harvard Medical School, Boston, MA

The role that variable number tandem repeat (VNTR) loci play in heritable traits remains largely unknown, primarily due to difficulties genotyping these genomic structural variants at population-scale. To investigate the phenotypic impact of VNTRs, we developed statistical methods to phase and impute VNTR lengths. We identified 257 genic VNTRs—ranging in size from 18-bp repeats within exons to megabase-scale gene family polymorphisms—by scanning the human reference genome for repetitive sequence. We estimated the repeat lengths of these VNTRs from whole-genome sequencing read depth in 519 quartet families in the Simons Simplex Collection. We inferred allelic contributions to these measurements by drawing upon surrounding SNPs to model haplotype sharing.

Using these haplotypes as a reference panel, we imputed VNTR lengths into a cohort of 337,488 unrelated British UK Biobank participants and tested for associations with 124 phenotypes including neurodevelopmental, cardiovascular, and metabolic disorders and other health-related traits. We found 22 VNTR-phenotype associations involving 11 distinct VNTRs that reached Bonferroni significance ( $p < 1.6 \times 10^{-6}$ ) and for which conditional association analyses showed that the VNTR, rather than nearby SNPs, was the likely causal variant.

The strongest association ( $p < 1 \times 10^{-13,800}$ ) in our data replicated a known association between Lipoprotein(a) concentration (Lp(a)) and repeat length of a VNTR that encodes *LPA* KIV2 domains, a VNTR implicated in cardiovascular diseases. Intriguingly, Mendelian randomization analyses of KIV2 VNTR length together with nearby variants indicated that the contribution of *LPA* variation to other molecular traits, including triglycerides, was not mediated solely by Lp(a).

Other strong associations ( $p < 2 \times 10^{-96}$ ) identified VNTRs as likely causal variants in previously reported GWAS loci. SNPs near *ACAN*, which encodes a component of the cartilage extracellular matrix, have previously been associated with height; SNPs near *MUC1*, which encodes a transmembrane protein expressed in secretory epithelial cells, have previously been associated with urea levels. In our analysis, imputed repeat lengths for VNTRs in exons of *ACAN* and *MUC1* explained these associations. In addition, multi-allelic copy number variation modifying dosage of genes in the *PGA* and *TP53TG3* gene families drove novel associations ( $p < 3 \times 10^{-10}$ ) with albumin levels and age at menopause respectively.

## PgmNr 2995: Comprehensive characterization of genetic and phenotypic heterogeneity of obstructive sleep apnea across multiple United States clinics.

### Authors:

S.A. Pendergrass<sup>1</sup>; O.J. Veatch<sup>2,3</sup>; C.R. Bauer<sup>1</sup>; D.R. Mazzotti<sup>2</sup>; B.T. Keenan<sup>2</sup>; J.D. Robishaw<sup>4</sup>; K. Bagai<sup>3</sup>; B.A. Malow<sup>3</sup>; A.I. Pack<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biomedical and Translational Informatics, Geisinger, Danville, PA.; 2) Center for Sleep and Circadian Neurobiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Sleep Disorders Division, Department of Neurology, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Biomedical Science, Charles E. Schmidt College of Medicine, Florida Atlantic University, Boca Raton, FL

---

Obstructive sleep apnea (OSA) is a common, complex sleep disorder defined by frequent episodes of reduced or complete cessation of airflow during sleep. OSA is linked to negative health outcomes, and understanding the shared genetic mechanisms contributing to expression of symptoms and comorbidities in OSA may improve precision medicine and lead to new treatment strategies for this condition. We hypothesized that electronic health record (EHR) datasets would provide power to validate robust previously reported genomic variation associated with OSA risk and detect novel genetic relationships between these OSA variants and other comorbidities. We leveraged EHRs and biorepositories from Geisinger, Vanderbilt University Medical Center, and the eMERGE Network to conduct a cross-sectional, retrospective OSA study, including evaluating associations between these genetic variants and OSA phenotypes as well as phenome-wide evaluations using EHR data. We used 51 candidate single nucleotide polymorphisms (SNPs) identified through systematic literature review, and evaluated associations between these SNPs and an EHR-derived phenotypic algorithm defined OSA diagnoses, polysomnography data, and ICD-9 defined diagnoses. In individuals with genetically informed European ancestry, nine candidate SNPs were associated ( $p < 0.05$ ) with EHR-derived OSA, controlling for age, sex and body mass index. In individuals with African ancestry, two different candidate SNPs were associated with EHR-derived OSA. Results from phenome-wide association studies indicated that the majority of OSA candidate SNPs were associated with other EHR-derived traits reflecting cardiovascular disease, diabetes, anemia, autoimmune thyroid diseases, obesity, genitourinary symptoms, mental health disorders, pulmonary diseases, eye diseases and headache syndromes. Notably, 74% of patients with OSA and the most prevalent associated comorbidity, essential hypertension ( $OR_{\text{Geisinger}} = 4.40$  [4.26, 4.55],  $p < 4.4 \times 10^{-301}$ ;  $OR_{\text{VUMC}} = 3.71$  [3.43, 4.02],  $p = 9.91 \times 10^{-253}$ ), had their first essential hypertension code at least 5 years before their first OSA code ( $\mu = 5.28 \pm 4.52$  years). Our results highlight robust OSA-associated genomic variants, and provide knowledge of convergent mechanisms influencing risk for multiple disorders in the same individual. This knowledge can lead to better biomarkers of OSA and more personalized treatments for preventing OSA and related comorbidities.

## PgmNr 2996: Accounting for cell type composition in GTEx data.

**Authors:**

F. Takeuchi

[View Session](#) [Add to Schedule](#)**Affiliation:** Research Institute, Nat Cntr for Global Health & Med (NCGM), Tokyo, Japan

---

The Genotype-Tissue Expression (GTEx) project investigated 449 human donors across 44 tissues, and catalogued the genetic effect on gene expression as eQTLs. Whereas some eQTLs were shared across most of the 44 tissues, some were specific to a small subset of tissues. Gene expression was measured in bulk tissues (such as, pancreas), which indeed are mixtures of cell types (such as, alpha cell, beta cell, etc.). Some eQTLs might vary by cell type, in which case the average is observed at tissue level. Here, I account for cell type composition in GTEx data and refine the eQTLs by cell types. Cell types and their gene expression profile were obtained by published single-cell RNA sequencing experiments of human and mouse. By using gene expression profile of cell types as reference, the cell type composition of each tissue sample in GTEx was inferred. For this deconvolution, I developed a novel fast algorithm of mixture modeling. Using the inferred cell type composition, eQTLs in the GTEx data were refined.

# PgmNr 2997: A novel method for genome-wide scale phenome-wide G×E analysis and its application to UK Biobank.

## Authors:

W. Bi <sup>1,2</sup>; Z. Zhao <sup>1,2</sup>; R. Dey <sup>1,2,3</sup>; L.G. Fritsche <sup>1,2</sup>; B. Mukherjee <sup>1</sup>; S. Lee <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Biostatistics, University of Michigan, Ann Arbor, MI.; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA

The etiology of most complex diseases involves genetic variants, environmental factors, and gene-environment interaction (G×E) effects. Compared with marginal genetic association studies, G×E analysis requires more samples and detailed measure of environmental exposures, which limits the possible discoveries. Large-scale population-based biobanks with detailed phenotypic and environmental information, such as UK-Biobank, can be an ideal resource to identify G×E effects. However, due to the large computation cost and the presence of case-control imbalance, existing methods often fail. For example, to analyze G×E for one phenotype and one environmental factor, standard Wald test requires 2 CPU years for 20 million variants, 400,000 samples and 15 covariates. For 1,500 phenotypes, the computation cost increases to 3,000 CPU years.

Here we propose a scalable and accurate method, SPAGE (SaddlePoint Approximation implementation of G×E analysis), that is applicable for genome-wide scale phenome-wide G×E studies. SPAGE fits a genotype-independent logistic model only once across the whole-genome analysis to reduce computation cost and uses a saddlepoint approximation (SPA) to calibrate the test statistics for analysis of phenotypes with unbalanced case-control ratios. Simulation studies show that SPAGE is 33-79 times faster than the Wald-test and 72-439 times faster than the Firth's test and can control type I error rates at the genome-wide significance level even when case-control ratios are extremely unbalanced.

We applied SPAGE to UK Biobank dataset of 344,341 white British European-ancestry samples. We totally analyzed 79 environmental factor × phenotype combinations and identified 34 significant G×E signals at a genome-wide significance level 5E-8. For example, in the analysis of chronic airway obstruction, we identified a significant G×E effect of the smoking status and a variant in gene *CHRNA5* (rs55781567,  $p=2.87E-8$ ). Smoking is an important risk factor to the chronic airway obstruction and the gene *CHRNA5* is well-known to be associated with the smoking behavior and some smoking-related diseases such as chronic obstructive pulmonary disease. In the analysis of cardiac dysrhythmias, we identified a significant G×E effect of gender and variants near gene *PITX2* (rs1906609,  $p=1.42E-12$ ). In the analysis of Hyperlipidemia, we identified a significant G×E effect of the vigorous physical activity and variants in gene *DNAH11* (rs10950866,  $p=3.64E-9$ ).

## PgmNr 2998: Parent-of-origin effect analyses provide evidence for maternal genetic and maternal environmental effects on the genetic variability in type 1 diabetes and rheumatoid arthritis.

### Authors:

I. Blunk<sup>1</sup>; H. Thomsen<sup>2,3</sup>; N. Reinsch<sup>1</sup>; M. Mayer<sup>1</sup>; A. Foersti<sup>2,4</sup>; K. Sundqvist<sup>4</sup>; J. Sundqvist<sup>4,5</sup>; K. Hemminki<sup>2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Genetics and Biometry, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany; 2) Division of Molecular Genetic Epidemiology, German Cancer Research Centre (DKFZ), Heidelberg, Germany; 3) GeneWerk GmbH, Heidelberg, Germany; 4) Center for Primary Health Care Research, Lund University, Malmö, Sweden; 5) Stanford Prevention Research Center, Stanford University School of Medicine, Stanford, California, USA

Numerous studies have indicated that the development of autoimmune disorders including type 1 diabetes (T1D) and rheumatoid arthritis (RA) is not only determined by genetic and environmental but also by epigenetic factors. Genomically imprinted genes, which contribute to parent-of-origin effects (POEs), as their alleles are epigenetically inactivated depending on their parental origin, have often been mentioned in this context. However, findings have been contradicting and other POEs including maternal genetic effects and maternal environmental effects may have biased earlier findings as they are known to mimic imprinting if not considered in the model. A special type of linear mixed model with two random parental gametic effects was used to investigate the existence of imprinting effects on the susceptibility to T1D and RA and to disentangle their variation from the variation of other POEs. The inclusion of fixed effects, such as gender, birth year, and social economic index allowed the investigation of environmental effects. Through linkage of the Swedish Hospital Discharge Register and the Multigeneration Register, a suitable population database was available. For T1D the dataset contained 27,255 patients with 208,114 ancestors; for RA 15,850 patients with 60,684 ancestors were available.

With regard to T1D the imprinting variance did not turn out to be significant, however, the significance of the maternal environmental variance indicated the importance of familial effects explaining 18.8% ( $\pm 1.81\%$ ) of the phenotypic variance. Similar results were observed for RA as imprinting effects were not significant but significant variances were found for maternal effects. However, they could not clearly be determined to be either genetic or environmental as although both variances were significant, a considerable ratio of the phenotypic variance was explained by maternal genetic effects (14%;  $\pm 4.14\%$ ) whereas only 5% ( $\pm 6.81\%$ ) was explained by the maternal environment. The results indicated that imprinting does not play a great role for the susceptibility to T1D and RA but that POEs due to maternal genetic effects as well as maternal environmental effects have a considerable impact. Fixed effects turned out to be significant, which further supports the importance of environmental factors. The results implicate that POEs other than imprinting effects may have biased earlier results and that they must be taken into account in future imprinting studies.



## PgmNr 2999: When can population based imputation be improved upon by complementary identity-by-descent sharing information?

### Authors:

A.F. Herzig <sup>1</sup>; M. Ciullo <sup>2,3</sup>; A.-L. Leutenegger <sup>4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Inserm, Univ Brest, EFS, UMR 1078, GGB, F-29200 Brest, France; 2) Institute of Genetics and Biophysics A. Buzzati-Traverso - CNR, Naples, Italy; 3) IRCCS Neuromed, Pozzilli, Isernia, Italy; 4) Inserm, U1141, NeuroDiderot, Equipe GenMedStroke, F-75010, Paris, France; 5) Université Paris-Diderot, Sorbonne Paris Cité, Paris, F-75010, France

---

In population based imputation, there is an intuitive preference for a reference panel that contains haplotypes specific to the study population; the benefits of choosing such a panel have now been widely demonstrated. In certain situations, such as in the study of isolated populations, it is even possible to include reference panel haplotypes from individuals who are closely related to the target group; resulting in exceptionally high imputation accuracy.

Results of this nature indicate the potential importance of long matches between target and reference haplotypes, in other words regions that are likely to be shared in a state of identity-by-descent (IBD), whose detection is typically the basis of family based imputation methods. However, in most circumstances, IBD-sharing information will not be sufficient to impute missing genotypes across the whole genome as identifiable shared regions will not cover entire chromosomes.

This has led to the idea of combining IBD and population based imputation methods. Whilst a complete integration of the two has yet to be successfully demonstrated, two recent software have been put forward that provide two-step approaches: Ped-Pop and Kinpute. These methods seek to significantly improve upon the exactitude of imputed genotypes from population based methods by overlaying inference from IBD-sharing information. Here, we report on the functionality of these two software and perform tests on simulation data based on the structure of the known genetic isolates of Cilento in Southern Italy. This allows us to determine and provide examples of specific scenarios where these methods may be most or least successful.



# PgmNr 3000: Human leukocyte antigen class II genes associated with chronic hepatitis B virus infection in Taiwanese population.

## Authors:

Y.H. Huang<sup>1</sup>; S.F. Liao<sup>1,2</sup>; S.S. Khor<sup>3</sup>; Y.J. Lin<sup>1</sup>; C. Huang<sup>1,4</sup>; B.J. Liu<sup>1,4</sup>; H.Y. Chen<sup>5</sup>; Y.H. Chang<sup>5</sup>; C.Y. Shen<sup>2</sup>; K. Tokunaga<sup>3</sup>; M.H. Lee<sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Institute of Clinical Medicine, National Yang-Ming University, Taipei, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 3) National Center for Global Health and Medicine, Tokyo, Japan; 4) Faculty of Medicine, National Yang-Ming University, Taipei, Taiwan; 5) Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

## Background

Hepatitis B virus (HBV) infection is highly prevalent in East and Southeastern Asia, which is well documented as a risk factor for hepatocellular carcinoma. The associations between genetic variants and HBV chronicity remains inconclusive.

## Methods

The study included 15,352 subjects previously acquired HBV, defined by seropositive for hepatitis B core antibody in Taiwan Biobank (TWB). Among them, 2,591 (16.9%) were seropositive for hepatitis B surface antigen (HBsAg) and were defined as chronic HBV infected cases. The study subjects were randomly divided into genome-wide association study (GWAS) and replication phase in a three-to-one ratio. All the participants were examined by Axiom Genome-wide TWB Array. Subjects with a call rate <97%, minor allele frequency <1%, and violation of Hardy-Weinberg equilibrium ( $P < 10^{-4}$ ) were excluded in the subsequent analyses. The associations between single nucleotide polymorphisms (SNPs) and HBV chronicity were examined under an additive model adjusted for age and sex. 2-field *HLA* imputation was carried out on *DRB1*, *DQA1*, *DQB1*, *DPA1*, *DPB1* using HIBAG R package. Haplotype analysis was done by Arlequin software. The frequencies of *HLA* alleles between cases and controls were compared by the Fisher's exact test with p-value corrected by the number of alleles tested. Multiple logistic regression models were used to obtain adjusted odds ratios ( $OR_{adj}$ ) with 95% confidence interval (CIs) for their associations with HBV chronicity.

## Results

A total of 583,383 SNPs were compared among 9,002 HBV resolved controls and 1,810 chronic HBV infection cases in GWAS phase. There were 450 SNPs within *HLA* gene loci significantly associated with HBV chronicity ( $P < 8.6 \times 10^{-8}$ ). All of these SNPs could be validated in the replication phase ( $P < 0.05$ ). In the imputation analysis, *DPA1\*02:02* and *DPB1\*05:01* were significantly associated with HBV chronicity, with the  $OR_{adj}$  (95% CI) of 1.41 (1.07-1.84) and 1.62 (1.30-2.01), respectively. On the other hand, *DQA1\*01:02*, *DQB1\*06:09*, and *DRB1\*13:02* showed inversely associations with HBV chronicity, with the  $OR_{adj}$  ranging from 0.32 (0.17-0.60) to 0.68 (0.55-0.85). Individuals who carried *DPA1\*02:02-DPB1\*05:01* haplotype had increased probability ( $OR_{adj}=1.59$  (1.28-1.97)) to HBV chronicity, whereas *DRB1\*13:02-DQB1\*06:09* showed a negative association ( $OR_{adj}=0.33$  (0.18-0.63)).

## Conclusion

*HLA* class II gene variants contributed to susceptibility for HBV chronicity after the acquisition of HBV

in an endemic area.

## PgmNr 3001: Multi-resolution localization of causal variants across the genome.

### Authors:

M. Sesia<sup>1</sup>; G. Katsevich<sup>1</sup>; S. Bates<sup>1</sup>; E. Candès<sup>1,2</sup>; C. Sabatti<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Statistics, Stanford University, Stanford, California.; 2) Department of Mathematics, Stanford University, Stanford, California.; 3) Department of Biomedical Data Science, Stanford University, Stanford, California.

---

We present *KnockoffZoom*, a flexible method for the genetic mapping of complex traits at multiple resolutions. *KnockoffZoom* localizes causal variants precisely and provably controls the false discovery rate using artificial genotypes as negative controls. Our method is equally valid for quantitative and binary phenotypes, making no assumptions about their genetic architectures. Instead, we rely on well-established genetic models of linkage disequilibrium. We demonstrate that our method can detect more associations than mixed effects models and achieve fine-mapping precision, at comparable computational cost. Lastly, we apply *KnockoffZoom* to data from 350k subjects in the UK Biobank and report many new findings.

## PgmNr 3002: Large-scale integration of gene-level knowledge in GWAS using a regression-based approach.

### Authors:

S. Biswas; S. Pal; S. Bhattacharjee

[View Session](#) [Add to Schedule](#)

**Affiliation:** National Institute of Biomedical Genomics, Kalyani, India

---

Genome-wide Association Studies (GWAS) are under-powered to detect all associated variants, largely due to a huge multiple-testing burden. There is recent interest for conducting prioritized GWAS primarily using SNP-level functional annotations and statistical methods have been developed to enable incorporation of SNP-level knowledge. Another, rich source of biological knowledge exists in the form of gene-level functional annotations (e.g. pathways). Gene expression regulation through cis or trans expression Quantitative trait loci (eQTLs) can also serve as biological knowledge to inform GWAS. It is imperative to use knowledge of pathways/networks and information from gene-expression/eQTLs a-priori to 'inform' the genome-wide search. Conventional pathway enrichment analysis can help in the interpretation of findings; but does not by itself improve the power of detecting additional novel loci. Here, we develop a statistical method to enable a prioritized GWAS using knowledge from a very large number of gene-level annotations. Our method being based on penalized logistic regression is computationally efficient and scalable for routine use with large number of annotations. We demonstrate using simulations and real GWAS data on various diseases that prioritized GWAS using gene-level annotations can indeed gain substantial power over unbiased GWAS. Unlike most prioritization tools using FDR, our method maintains global false-positive rates at the same stringency as in a usual GWAS (i.e.  $p < 5 \times 10^{-8}$ ). We have demonstrated correct type I error as well as increase in overall power using simulations. We have created an R package **'GKnowMTest' (Genomic Knowledge-guided Multiple Testing)** for flexible incorporation of gene-level knowledge into GWAS. Further, we incorporated knowledge from other sources such as Gene-Expression Omnibus (GEO) to get differential-expression information or GTEx to extract eQTL information in order to prioritize GWAS. Using a novel whole genome simulation strategy, we were able to show substantial improvement in power of detection by exploiting gene-level knowledge. Further the method was applied to several real GWAS datasets and gave encouraging results in terms of detecting genomic regions missed by standard unweighted GWAS.

## PgmNr 3003: Leveraging pleiotropy in genome-wide association studies across multiple traits with per trait interpretations.

### Authors:

K. Collins <sup>1</sup>; E. Eskin <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, University of California, Los Angeles, CA; 2) Department of Human Genetics, University of California, Los Angeles, CA; 3) Department of Computational Medicine, University of California, Los Angeles, CA

---

Genome-wide association studies (GWAS) have been instrumental in identifying genetic variants associated with complex traits. As sample sizes in GWAS have increased, more and more variants have been discovered which suggests the presence of a large number of variants at low effect size. Additionally, many of these variants are shown to be associated with multiple traits. With this in mind, it is natural to assume there is more statistical power for identifying additional variants when traits are analyzed together.

We present PAT, a pleiotropic association test, for joint analysis of summary statistics across multiple traits. Our method assumes genetic effect sizes follow Fisher's polygenic model and are correlated according to genetic correlation estimates obtained from GWAS. We perform hypothesis testing to determine whether a variant is associated with at least one of the traits. Under the null model, summary statistics are derived from a variant with no genetic effect; therefore, any deviation in the summary statistic is fully attributed to environmental effects. Under the alternative model, however, there is at least one trait for which the variant has a non-zero genetic effect. This results in variation in the summary statistics to be the sum of genetic and environmental effects. For simulations that follow our assumed model, we find PAT has a substantial increase in power. We also find our method still has a moderate power increase in many scenarios where our model assumptions are violated.

While PAT is able to determine whether a variant is associated with the set of traits, it cannot state for which of the trait(s), the variant is associated. We extend the M-value framework to account for estimated genetic covariation to interpret PAT results. If we consider all possible models of genetic effect (ie. the variant affects no traits, all traits, or the variant affects only the first trait, etc.), we have  $2^m$  total models where  $m$  is the number of traits. We take the sum over all models as well as sum over the models that are compatible with the variant affecting a particular trait to compute the posterior probability of a variant affecting that trait. We analyze traits from the UK Biobank which shows that utilizing multiple traits increases power for detecting variants that affect more than one trait in real data.

## PgmNr 3004: Complex-Traits Genetics Virtual Lab: A community-driven web platform for post-GWAS analyses.

### Authors:

G. Cuellar Partida <sup>1</sup>; M. Lundberg <sup>1</sup>; A.I. Campos Gonzalez <sup>2</sup>; P.F. Kho <sup>2</sup>; S. D'Urso <sup>1</sup>; L.D. Hwang <sup>1</sup>; T.T. Ngo <sup>1</sup>; M.E. Renteria <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Queensland, Brisbane, Queensland, Australia; 2) QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

---

**Motivation:** For over a decade, genome-wide association studies (GWAS) have been an important method for mapping genetic variation underlying complex traits. With the ever-increasing volume of data generated however, new tools are needed to integrate the vast array of GWAS results and perform further analyses to maximize their utility for biological discovery.

**Results:** Here we present the Complex-Traits Genetics Virtual Lab (CTG-VL) — <https://genoma.io>. CTG-VL integrates these key components: **(i)** publicly available GWAS summary statistics; **(ii)** a suite of analysis and visualization tools; and **(iii)** several data sets for genomic annotations. This platform makes available results from post-GWAS analyses of >1,500 complex traits for assessing pleiotropy at the genetic variant through to gene-, pathway- and tissue-based levels. Further, CTG-VL integrates many analysis functions from other widely used web platforms (e.g., FUMA GWAS, LD-Hub). Here we demonstrate CTG-VL using summary statistics from a chronic pain GWAS (i.e., pain at any site >3 months) derived from UK Biobank (N=288K). Using Mendelian Randomization (MR) approaches implemented in the platform, we performed a phenome-wide MR analysis of chronic pain and found evidence for causal relationships with dozens of traits including inflammation biomarkers, type of diet and sleep.

**Conclusion:** CTG-VL is a freely available online web application to further harness GWAS data for research reproducibility, collaboration and translation.

## **PgmNr 3005: MF-TOWmuT: Testing an optimally weighted combination of common and rare variants with multiple traits using family data.**

**Authors:**

C. Gao; Q. Sha; S. Zhang; K. Zhang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Mathematical Sciences, Michigan Technological University, Houghton, Michigan.

---

With rapid advancements of sequencing technologies and accumulations of electronic health records, multiple disease phenotypes and a large number of genetic variants become available in many genetic association studies. Thus, it becomes necessary and important to develop new methods that can jointly analyze multiple traits and multiple genetic variants and help us to better understand genetic etiology of complex human diseases. Comparing with methods that only use a single trait or marker, the joint analysis of multiple genetic variants and multiple phenotypes is more powerful since such analysis can fully incorporate correlation structure of genetic variants and/or phenotypes and their mutual dependence pattern. However, most of existing methods are only applicable to unrelated samples. We propose a new method called MF-TOWmuT to detect association of multiple phenotypes and multiple genetic variants in a genomic region using an optimally weighted combination of variants with family samples. Our method can be applied to both qualitative and quantitative phenotypes and both rare and common variants. We evaluate and compare our method with MFQLS and mFARVAT-SKAT-O. Our simulation results show that (1) the type I error of our method is well preserved; (2) MF-TOWmuT is robust to the proportion of dichotomous or continuous traits and to the proportion of risk or protective variants; (3) MF-TOWmuT outperforms MFQLS and mFARVAT-SKAT-O in terms of power in extensive scenarios. We apply MF-TOWmuT to GoKind study - a study for searching susceptibility genes for diabetes and is able to find some novel susceptibility genes.

## PgmNr 3006: Combining case-control status and family history of disease increases association power.

### Authors:

M.L.A. Hujoel<sup>1</sup>; S. Gazal<sup>2,3</sup>; P.-R. Loh<sup>3,4</sup>; N. Patterson<sup>3</sup>; A.L. Price<sup>1,2,3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA.; 2) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA.; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Brigham and Women's Hospital / Harvard Medical School, Boston, MA.

---

Family history of disease can provide valuable information about an individual's genetic liability for disease in case-control association studies, but it is currently unclear how to best combine case-control status and family history of disease. We developed a new association method based on posterior mean liabilities under a liability threshold model, conditional on both case-control status and family history (LT-FH); association statistics are computed via linear regression of genotypes and posterior mean liabilities, equivalent to a score test. LT-FH captures the fact that controls with family history have higher genetic liability than controls without family history (and likewise for disease cases). In contrast to GWAX (Liu et al. 2017 Nat Genet), which assigns the same binary phenotype to all cases and "proxy cases" (controls with family history of disease) to increase power for diseases of low prevalence, LT-FH accurately models a broad range of case-control status and family history configurations, greatly increasing power for diseases of both low and high prevalence.

We applied LT-FH to 12 diseases from the UK Biobank for which family history (parental and sibling history) was available for most target samples (average  $N=350K$ ). We compared LT-FH to genome-wide association without using family history (GWAS) and GWAX. LT-FH was +56% (block-jackknife s.e. 6%) more powerful than GWAS and +36% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX (+37% for lower-prevalence diseases and +35% for higher-prevalence diseases), based on the number of independent genome-wide significant loci detected across all diseases (e.g. 693 independent loci for LT-FH vs. 445 for GWAS). The second best method was GWAX for lower-prevalence diseases and GWAS for higher-prevalence diseases, consistent with simulations. We also confirmed that LT-FH was well-calibrated (assessed via the stratified LD score regression attenuation ratio), consistent with simulations. When using BOLT-LMM (Loh et al. 2018 Nat Genet) instead of linear regression to compute association statistics for all three methods (increasing the power of each method), LT-FH was +59% (s.e. 6%) more powerful than GWAS and +38% (s.e. 4%) more powerful than the trait-specific maximum of GWAS and GWAX. In summary, LT-FH greatly increases association power in case-control association studies when family history of disease is available.



## PgmNr 3007: BLUP-based mixed-model approach for biobank-scale GWAS.

### Authors:

J. Jiang<sup>1</sup>; L. Ma<sup>2</sup>; P.M. VanRaden<sup>3</sup>; J.R. O'Connell<sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 2) Department of Animal and Avian Sciences, University of Maryland, College Park, MD; 3) AGIL-ARS-USDA, Beltsville, MD

Linear mixed models (LMM) are a preferred approach to account for population structure and/or relatedness in genome-wide association studies (GWAS); however, scaling LMM approaches to biobank-scale sequence data is challenging. Software such as BOLT and SAIGE address the challenges of big sample size,  $N$ , and large numbers of sequence variants,  $M$ , by using a corrected linear regression to approximate the LMM score statistic to achieve time complexity linear in  $N$ . The approach is not without issues. Fitting the null model and computing the correction factor(s) are computationally intensive and the approximation may result in inaccurate test statistics and lose power. To address these problems, we develop a fast and accurate LMM approach based on best linear unbiased prediction (BLUP).

Our approach has three steps, 1) variance component estimation (VCE), 2) SNP-BLUP, and 3) association testing. The VCE step makes use of a recently proposed randomized Haseman-Elston (HE) regression estimator with sub-linear time complexity in  $N$ . The SNP-BLUP step fits a smaller set of  $M_p$  markers as random effects in LMM, based on HE estimates. We use a block-wise Gauss-Seidel (GS) method to compute BLUP (?) of  $M_p$  marker effects. To avoid inversion of the large  $M_p \times M_p$  matrix to compute  $\text{var}(?)$ , we set our GS block size ranging from 1K to 5K, which is less than a few percent of  $M_p$ . Our  $?^2/\text{var}(?)$  is equivalent to LMM score statistic. To perform association tests, we scan the variants block by block, and compute ? and  $\text{var}(?)$  for each of them based on the SNP-BLUP output. Association testing has time complexity that scales linearly in block size,  $N$  and  $M$ .

To test the accuracy of our approach, we simulated many data sets ranging from 5K to 20K subjects under different genetic models. Genotypes were simulated in unrelated, admixed, and related samples using HAPGEN2, bnpsd, and genosim, respectively. We then simulated major QTL effects, polygenic effects, and random errors, and summed them to construct quantitative phenotypes. Our approach produced nearly identical results ( $r^2 \approx 0.999$  for chi-square statistics) compared with exact results from EMMAX. We also compared our approach with SAIGE using dairy cattle data with 300K cows and  $M_p=60K$  SNPs. We identified all the QTLs that SAIGE found but also a few known QTLs that failed to pass Bonferroni threshold in SAIGE. More importantly our software took <2 hours on a typical PC while SAIGE used >200 hours.

## PgmNr 3008: A unified framework for detecting pleiotropic loci.

### Authors:

C. Lee <sup>1,2</sup>; H. Shi <sup>3</sup>; P. Bogdan <sup>4,5</sup>; H. Lee <sup>6</sup>; E. Eskin <sup>4,7</sup>; B. Han <sup>1</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Medical Sciences, Seoul National University College of Medicine, Seoul, Republic of Korea; 2) Department of Convergence Medicine, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea; 3) Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, California, USA; 4) Department of Human genetics, University of California, Los Angeles, Los Angeles, California, USA; 5) Department of Pathology and Laboratory Medicine, University of California, Los Angeles, Los Angeles, California, USA; 6) Department of Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, Republic of Korea; 7) Department of Computer Science, University of California, Los Angeles, Los Angeles, California, USA

---

Pleiotropy refers to a phenomenon that the same genes or variants affect more than one trait simultaneously. Pleiotropic loci can give us intuition about shared pathways and etiology of multiple traits. To detect pleiotropic loci, several methods have been proposed. These cross-trait analysis methods aggregate summary statistics of multiple traits together, of which an added benefit is the increased effective sample size. However, these methods have limited flexibility in that the traits have to be in the same category, either qualitative or quantitative. For example, no systematic methods have been suggested to aggregate odds ratios for dichotomous traits and regression coefficients for continuous traits together. Moreover, even in the same continuous category, traits may differ in their units. Some traits may be morphological indices, some may be lipoprotein densities, and some may be confidence scores for questionnaires in the survey. In these situations, a method assuming a shared underlying distribution of effect sizes can suffer from the inevitable violation of the assumption. Furthermore, the results will be dependent on the unit scales, such as kilogram versus gram. Here, we introduce DELPY (DEtecting Loci with PleiotropY), a novel unified framework for joint analysis of multiple traits to detect pleiotropic loci. Given summary statistics of multiple traits, our method projects the statistics into the space representing genetic liabilities for dichotomous traits and the space representing standardized phenotypes for continuous traits. This way, different categories are seamlessly put together and the outcomes are independent of the units and scales. A benefit of using this projection is that this space is the basis for heritability and coheritability calculations. Therefore, we can utilize the estimated genetic covariance as the prior information for effect sizes to increase statistical power. In addition to genetic covariance, our method estimates and accounts for the environmental covariance due to sample overlap or hidden relatedness. Given the genetic and environmental covariance matrices, our method tests the genetic contribution of a variant toward the overall heritability of multiple traits using likelihood ratio test. Since the asymptotic p-value is only accurate when the number of traits is large, which is rarely the case, we developed an efficient importance sampling procedure to rapidly assess the significance of test.

## PgmNr 3009: Mapping of novel protein quantitative trait loci implicates new biology and therapeutic targets in complex diseases: The SCALLOP consortium.

### Authors:

A. Malarstig<sup>1</sup>; S. Gustafsson<sup>2</sup>; Q. Wang<sup>3</sup>; J. Peters<sup>4</sup>; E. McDonald-Dunlop<sup>5</sup>; N. Ericsson<sup>2</sup>; S. Enroth<sup>2</sup>; A. Kalnainen<sup>6</sup>; D. Aaen-Hansen<sup>7</sup>; D. Zhernakova<sup>8</sup>; A. Schork<sup>9</sup>; Y. Wu<sup>10</sup>; E. Fauman<sup>11</sup>; T. Esko<sup>6</sup>; J. Fu<sup>8</sup>; U. Gyllenstein<sup>2</sup>; C. Hayward<sup>5</sup>; M. Landen<sup>12</sup>; O. Melander<sup>13</sup>; A. Siegbahn<sup>2</sup>; G. Smith<sup>13</sup>; L. Wallentin<sup>2</sup>; J.F. Wilson<sup>5</sup>; A. Butterworth<sup>4</sup>; M. Holmes<sup>14</sup>; E. Ingelsson<sup>15</sup>; L. Folkersen<sup>9</sup>; SCALLOP consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Karolinska Institute, Stockholm, Sweden; 2) Uppsala University, Uppsala, Sweden; 3) Baker Heart & Diabetes Institute, Melbourne, Australia; 4) Cambridge University, Cambridge, United Kingdom; 5) University of Edinburgh, Edinburgh, United Kingdom; 6) University of Tartu, Tartu, Estonia; 7) Intomics, Copenhagen, Denmark; 8) University of Groningen, Groningen, Netherlands; 9) St Hans Hospital, Copenhagen, Denmark; 10) University of Brisbane, Brisbane, Australia; 11) Pfizer Worldwide R&D, Cambridge, MA, USA; 12) University of Gothenburg, Gothenburg, Sweden; 13) Lund University, Malmö, Sweden; 14) Oxford University, Oxford, United Kingdom; 15) Stanford University, Stanford, CA, USA

---

Circulating proteins play vital roles in human physiology but their characterisation at scale, including inter-individual differences due to DNA variation, regulatory pathways, and potential causal roles in disease is incomplete. We mapped protein quantitative trait loci (pQTL) for 90 circulating proteins, measured using the Olink proximity extension assay CVD-I panel, by meta-analysing 13 genome-wide association studies encompassing up to 21,758 participants. Primary signals were replicated in two independent cohorts encompassing 9,500 subjects. A total of 467 pQTLs for 85 proteins were identified and replicated, including 123 signals significant after conditioning for lead variants in the same loci. By integrating pQTLs with curated protein-protein interaction networks and text-mining, we show examples of molecular processes involved with the regulation of circulating proteins. Using Mendelian randomization and mining of drug databases, we highlight several protein-outcome associations with established development (e.g. IL1RA in rheumatoid arthritis), with predicted successful drug development (e.g. IL6RA in coronary heart disease), and ?drugs in early development that target proteins with a causal role in biology (e.g. DDK1 and osteoporosis).

## PgmNr 3010: Considerations for tests of disease risk and age at onset.

### Authors:

E.R. Martin <sup>1,2</sup>; X.R. Gao <sup>3,4</sup>; Y. Li <sup>5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dr. John T. Macdonald Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Ophthalmology and Visual Science, The Ohio State University, Columbus, OH; 4) Department of Biomedical Informatics, Division of Human Genetics, The Ohio State University, Columbus, OH; 5) Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC; 6) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC

---

Genetic studies often look for genes that influence one's *risk* of developing disease, typically by comparing genotype (or allele) frequencies between affected and unaffected individuals. Instead of modeling disease risk, we may look for genes that control the timing of disease onset, or *age at onset* (AAO). We can test for AAO effects by looking for association between genotypes (or alleles) and AAO as a quantitative variable in a group of cases. These two categories are not disjoint; a gene that influences AAO might also appear to influence risk. For example, if an allele lowers AAO, then we would expect that allele to be enriched in a set of cases (relative to controls) taken at any cross-sectional timepoint. In this scenario, when an allele influences both AAO and risk, a reasonable question is whether we would have more power to detect association using a case-control design to test risk or a case-only design to test AAO, or some alternative.

To address this question, we propose a genetic model for AAO, and use this to analytically compute the required sample size for 80% power for the Cochran-Armitage trend case-control test for risk and a linear regression case-only test for AAO. We also use simulations to compare power of these tests with a 2-df joint test (which combines the risk and AAO statistics) and the Cox Proportional Hazards (CoxPH) model testing AAO with censored data in controls.

We find that neither risk nor AAO test is most powerful in all situations. In general, when there is little heterogeneity, the case-control risk test has more power than the case-only AAO test (with equivalent sample sizes), but when the model is complex (e.g., with heterogeneity or reduced penetrance), this pattern reverses and the case-only AAO test tends to have more power than the risk test. Our simulations show that in many scenarios, the 2-df joint test outperforms either test alone, particularly as heterogeneity increases. When there is reduced penetrance, case-only tests are to be preferred, and we find that the CoxPH test with cases only has more power than the linear-regression based case-only test.

The results have implications for choice of test statistic(s), study design and interpretation. If genotype data from both cases and control are available, then using all data in a joint test is generally most powerful. However, with substantial heterogeneity, putting resources into a case-only design might be prudent, and the CoxPH approach may be preferred.



## PgmNr 3011: Multi-allelic method for genome-wide association studies: The kernel trick.

### Authors:

K. Misawa <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Section of Genome Analysis, Institute of Biomedical Science, Kansai Medical University, Hirakata, Osaka, Japan; 2) Center for Advanced Intelligence Project, Riken, Tokyo, Japan; 3) Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan

---

Genome-wide association studies (GWAS) typically focus on associations between single-nucleotide polymorphisms (SNPs) and traits like major human diseases. In many cases of GWAS, the reference allele is defined as “0”, whereas the alternative allele as “1”. In datasets of large samples, many variant sites are observed to be multi-allelic. For example, 274,620 sites are multi-allelic among genomic sequences of 2,504 people determined by 1000G. The analysis of multi-allelic sites is currently often ignored in GWAS and sequence-based association studies. In this study, a method for conducting statistical tests on multi-allelic sites is proposed. This method treats sequence identity between two sequences as a kernel function. Kernel function defines inner product in the sequence space. It is proven that sequence identity kernel is equivalent to previous methods when all the sites are biallelic. Thus, the proposed method is a natural extension of previous methods. This kernel method can be used in support vector machine, analysis of principal components, and sequential kernel association test. The proposed method enables to study large samples with multi-allelic sites for GWAS.

## PgmNr 3012: Meta-PheWeb: Meta-analysis and cross-comparison between PheWAS datasets.

**Authors:**

P. VandeHaar; S. Gagliano Taliun; A. Boughton; D. Taliun; G. Abecasis

[View Session](#) [Add to Schedule](#)

**Affiliation:** Biostatistics, Univ MICHIGAN, Ann Arbor, Michigan.

---

To allow intuitive browsing, comparison and interpretation of continuously growing publicly-available genetic association results across a wide range of complex traits we developed a free online tool, PheWeb. This web-based tool allows users to view summary statistics for thousands of phenotypes in datasets such as the UK Biobank and the Michigan Genomics Initiative. Users can also view associations to a gene or variant across all phenotypes in several datasets and can also compare and meta-analyze similar phenotypes across datasets. Meta-PheWeb is available at <http://pheweb.sph.umich.edu> and source code is at <https://github.com/statgen/pheweb> .

## PgmNr 3013: Gene-based analysis using gene expression data and polygenic risk score.

**Authors:**

S. Yan; Q. Sha; S. Zhang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Mathematical Sciences, Michigan Technological University, Houghton, MI.

---

Conventional genome-wide association studies (GWASs) have been successfully applied to identify the association of genetic variants with phenotypes. In spite of many successes, there existing two major challenges: one is missing heritability of complex diseases, due to the polygenic effects and small effect sizes; the other one is the difficulty of interpretation because some identified genetic variants are not in protein-coding regions.

Recently, gene-based association studies show that integrating GWAS with expression quantitative trait locus (eQTL) data can boost statistical power and many studies also have shown the advantage of using polygenic risk scores (PRS) to evaluate the putative association to biologically informative features. Since many complex diseases are associated with gene expression which is influenced by genetic variants, and the genetic liability of phenotypes can be captured by PRSs, using both eQTL data and PRS may gain more statistical power to identify the underlying association. In this presentation, we propose a set-based statistical method that use integrated gene expression measurements and PRS to identify genes that are associated with a phenotype of interest. We use a generalized linear model to associate a phenotype with gene expression and PRS, use a score test statistic to test the association between the phenotype and the gene set.

Simulation studies, as well as a real data application, are used to compare the proposed method with some existing methods that use either gene expression data or PRS to test the association between a gene set and a phenotype. Our simulation studies show that the proposed method has correct type I error rates, the power is boosted by using both integrated gene expression measurements and PRS comparing to other methods. The real data analysis shows that the proposed method is applicable to GWAS and is more powerful than the methods we compare with.



## **PgmNr 3014: Imputation of missing genotypes and estimation of relatedness between subjects without genetic data across pedigrees.**

### **Authors:**

M. Saad <sup>1</sup>; E. Ullah <sup>1</sup>; E.M. Wijsman <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar; 2) Division of Medical Genetics, Department of Medicine, and Department of Biostatistics, University of Washington, Seattle, USA

---

Family-based imputation allows better imputation of rare variants compared to population-based imputation. With whole genome sequencing becoming inexpensive, there is an opportunity to investigate rare variation. In the search for disease-associated rare variation, family-based designs have again become common, because rare, highly-penetrant genotypes can segregate in pedigrees.

In pedigrees, for subjects without genotype data, phenotype data can be available. Family-based imputation has the strength of imputing genotypes on such subjects using the genotypes of the available relatives. This likely increases the sample size and therefore the statistical power if imputation is accurate. In association testing, when related subjects are included, a kinship matrix must be used to account for relatedness. Relationships may be known from pedigree structure, or can be inferred using observed genotypes. For subjects without genotype data, however, no approach is able to infer relatedness, which is crucial to control the type 1 error.

In this work, we assess the performance of family-based imputation (GIGI2) for imputing subjects with no genotype data. We also propose a solution for inferring relatedness between such subjects by incorporating posterior probabilities of missing genotypes in an Expectation-Maximization approach. Through simulation, we obtained an average correlation of 0.6 between imputed and observed subjects. Correlation reached one for some subjects. Imputation performance decreased with size of imputation reference panel. Our approach succeeded in inferring many relationship types. The average kinship estimates were 0.19, 0.1267, 0.0641, and 0.03068 for underlying kinships of 0.25, 0.125, 0.0625, and 0.03125, respectively, for pairs of subjects without genotype data.

## PgmNr 3015: Evaluation of polygenic risk scores and machine learning methods for the prediction of complex diseases.

### Authors:

Gao<sup>1,2,3</sup>; Y. Li<sup>4,5</sup>; E. Martin<sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Ophthalmology and Visual Science, The Ohio State University, Columbus, Ohio.; 2) Department of Biomedical Informatics, The Ohio State University, Columbus, Ohio.; 3) Division of Human Genetics, The Ohio State University, Columbus, Ohio.; 4) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; 5) Duke Molecular Physiology Institute, Duke University, Durham, NC; 6) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL

---

Polygenic risk scores (PRSs) examine the cumulative effect of genetic variants on a disease or trait by aggregating many individual variants, typically weighted by their effect sizes, into a single measure. PRSs provide a convenient way to quantify the overall genetic risk estimation for a particular disease. Suppose  $m$  independent single nucleotide polymorphisms (SNPs) are obtained from genome-wide association studies, weighted PRS is defined as  $PRS = \sum_m (\beta_i * G_i)$ , where  $\beta_i$  and  $G_i$  are the effect size from a genetic association analysis and the allele dosage for each SNP  $i$ , respectively. Though the PRS calculation is straightforward, presently there is no consensus on how to select the best set of SNPs for optimal prediction accuracy. Many reports have used the genome-wide significance threshold,  $P < 5 \times 10^{-8}$ , to select SNPs for calculating PRSs. In other reports,  $P < 5 \times 10^{-5}$  was used for selecting SNPs. There are also a number of methods that use Bayesian methods, including LDpred, to derive the proportion of SNPs that have non-zero effect sizes. However, they also may not give the best prediction accuracy in certain situations. Furthermore, the majority of predictive models have been based on classic regression models. We conducted this study with 435,678 European participants from the UK Biobank dataset. We evaluated a number of PRS calculation approaches using a grid of  $p$ -values (0.01, 0.001,  $10^{-4}$ ,  $5 \times 10^{-5}$ , and  $5 \times 10^{-8}$ ) and LD-clumping cutoffs ( $r^2 = 0.2$  and 0.5) for selecting SNPs based on the results from our previous report on intraocular pressure (IOP) and glaucoma risk prediction (Gao *et al.* 2019). We also compared different predictive models, such as logistic regression and Xgboost (an efficient machine learning approach). The weighted IOP PRS constructed based on SNPs with  $P < 10^{-4}$  and LD-clumping  $r^2 = 0.2$  gave the best performance in predicting glaucoma among the  $p$ -value and LD-clumping cutoffs that we tested. Furthermore, Xgboost gave better prediction accuracy than logistic regression in our tests using cross-validation and independent testing samples (AUC increased by 1%,  $P < 0.001$ ). We are further evaluating these methods using the Alzheimer Disease Genetics Consortium (ADGC) dataset. Our results provide guidelines for constructing PRSs and apply extrapolative models for the prediction of complex diseases.

## **PgmNr 3016: A unifying framework for region-based association testing in family-based designs without the need for asymptotic assumptions or approximations, including higher criticism approaches, SKATs, multivariate and burden tests.**

### **Authors:**

J. Hecker <sup>1,2</sup>; F. William Townes <sup>1</sup>; P. Kachroo <sup>2</sup>; M. Cho <sup>2</sup>; J. Lasky-Su <sup>2</sup>; S. Weiss <sup>2</sup>; N. Laird <sup>1</sup>; C. Lange <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Harvard T.H. Chan School of Public Health, Boston, MA, USA; 2) Brigham and Women's Hospital, Boston, MA, USA

---

For the region/set-based association analysis of rare variants in family-based studies, we propose a general methodological framework that integrates higher criticism approaches, SKATs, burden, and multivariate tests into the FBAT approach. Using the haplotype algorithm for FBATs to compute the conditional genotype distribution under the null hypothesis of Mendelian transmissions, virtually any association test statistics can be implemented straightforwardly in our approach, requiring no assumptions about the asymptotic distribution or approximations. Given the conditional joint distribution of the rare variants, simulation-based or exact p-values can be computed without the need to estimate the genetic correlation/linkage disequilibrium (LD) structure between variants empirically. Using simulations, we compare the features of the proposed test statistics in our framework with the existing region-based methodology for family-based studies under various scenarios. Under realistic assumptions, the tests of our framework outperform the existing methodology. We provide general guidelines for which scenarios, e.g., the sparseness of the signals or local LD structure, which test statistic will provide distinct power advantages over the others. We illustrate our approach in an application to a whole-genome sequencing dataset with 900 asthmatic trios.

## PgmNr 3017: Haplotype analysis of more than 450,000 US veterans in the VA's Million Veteran Program (MVP), and its impact to imputation and GWAS.

### Authors:

S.G. Ji <sup>1,2</sup>; B. Gorman <sup>1,3</sup>; Y. Shi <sup>1</sup>; M. Xu <sup>1,3</sup>; W. Mulla <sup>1,3</sup>; A. Liem <sup>1,3</sup>; P. Tsao <sup>4,5</sup>; S. Pyarajan <sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) VA Boston Healthcare System, Boston, MA; 2) Seven Bridges Genomics Inc., Charlestown, MA; 3) Booz Allen Hamilton Inc., McLean, VA; 4) VA Palo Alto Healthcare System, Palo Alto, CA; 5) Department of Medicine, Stanford University School of Medicine, Stanford, CA; 6) Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA

---

The Million Veteran Program (MVP) has recently released the raw and imputed genotypes of more than 450,000 ancestrally diverse individuals. Here, we conducted a haplotype analysis to understand its impact on phasing, imputation, and ultimately GWAS across 275,000 European Americans (EA), 80,000 African Americans (AA), and 34,000 Hispanic/Latinos (H/L). As expected, significantly different distribution of identical-by-descent (IBD) segments were observed, where H/L had the greatest enrichment of large IBD segments. The decay of phasing accuracy as a function of distance between two phased variants were slower for AA and H/L compared to EA, showing the positive effect of longer haplotype blocks from recent admixture to better phasing performance. Within-cohort phasing showed better consistency and accuracy for AA compared to EA when sample size was matched and larger than 25,000. A mean switch error rate (SER) of 15% was observed when using only EA to phase AA, whereas using only AA to phase EA showed a 1% increase in mean SER. For the large EA cohort, the effect of increasing the number of conditioning haplotypes (k-hap) plateaued with increasing number of samples, whereas phasing performance of AA increased consistently with increasing k-hap up to 80,000 individuals. However, the increase in phasing performance did not directly lead to a meaningful improvement in imputation performance. A significant difference was observed depending on the reference panel used for imputation. We compared the 1000 Genomes Project Phase 3 (G1K), the Haplotype Reference Consortium (HRC), and the African Genome Resource (AGR). Imputation quality was assessed using precision-recall on a subset of samples that were whole genome sequenced. Although HRC includes G1K, using HRC to impute AA samples showed worse performance than using G1K alone. The incorrectly imputed variants were enriched in local segments of non-European ancestry. On the other hand, using AGR was consistently better than G1K in all MVP ancestries, with the largest improvement on AA individuals. Although ancestry-matched reference panels have the best imputation performance, we show that AGR is a more generic panel that performs better than G1K across all ancestries considered in MVP. Finally, we tested the effect of the optimized imputation to the GWAS of height. The effect of improved imputation metrics correlated well with improved heritability estimates and inflation metrics across all ancestries.

## PgmNr 3018: MetaMinimac2: An efficient tool to combine genotype data imputed against multiple reference panels.

### Authors:

K. Yu <sup>1</sup>; S. Das <sup>2</sup>; G. Abecasis <sup>1</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) 23anMe, Mountain View, CA

---

Genotype imputation has become an integral tool in genomewide association studies. Imputation uses a reference panel of densely sequenced samples to infer unobserved genotypes for study samples with genotype array data. With increasing availability of whole genome sequence datasets, investigators have access to a multitude of reference panel choices. It is desirable to construct a combined reference panel from all available sequencing studies to achieve optimal imputation accuracy, but challenging in practice due to data sharing and use limitations.

Here, we describe MetaMinimac2, a novel method which allows for imputation results generated using different reference panels to be combined into a consensus dataset without requiring the sharing of individual-level data for underlying panels. The meta-imputed results are a weighted average of the imputed allele counts from different panels. The weights are individual and region specific, reflecting the optimal choice of the reference panel which varies along the genome. Weights for each region and individual are estimated through a hidden Markov model, utilizing corresponding diagnostic metrics reported by our minimac4 imputation software.

We tested our method by meta-imputing 61 African-American genomes against two homogenous reference panels comprised of 600 African samples and 503 European samples. The results show that meta-imputation consistently outperforms imputation using a single reference panel and performs as well as imputation using a combined panel. For very rare variants (with minor allele frequency of ~0.1%), the imputation  $R^2$  was 42.6% using meta-imputation versus 42.4% using a combined panel, and 0.9% using only the European panel or 33.3% using only the African panel.

An additional benefit of meta-imputation is in reduced computation time with no decrease in variant quality when imputing against large panels. We randomly split the 60,039 genomes in the TOPMed Freeze 5 panel into three non-overlapping subsets, imputed against each subset, and then merged the imputation results. We saved 45.7% of the net computational time with negligible loss in imputation accuracy, compared with imputing directly against the full panel.

In summary, we have developed a flexible method that enables researchers to efficiently merge imputation results from multiple panels to improve the overall accuracy of imputed genotypes without the need to access individual genotype data for the underlying reference samples.

## PgmNr 3019: Defining and characterizing pleiotropy across 57 common diseases and complex traits.

### Authors:

L. O'Connor<sup>1,2</sup>; A. Price<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Harvard T. H. Chan School of Public Health, Boston, Massachusetts; 2) Broad Institute of MIT and Harvard, Cambridge, Massachusetts

---

Genetic variants implicated by GWAS often affect multiple traits, a phenomenon known as pleiotropy. Existing methods to quantify pleiotropy capture only a fraction of pleiotropic signal, either because they only measure signed relationships (Bulik-Sullivan et al. 2015b Nat Genet) or because they restrict to genome-wide significant SNPs (Pickrell et al. 2016 Nat Genet). Thus, new approaches are needed to define and estimate genome-wide pleiotropy.

We define the *excess squared effect size correlation* ( $Q_e$ ), quantifying the excess correlation in squared standardized effect sizes between two traits beyond what is expected based on their signed genetic correlation ( $r_g$ ). Intuitively;  $Q_e$  quantifies the similarity between Manhattan plots for each trait. We developed a method, cross-trait LD 4th moments regression (C-LD4M), to estimate  $Q_e$  by regressing products of  $\chi^2$  statistics for two traits on LD 4th moments (sums of  $r^4$ ). We confirmed via extensive simulations that C-LD4M produces unbiased estimates of  $Q_e$ .

We applied C-LD4M to 57 diseases and complex traits (max  $N=460K$ ), yielding three main observations. First, a majority (>60%) of all trait pairs have  $Q_e$  estimates that are significantly positive (FDR<5%) but very small ( $Q_e<0.05$ ). Thus, shared genetic risk variants do not imply meaningful shared biology. Second, complex traits with shared causal tissues consistently have high  $Q_e$ , even when  $r_g$  is low. In particular, we detected strong pleiotropy among 12 brain-related traits (avg  $Q_e=0.41$ ; e.g.  $Q_e=0.80$  vs.  $r_g=0.16$  for # children and # sexual partners), and among 12 immune-related traits (avg  $Q_e=0.19$ ; e.g.  $Q_e=0.37$  vs.  $r_g=-0.04$  for celiac disease and primary biliary cirrhosis). Although there was little pleiotropy between brain- and immune-related traits (avg  $Q_e=0.03$ ), an exception was ulcerative colitis, which had high  $Q_e$  with several brain-related traits, consistent with the hypothesized role of psychological stress. Third, highly polygenic traits are highly pleiotropic, with a correlation of 0.77 across traits between polygenicity and average  $Q_e$ . We recently showed that extreme polygenicity is explained by negative selection, which flattens the distribution of heritability across the genome (O'Connor et al. in press AJHG). We hypothesize that negative selection likewise amplifies pleiotropy between traits by flattening strong genetic signals in trait-specific genes and pathways, leaving behind weaker genetic signals that are often tissue specific and pleiotropic.

# PgmNr 3020: A powerful gene-set analysis method identifies novel associations and improves interpretation in UK Biobank.

## Authors:

D. Dutta<sup>1,2</sup>; S. Lee<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Biostatistics, University of Michigan, Ann Arbor, Michigan.; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan

---

Biologically functional gene-set association (GSA) analysis can be complementary to single-variant or single-gene test and can provide insights into the genetic architecture of complex diseases. Existing GSA methods have low statistical power, especially when only a small fraction of the genes are associated with the phenotype. Additionally, since most of the existing methods cannot identify active genes within the gene-set, interpreting results is challenging. Here, we introduce Gene-Set association Using Sparse Signals (GAUSS), a method for GSA with summary statistics which additionally selects the subset of genes with maximum association signals as the active genes. The simulation-based p-value for GAUSS can be efficiently calculated by using pre-computed correlation structure of test statistics from a reference data. Numerical experiments show that GAUSS can increase power over several existing methods while controlling type-I error. We analyzed summary statistics from the UK-Biobank data for 1201 phenotypes with 10679 gene-sets to demonstrate that GAUSS can identify novel associations across a large number of phenotypes and gene-sets. For example, GAUSS detected two fatty-acid related gene-sets associated with E.Coli-infection (p-values  $< 10^{-06}$ ) which illuminate the antibacterial role of fatty-acids. Additionally, GAUSS allows us to investigate active genes for different phenotypes using phenome-wide analysis of a given gene-set, which has been unexplored till now. This highlights genetic and mechanistic similarities or differences between the phenotypes. For example, the association of ATP-binding-cassette transporter gene-set in KEGG database with digestive diseases like Cholelithiasis is driven by *ABCG5* while those with Celiac disease and other intestinal malabsorptions are driven by *TAP2*. The novel associations detected by GAUSS along with the information on active genes and its computational scalability make it an attractive choice to perform GSA.

## PgmNr 3021: Multi-phenotype transcriptome-wide association study using summary statistics identifies new genes associated with lipid phenotypes.

### Authors:

H. Feng <sup>1,2</sup>; N. Mancuso <sup>3,4</sup>; B. Pasaniuc <sup>3,4</sup>; P. Kraft <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biostatistics, Harvard Chan School of Public Health, Brookline, Massachusetts.; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States of America; 3) Department of Human Genetics, University of California Los Angeles, Los Angeles, California, United States of America; 4) Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, California, United States of America

---

Tests that combine evidence for association between single SNPs and multiple phenotypes are well established. Here, we extend these methods to transcriptome-wide association (TWAS) to improve power of detecting genes associated with multiple phenotypes regulated through similar pathways. We show analytically that the multi-phenotype TWAS has a similar form as a standard multi-phenotype marginal SNP association statistic. Thus, established methods for combining marginal-SNP association tests across multiple traits (e.g., SUM, Wald, or ASSET) can be directly extended to TWAS. We evaluate the Type I error rate and power of multi-phenotype methods via simulation across different residual covariance structures. We show that multi-phenotype TWAS improves the power of traditional single-phenotype TWAS, with the Wald test being near-optimal in most of the scenarios. However, we observed no uniformly best multi-phenotype TWAS method, since the relative power of the methods varies across different alternatives. We showcase these methods in empirical data from the Global Lipids Genetics Consortium of 4 traits (LDL, HDL, TG, and total cholesterol) with cross-tissue weights built with sparse canonical correlation analysis on GTEx gene expression data. The joint analysis identified additional trait-associated genes (222%, 137%, 198% and 132% for the 4 traits respectively and 14% compared to Bonferroni correction). The set of genes identified by joint testing methods also provided new information into the gene regulation architecture for these traits.



## PgmNr 3022: Mendelian randomization accounting for horizontal and correlated pleiotropic effects using genome-wide summary statistics.

### Authors:

J. Morrison; N. Knoblauch; J. Marcus; M. Stephens; X. He

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, University of Chicago, Chicago, Illinois.

---

Mendelian randomization (MR) is a valuable tool for detecting evidence of causal relationships between pairs of traits. Opportunities to apply MR are increasing rapidly as the number of genome-wide association studies (GWAS) with publicly available summary statistics grows. Unfortunately, existing MR methods are prone to false positives caused by pleiotropic variants. *Horizontal pleiotropy* occurs when a single variant is associated with two traits through different biological pathways. Many methods have been proposed to account for horizontal pleiotropy, which has been shown to be a widespread phenomenon. *Correlated pleiotropy* occurs when a variant affects two traits through a shared pathway/factor. If multiple variants act through the same shared factor, a subset of variants will have correlated effects on both traits. This pattern can result in false positives using most existing approaches.

We propose a new approach, Causal Analysis Using Summary Effects (CAUSE) that accounts for both types of pleiotropy using variants genome-wide. The key idea of CAUSE is that, in the presence of a causal effect, all variants associated with the causative (up-stream) trait should have correlated effects on the outcome (down-stream) trait. This is a stronger criterion than most alternative methods use and allows CAUSE to avoid false positives.

We demonstrate in simulations that CAUSE is substantially better at controlling false positive rate in the presence of correlated pleiotropic effects than other methods. CAUSE is also able to achieve better power when GWAS are under-powered. This is a result of using all variants rather than only those that are genome-wide significant.

We apply CAUSE to study relationships between pairs of complex traits and between blood cell composition and autoimmune disorders. CAUSE makes fewer discoveries than alternative methods. However, many of the pairs that CAUSE determines to be consistent with a causal effect have strong literature support. These including effects of LDL cholesterol on heart disease and of blood pressure on stroke risk. CAUSE is also able to identify an effect of blood pressure on heart disease risk that is not found using other methods. We find that many of the discoveries made by less robust methods may be explained by unmeasured shared factors rather than causal effects, suggesting that correlated pleiotropy may be a major contributor to MR false discoveries.

# PgmNr 3023: Assessment of performance of methods for polygenic Mendelian randomization analysis under realistic genetic architecture of complex traits.

## Authors:

G. Qi <sup>1</sup>; N. Chatterjee <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD

## Background

Mendelian randomization (MR) has provided major opportunities for understanding the causal relationship across phenotypes. Previous studies have often evaluated MR methods based on simulations that do not incorporate realistic models for genetic architecture of complex traits, and/or ignore the process through which instruments are selected based on genome-wide association analysis. These issues can lead to discrepant performance of MR methods in simulations and in real studies.

## Methods

We use a simulation framework that closely mirror what is empirically observed in genome-wide association studies. We simulate data on full GWAS for the exposure and outcome of interest under realistic model for effect-size distribution coherent with heritability, coheritability and polygenicity typically observed for complex traits. We select instruments based on SNPs which reach genome-wide significance in the underlying study of exposure. We then compare performance of 10 MR methods under different sample sizes of underlying GWAS, proportion of valid instruments and mechanism of pleiotropy.

## Results

Simulations show that when the InSIDE assumption holds, i.e. the instrument strength and pleiotropic effect are uncorrelated, IVW, MR-Robust, MR-RAPS, Egger regression, mode-based estimator (MBE) and MRMix have reasonably well controlled type I error rate. Among them, IVW, MR-Robust and MR-RAPS lead to the most accurate point estimates in terms of mean squared error (MSE). MRMix estimator has comparable MSE to the above three methods when the sample size is large (e.g.  $n_x > 200k$ ,  $n_y > 100k$ ), but large MSE when the sample size is smaller. The MBE and Egger regression lead to large MSE throughout the settings. Other methods, including median, MR-PRESSO, MR-Lasso and contamination mixture, have inflated type I error rate especially when  $>50\%$  of the instruments are invalid. When the InSIDE assumption is violated, we observe severe type I error inflation for most of the methods but that of MRMix and the MBE remain well controlled. MRMix leads to smaller MSE than MBE when the sample size is moderately large ( $n_x > 100k$ ,  $n_y > 50k$ ).

## Conclusions

Our simulation studies indicate different performance of MR methods in difference scenarios. Implementing multiple methods can be helpful to obtain robust conclusions.

## PgmNr 3024: Impact of fine-scale population structure in the UK Biobank on Mendelian disease variants.

### Authors:

S. Cullina<sup>1,2</sup>; G.M. Belbin<sup>1,2</sup>; R. Shemirani<sup>3</sup>; J.L. Ambite<sup>3</sup>; C.R. Gignoux<sup>4</sup>; E.E. Kenny<sup>1,2,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genomic Health, Icahn School of Medicine at Mount Sinai, New York, NY; 2) The Charles Bronfman Institute for Personalized Medicine, Mount Sinai, New York, New York.; 3) Information Sciences Institute, University of Southern California, Marina del Rey, CA, United States.; 4) Colorado Center for Personalized Medicine, Anschutz Medical Center, University of Colorado.; 5) Department of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.; 6) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

---

Massive, unselected population-based biobanks offer new opportunities to explore the segregation, penetrance and expressivity of variants thought to be underlying Mendelian forms of diseases. Here we examine Mendelian disease variants segregating in the UK Biobank (UKBB). We hypothesize that patterns of settlement will result in fluctuations in variant frequencies, particularly in historically rural and recent diaspora populations.

Of 6184 one star or above variants asserted to be unconflicted, and pathogenic or likely pathogenic in ClinVar, 1578 segregate in the UKBB. Variants were linked to one or more diseases in OMIM, and the variant-disease pairs (VDP) were binned according to mode of inheritance; Autosomal Dominant (AD) or Autosomal Recessive (AR). We extracted Easting and Northing coordinates for region of birth for UK born participants and mapped these to 167 county-level regions across Britain (ranging in size from N=35 for the Shetland Isles to N=26733 for Manchester). Variant frequencies were calculated within geocoded regions, and to account for variation in population size we report the lower bound of the 95% confidence interval of the minor allele frequency (MAF). We observe 34 AR VDPs segregating at a MAF of >1% in at least one region of the UK, and 39 AD VDPs segregated at a MAF >0.01%. We observed a significant enrichment of these VDPs (3) in the Outer Hebrides ( $p < 6.3 \times 10^{-7}$ , Fisher's Exact), a region we have previously shown has elevated cryptic relatedness in the form of Identity-By-Descent sharing (median pairwise IBD in Outer Hebrides=22cM ; versus 4.03 cM in UK-born participants overall), indicative of a founder effect in this region. Specifically, we see AR VDP for PYGM (p.R50X), PPA2 (p.R127L) and PRKN (p.R275W) each segregating at MAF of >1% in this region, associated with glycogen storage disease, sudden cardiac failure and juvenile parkinsons respectively. We are currently extending this work to >100 recent diaspora populations to the UK, and 18 known or hidden founder populations we have previously detected in the UKBB. Where possible, we will link to relevant ICD10 codes and biomarkers to examine evidence for expressivity and penetrance.

This work shines a light on population health implications of Mendelian disease variants, and a framework for generating evidence of pathogenicity, potential links to local environmental effects, and understanding populations at risk.

## PgmNr 3025: Structure informed clustering for population stratification and genetic risk prediction.

### Authors:

A. Bose <sup>1</sup>; M. Burch <sup>1</sup>; A. Chowdhury <sup>2</sup>; P. Paschou <sup>3</sup>; P. Drineas <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, Purdue University, West Lafayette, Indiana.; 2) Department of Statistics, Purdue University, West Lafayette. Indiana; 3) Department of Biological Sciences, Purdue University, West Lafayette, Indiana

---

Genome-wide association studies (GWAS) have been extensively used to estimate the signed effects of trait-associated alleles and compute polygenic risk scores. Recently, two independent studies failed to replicate the strong evidence for selection for height across Europe, implying that standard population structure correction approaches may not be enough, and that more rigorous and sophisticated methods are required. Here, we provide a correction technique for complex population structure while leveraging the linkage disequilibrium (LD) induced distances between individuals. We implement CluStrat, which performs agglomerative hierarchical clustering using the Mahalanobis distance based Genetic Relationship Matrix (GRM) representing the population-level covariance (LD) for the SNPs. With the growing size of data, computing and storing the genome wide covariance matrix is a non-trivial task. We get around this overhead by computing the GRM directly using a connection between statistical leverage scores and the Mahalanobis distance. For biobank-scale datasets, we also implement a fast algorithm to approximate all leverage scores, therefore approximating the GRM. We test CluStrat on a large simulation study of discrete and admixed, arbitrarily-structured subpopulations with allele frequencies simulated from the Balding-Nichols and Pritchard-Stephens-Donnelly models respectively for 100,000 SNPs and 5,000 individuals across 21 different scenarios. We simulated a quantitative (and its binary equivalent) trait with genetic effects at causal loci drawn from the normal distribution and varied the genetic, environmental and noise variances. CluStrat not only observed the lowest number of spurious associations for all the scenarios, but also identified two to three-fold more rare variants at causal loci as obtained by other Linear Mixed Model (LMM) or Principal Component (PC) based stratification methods. CluStrat returned similar results when applied to the Parkinson's Disease data set from WTCCC, identifying less spurious associations than LMM or PCA-based approaches. Harnessing the LD structure by fast approximation of the Mahalanobis distance is also useful in calculating the kinship matrix in LMM for heritability estimation in tera-scale datasets as well as large GWAS summary statistics. Here, we provide a comprehensive guide to stratification and subsequent disorder trait prediction or estimation utilizing the underlying LD structure of the genotypes.

## PgmNr 3026: Existence and implications of population variance structure.

### Authors:

S. Musharoff <sup>1,2</sup>; D. Park <sup>3,4</sup>; A. Dahl <sup>3</sup>; J. Galanter <sup>1,3,4</sup>; X. Liu <sup>5</sup>; S. Huntsman <sup>1</sup>; C. Eng <sup>3</sup>; E. Burchard <sup>1,3</sup>; J. Ayroles <sup>6</sup>; N. Zaitlen <sup>3,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Medicine, University of California, San Francisco, San Francisco, CA; 2) Genetics, Stanford University, Stanford, California.; 3) Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA; 4) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 5) Department of Human Genetics, University of Chicago, Chicago, IL; 6) Ecology and Evolutionary Biology and Lewis-Sigler for Integrative Genomics, Princeton University, Princeton, NJ; 7) Department of Neurology, University of California, Los Angeles, Los Angeles, CA

---

Identifying the genetic and environmental factors underlying phenotypic differences between populations is fundamental to multiple research communities. To date, studies have focused on the relationship between population and phenotypic mean. Here we consider the relationship between population and phenotypic variance, i.e., “population variance structure.” In addition to gene-gene and gene-environment interaction, we show that population variance structure is a direct consequence of natural selection. We develop the ancestry double generalized linear model (ADGLM), a statistical framework to jointly model population mean and variance effects. We apply ADGLM to several deeply phenotyped datasets and observe ancestry-variance associations with 12 of 44 tested traits in ~337K UK Biobank British individuals, 19 of 21 tested traits in a multi-ethnic sample of ~464 UK Biobank individuals, and 3 of 14 tested traits in ~3K recently admixed Mexican, Puerto Rican, and African-American individuals. We show through extensive simulations that population variance structure can both bias and reduce the power of genetic association studies, even when principal components or linear mixed models are used. ADGLM corrects this bias and improves power relative to previous methods in simulated datasets. Additionally, ADGLM identifies 4 novel genotype-variance associations across 4 phenotypes in admixed datasets and numerous such associations across 8 phenotypes in the UK Biobank British dataset. By focusing primarily on the effect of genetic variation on phenotypic mean and ignoring its effect on phenotypic variance, we have been missing an important axis contributing to phenotypic variation and disease emergence. Modeling phenotypic variance with ADGLM will enable discoveries along this axis, particularly as human studies continue to increase in their size and diversity.

## PgmNr 3027: Rare variants are disproportionately impacted by population structure in large genome-wide association studies.

### Authors:

A. Zaidi; I. Mathieson

[View Session](#) [Add to Schedule](#)

**Affiliation:** University of Pennsylvania, Philadelphia, PA.

---

Population structure leads to spurious associations in genome-wide association studies (GWAS). While many methods have been developed to detect and correct for population structure in GWAS, there are still lingering questions regarding the efficacy of these methods, especially as we move into the Biobank era, with such large sample sizes that even subtle substructure can create false positive associations. Here, we investigated the impact of population structure in the UK Biobank (UKB) study by simulating a large GWAS in a population with spatial structure that closely mimics geographic structure in the UKB.

The population structure in the UK is relatively recent, following almost complete population replacement in the last 5,000 years. We therefore assumed that demes become panmictic 100 generations ago, which is roughly the time to the last population replacement in the UK, and chose the migration rate among demes to match the  $F_{st}$  ( $\approx 0.001$ ) and genomic inflation ( $\lambda \approx 10$ ) observed in the UKB. We model phenotypes as a function of the environment—either smoothly distributed across demes (clinal effects e.g. latitude)—or affecting only a single deme (i.e. local effects). We show that when environmental effects are sharply distributed, rare variants exhibit more inflation compared to common variants, consistent with previous results in simpler models. This is because rare variants, which are more geographically structured, are more likely to be correlated with localized environmental variables.

We further show that because of the fact that demes become panmictic in the relatively recent past, the population structure is driven primarily by rare variants. Thus, principal component correction does not fully account for environmental confounding, even in common variants, and even when the environment is smoothly distributed across demes. This explains empirical observations that residual population structure continues to confound estimated effect sizes even after correcting for hundreds of PCs. We show that this effect also applies to rare variant burden tests. These results emphasize the difficulty in performing variant association tests under complex population structure, particularly for rare variants, highlighting the need for alternative methods of correction.

## PgmNr 3028: Negative selection on complex traits limits genetic risk prediction accuracy between populations.

**Authors:**

A. Durvasula; K. Lohmueller

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, UCLA, Los Angeles, California.

---

Genetic risk prediction is a key goal for medical genetics. Along these lines, great progress has been made towards identifying individuals with extreme risks across several traits and diseases using polygenic risk scores (PRS). However, many of these studies have been done in predominantly European populations, and the portability of PRS between populations is poor. The reasons for this poor portability of PRS across populations remain elusive. Here, we use population genetic simulations to show that negative selection maintains private variants at low frequency, creating population-specific genetic architectures. We find that variants present in Europe but not Africa are predicted to explain ~30% of the heritability when traits are not tied to fitness. When there is negative selection, these population-specific variants can explain 50-70% of the heritability. We show that the heritability accounted for by population-specific variants directly leads to inferior performance in risk prediction between populations, especially in the tails of the risk distribution where individuals have the most extreme phenotypes. To evaluate the impact of population-specific variants for real traits, we built a Bayesian model to stratify heritability between private and shared variants and applied it to 43 traits and diseases in the UK Biobank. Across these phenotypes, we find ~60% of the heritability is attributable to private variants, setting an upper bound on the accuracy of genetic risk predictions between populations. We conclude that the interplay between recent population growth and purifying selection contributes to the decreased accuracy of PRS in non-European populations. Consequently, genetic association studies need to include more diverse populations to enable the utility of genetic risk prediction in all populations.



## PgmNr 3029: A catalog of likely-causal, coding variant associations from imputed UK Biobank exome data.

### Authors:

A. Barton <sup>1,2</sup>; P.-R. Loh <sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics and Integrative Genomics PhD Program, Harvard Medical School, Boston, MA; 2) Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA

---

Determining the causal effects of coding variants on complex traits is challenging because most coding variants are rare. Imputation of rare, coding variants is generally limited by the size of sequenced reference panels and by imperfect matching to GWAS cohorts, while direct exome sequencing is typically only feasible in smaller cohorts. To overcome this challenge, we leveraged recently-released whole-exome sequence data for 49,960 UK Biobank participants to serve as a reference panel for imputation into the remaining ~90% of the remaining UK Biobank cohort. This approach successfully imputed variants with minor allele frequencies as low as 0.005-.01% at an accuracy of  $R^2 > 0.5$ .

To assess which of these variants might causally affect phenotypes, we performed association analyses using linear regression with principal component covariates on 34 heritable traits measured in  $N \approx 337K$  unrelated UK Biobank participants of British ancestry. Among coding variant associations that reached genome-wide significance ( $p < 5e-8$ ), we identified a subset of potentially-causal associations not explained by linkage disequilibrium with nearby SNPs. Most coding variant associations that reached significance were unlikely to be causal: our approach filtered ~90% of associated variants that were annotated as having a functional impact on a gene (missense, frameshift, etc.). Upon restricting to variants with CADD score  $\geq 20$ , most high-CADD, coding variant associations (~69%) were still filtered.

We identified 1,036 remaining associations involving coding variants with CADD  $\geq 20$  as putatively causal (across the 34 traits). To assess the extent to which our imputation approach improved our ability to detect these associations, we compared our association results to analogous results using imputation data provided by UK Biobank on the same set of samples but imputed using the Haplotype Reference Consortium (HRC) panel. We found that ~24% of variants we identified as putatively causal did not reach genome-wide significance in the UK Biobank HRC-imputed data release; most such variants were rare. Furthermore, approximately half of all the genes implicated by these variants were not previously reported in the NHGRI-EBI GWAS Catalog for the corresponding trait.

Our results demonstrate the utility of within-cohort imputation in population-scale GWAS cohorts and provide a resource of likely-causal, coding variant associations for follow-up analysis and fine-mapping method development.



## PgmNr 3030: Efficient gene-environment interaction tests for large biobank-scale sequencing studies.

### Authors:

H. Chen <sup>1,2</sup>; X. Wang <sup>3</sup>; A.K. Manning <sup>4,5</sup>; Y.J. Sung <sup>6</sup>; D.C. Rao <sup>6</sup>; A.C. Morrison <sup>1</sup>; E. Boerwinkle <sup>1,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics Center, Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 2) Center for Precision Health, School of Public Health and School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX; 3) Department of Biostatistics and Data Science, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX; 4) Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA; 5) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 6) Division of Biostatistics, Washington University School of Medicine, St. Louis, MO; 7) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX

---

Complex human diseases are affected by both genetic and environmental risk factors. Gene-environment interaction (GEI) tests for aggregate genetic variant sets have been developed in recent years. With the advance of sequencing technology, whole exome and genome sequencing data are now being collected at an unprecedented scale in biobanks. However, existing statistical tests become computationally intensive in such large-scale sequencing studies with correlated samples. We propose efficient Mixed-model Association tests for GEne-Environment interactions (MAGEE), for testing the GEI between an aggregate genetic variant set and environmental exposures on quantitative and binary traits in large-scale sequencing studies. Joint tests for genetic main effects and GEI are also developed. A null generalized linear mixed model adjusting for covariates but without any genetic effects is fitted only once in a whole genome GEI analysis, thereby vastly reducing the overall computational burden. Score tests for variant sets are performed as a combination of genetic burden and variance component tests by accounting for the genetic main effects using matrix projections. The computational complexity is dramatically reduced in a whole genome GEI analysis, which makes MAGEE scalable to hundreds of thousands of individuals. Furthermore, MAGEE allows flexible weighting schemes to incorporate functional genomic information. We applied MAGEE to the exome sequencing data of 49,959 related individuals from the UK Biobank, and the analysis of 19,442 protein coding genes finished within a few hours on a single computing node.

## PgmNr 3031: Carriers-only tests of association of rare genetic variants with a binary outcome.

### Authors:

T. Sofer <sup>1,2</sup>; J. Lee <sup>2</sup>; E. Schifano <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine, Harvard Medical School, Boston, Massachusetts.; 2) Sleep Medicine, Brigham and Women's Hospital, Boston, Massachusetts; 3) University of Connecticut, Storrs, Connecticut

---

**Background:** Rare variants association analyses with binary outcomes that use likelihood-based tests such as the Score and Wald tests, suffer from inflated type 1 error. The problem is exacerbated when pooling together individual-level data from diverse parent studies, with potentially different proportions of affected individuals and different frequencies of genetic variants. For example, the analysis of short sleep (< 6 hours/night) considers ~22,000 individuals from TOPMed of European, African, Hispanic, and Asian descent, and the proportion of short sleepers is highest among African Americans. Here, the Score test detected many spurious associations

**Methods:** We propose a class of semi-parametric tests for association of rare variants with binary traits. Given a vector of probabilities of affection status among a set of (possibly correlated) carriers of a rare variants, we test whether the number of affected individuals among these carriers is consistent with their probability vector. Our carriers-only tests use the Poisson-Binomial (BinomiRare test) or Conway-Maxwell-Poisson (CMP test) distributions.

**Results:** We performed genetic analysis of short sleep in TOPMed. The Score test did not control the type 1 error. The SAIGE test improved upon the Score test, but had also high inflation when there were less than 30 carriers of the rare variant, and was slower than the carriers-only approach. When there are at least 30 carriers, the SAIGE and carriers-only tests had similar performance (correlation between p-value = 0.96), with SAIGE being often slightly more powerful. Still, in the single region that passed the genome-wide significance threshold, the most significant associations were detected by BinomiRare. A variant chr2:33765596:A:G (chr2 p22.3; 65 carriers; CADD score=15.3) had BinomiRare p-value  $7.4 \times 10^{-9}$ , SAIGE p-value =  $2.6 \times 10^{-8}$ .

**Conclusions:** carriers-only tests offer a computationally efficient alternative to other tests of rare variant associations in the mixed model settings, in terms of both time and memory. They protect the type 1 error for arbitrary number of carriers of the rare variant and for arbitrary case-control proportion. These tests can be applied on single variant and on a burden of multiple variants.

## PgmNr 3032: Polygenic risk modeling with latent trait-related genetic components.

### Authors:

M. Aguirre <sup>1,2</sup>; Y. Tanigawa <sup>1</sup>; G. Venkataraman <sup>1</sup>; J.M. Justesen <sup>1,3</sup>; R. Tibshirani <sup>4</sup>; T. Hastie <sup>4</sup>; M.A. Rivas <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Data Science, Stanford University, Stanford, California.; 2) Department of Pediatrics, School of Medicine, Stanford University, Stanford, CA; 3) Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen; 4) Department of Statistics, Stanford University, Stanford, CA

---

Polygenic risk models have led to significant advances in understanding complex diseases and their clinical presentation. While simple linear models of genetic risk like polygenic risk scores (PRS) can be quite accurate, they do not generally account for disease subtypes or pathways which underlie within-trait diversity. Meanwhile, etiology- and pleiotropy-aware models of genetic effects offer the opportunity to more deeply analyze the underpinnings of complex traits within individuals and across populations. Here, we introduce a latent factor model of genetic risk based on components from matrix DEcomposition of Genetic ASSociations (DEGAS), and compute DEGAS risk scores for high body mass index, diabetes, and renal failure in a cohort of array-genotyped individuals from UK Biobank. We assess the predictive performance of DEGAS relative to traditional PRS, and validate its characterization of genetic risk profiles for each trait.

## PgmNr 3033: Methodology for differential gene expression analysis in large samples: Application to cystic fibrosis lung disease.

### Authors:

L. Zhong<sup>1,2</sup>; L.J. Strug<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

---

Read counts from RNA-seq experiments are a measure of gene expression. After quality control and normalization procedures, the analysis goal is to identify truly differentially expressed genes in the presence of additional biological or technical variation that remains in the count data. While overdispersion is most common, some genes have read counts that are under dispersed. Existing statistical methods model the mean-variance relationship by either assuming a negative binomial distribution or by transforming the counts to a continuous scale, but the tools to estimate dispersion as well as outliers can bias the results. As sequencing technology decreases in price, RNA-seq experiments using larger sample sizes is becoming more accessible, and thus so are more robust tools that rely on large sample statistical properties. We propose a novel method for differential gene expression analysis using robust profile likelihoods, RPLR. Poisson log-linear regression is used as the working model, and after a data-driven robust adjustment, the profile likelihood ratio displays accurate inference on the fold change of gene expression under binary or quantitative conditions of interest. We simulated count data genome-wide from a negative binomial, generalized poisson and rounded gamma distribution and showed the RPLR performs at least as well as edgeR (Robinson et al. 2010), DESeq2 (Love et al. 2014) and limma-voom (Law et al. 2014) for a binary condition of 2 fold change at sample sizes as low as 25 per group, resulting in a sensitivity of 0.8 controlled at a 10% nominal FDR. By sample sizes of 40 per experimental group, RPLR had the highest AUC, and performed uniformly better under under-dispersion, presence of outliers and for different mean-variance relationships. In RNA-Seq of nasal epithelia from 80 individuals with Cystic Fibrosis (CF) the RPLR method identified differentially expressed genes associated with CF lung disease. These genes included those identified by existing methods (eg *MFHAS1*, *CYP2T3P* and *THG1L*) and novel findings such as *UBE2G1* (FDR of 0.03) whose 6th exon has complete sequence homology with a lncRNA at the known *SLC26A9* CF modifier locus. RPLR provides a flexible approach to differential gene expression analysis that is designed for larger sized studies. It can accommodate any unknown mean-variance relationship, is robust to outliers, and is applicable to categorical or quantitative variables of interest and complex experimental designs.

## PgmNr 3034: TWO-SIGMA-geneset: TWO-component SInGle cell model-based association method for gene set testing.

### Authors:

E. Van Buren <sup>1</sup>; M. Hu <sup>3</sup>; B. Babcock <sup>2</sup>; K. Wilhelmsen <sup>2</sup>; D. Wu <sup>1,4</sup>; Y. Li <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 3) Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation; 4) Department of Oral and Craniofacial Health Sciences, Adams School of Dentistry, The University of North Carolina at Chapel Hill

---

Gene set testing (GST) methods allow omnibus testing for a pre-specified set of genes, often by assigning an overall  $p$ -value to the set of genes. By considering such sets of genes, GST methods can improve statistical power and help to understand how genes in a common pathway jointly regulate biological processes. There are many GST methods for bulk RNA-seq data, however no such methods are specifically designed for use with single-cell RNA-seq data, which often exhibits excess zeros and overdispersed counts compared to traditional bulk RNA-seq data. Here, we propose TWO-SIGMA-geneset to conduct gene set testing using single-cell RNA-seq data. We utilize our previously developed TWO-SIGMA, a flexible TWO-component SInGle cell Model-based Association method. The first component of TWO-SIGMA models the “drop-out” probability with a mixed effects logistic regression, and the second component models the (conditional) mean read count with a log-linear negative binomial mixed-effects regression. TWO-SIGMA accommodates excess zeros and overdispersed counts, and can, using random effects, explicitly account for the possible dependency introduced by measuring multiple single cells from the same sample. TWO-SIGMA-geneset focuses on conducting competitive gene set testing, in which the genes in a given set are compared to the remaining collection of genes. Each gene in the set is assigned an individual statistic, such as a likelihood ratio statistic from the joint test of both components, produced from TWO-SIGMA. These gene-level statistics are then combined into one set-level statistic, with options including the average, median, or sum of the ranks. Finally, the set-level statistic is compared to the complement set of genes using, for example, the t-test, Wilcoxon rank-sum test, or a Monte Carlo  $p$ -value. Simulation studies show that type-I error is well controlled in a variety of representative scenarios. Power is improved over state-of-the-art methods, including CAMERA, for a variety of scenarios consistent with real single-cell RNA-seq data, including when varying: gene set sizes, the scales of the differential expression, the proportion of drop-out events, or the presence of individual-level random effects. Extensions to allow for inter-gene correlation are explored. We demonstrate our method with an application to a real dataset consisting of 36,849 single cells from 10 mice and gene sets from the Molecular Signatures Database.

# PgmNr 3035: Robust filtering of cell-free droplets from single-cell RNA sequencing.

## Authors:

J. Xi; H. Kang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI

---

The advances of single-cell RNA sequencing (scRNA-seq) technologies and analytic methods enable massively parallel profiling of single cell transcriptomes of thousands of cells using droplet-barcoding microfluidic devices. Due to the stochastic nature of the technology, it is expected that a large fraction of droplets does not contain cells. Ideally, the barcodes of these cell-free droplets should not present in the sequenced scRNA-seq library. However, in actual scRNA-seq data, the microparticles in cell-free droplets occasionally react with transcripts after the droplets are broken for library preparation. This results in the sequencing of cell-free droplets with typically low UMI read counts. As a result, the number of barcoded droplets is orders of magnitudes larger than the droplets representing individual cells. A very first step to analyze droplet-based scRNA-seq data is to filter out cell-free droplets to focus on droplets representing actual cells. However, typical procedures to determine cell-free droplets is very naive, for example, relying only on the total number of UMIs per droplet. Such a heuristic procedure may result in serious bias and systematic underrepresentation of cell types with lower UMI counts. We propose a robust method to distinguish cell-free droplets from cell-containing ones in scRNA-seq datasets by integrating transcriptomic profiles, genetic information, and the UMI distribution together. We use Dirichlet multinomial distribution to model the likelihood of cell-free droplet and combine them with genetic variants (if multiplexed) to determine whether the droplet is likely cell-free or not. We apply our method to various scRNA-seq datasets generated by DropSeq and 10x Chromium technologies and show that our method rescues more droplets that are likely to represent individual cells compared to existing methods. For example, among 247,027 droplet barcodes obtained from a DropSeq experiment across a mixture of 3 cancer lines, we observed that Cell Ranger's UMI-count method and EmptyDrops software filtered 938 and 2,407 as valid barcodes, while our proposed method filtered 3~7-fold more (7,179) barcodes. False negative rates based on carefully chosen high-quality droplets were 54% for Cell Ranger, 24% for EmptyDrops, and 2.4% for our method. Fraction of apparent doublets among filtered droplets (which is enriched for cell-free droplets) was estimated to be 8.4% for Cell Ranger, 18.8% for EmptyDrops, and 11.2% for our method.

## PgmNr 3036: Japanese specific imputation reference panel using 7,000 whole genome sequences.

### Authors:

J. Flanagan<sup>1,2</sup>; N. Matoba<sup>3</sup>; M. Akiyama<sup>3,4</sup>; K. Ishigaki<sup>3</sup>; Y. Momozawa<sup>5</sup>; K. Ito<sup>6</sup>; K. Matsuda<sup>7,8</sup>; Y. Murakami<sup>9</sup>; Y. Kamatani<sup>3,10</sup>; A.P. Morris<sup>2,11</sup>; M. Horikoshi<sup>1</sup>; C. Terao<sup>3,12,13</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory for Endocrinology, Metabolism and Kidney diseases, RIKEN, Center for Integrative Medical Sciences, Yokohama, Japan; 2) Department of Biostatistics, University of Liverpool, Liverpool, UK; 3) Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 5) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 6) Laboratory for Cardiovascular Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 7) Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 8) Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan; 9) Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; 10) Kyoto-McGill International Collaborative School in Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 11) School of Biological Sciences, University of Manchester, Manchester, UK; 12) Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan; 13) The Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan

---

The inclusion of population-specific whole genome sequences (WGS) into established reference panels has been demonstrated to improve imputation quality at minor allele frequency (MAF) thresholds of <5%. While population-specific WGS at deep depth is expected to improve imputation both quantitatively and qualitatively, the impact of combining WGS at different depth (deep and low depth vs deep depth) remains unexplored.

We created three Japanese population specific reference panels using combinations of WGS from the BioBank Japan Project to augment the 1000 Genomes Project (1KG): 1000 Genomes Project augmented with 1000 Japanese WGS (1KG+1K) at 30x depth; 1000 Genomes Project augmented with 3000 Japanese WGS (1KG+3K) at 30x depth; and 1000 Genomes Project augmented with 7000 Japanese WGS (1KG+7K) of mixed depth (3000 at 30x and 4000 at 3x depth).

A cohort of 30,042 Japanese individuals genotyped on the Illumina OmniExpress and HumanExome chips was prephased with EAGLE2 and imputed up to each of the reference panels using minimac3. Comparisons of imputation quality (based on estimated  $R^2$  provided by minimac3) were conducted on a subset of imputed variants shared across the three reference panels and 1KG alone.

Analyses of autosomes demonstrated that the addition of WGS data substantially improved imputation quality, particularly for SNPs with MAF <5%. A total of 42,380,426 SNPs were shared across reference panels. Of these, after imputing up to the 1KG panel alone, 10,850,678 (25.6%) SNPs were 'well imputed' ( $R^2 > 0.3$ ), including 3,172,237 (9.3%) SNPs with MAF <0.5%. After imputing up to the 1KG+3K panel, the number of well imputed SNPs increased to 16,452,081 (38.8%). We also

observed a dramatic increase in the number of well imputed SNPs with MAF <0.5% to 8,499,108 (24.8%). The highest quality imputation was derived from imputing up to the 1KG+7K panel: 18,065,768 (42.6%) of SNPs were well imputed, including 10,114,995 (29.5%) with MAF <0.5%. The mean  $R^2$  of SNPs with MAF <0.5% increased from 0.094 after imputation up to the 1KG panel alone to 0.203 after imputation up to the 1KG+7K panel. The addition of 7000 Japanese WGS of mixed depth to the 1KG panel resulted in a much-improved imputation into Japanese individuals in terms of both the number of variants and their imputation quality, and would be expected to increase power for discovery of rare variant associations with complex human disease.



## PgmNr 3037: Equal Local Levels (ELL): A test statistic for trans eQTL detection.

### Authors:

T. Akinbiyi <sup>1</sup>; N. Gonzales <sup>2</sup>; A. Palmer <sup>3</sup>; M.S. McPeck <sup>1</sup>; M. Abney <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Statistics Dept, University of Chicago, Chicago, IL.; 2) Human Genetics Dept, University of Chicago, Chicago, IL; 3) Palmer Lab, Biomedical Research Facility II (BRF2), La Jolla, CA

In trans eQTL mapping it is common to have to test the association between a gene expression level and a genetic variant simultaneously for 1000s of genes and 100,000s of genetic variants. Properly maintaining type 1 error without sacrificing power becomes challenging. One strategy is, for a single SNP, to aggregate all tests of that SNP against each gene and test whether it is associated with - and thus an eQTL for - at least one gene, repeated over all SNPs in the genome. Global Type 1 error correction then adjusts for only the number of genetic variants not the product of that and the number of genes. Once SNPs that are eQTLs for at least one gene are found, further detailed analysis of those specific SNPs can be performed

We propose the Equal Local Levels (ELL) test statistic which aggregates test statistics -- Z-scores -- from the tests of association between a SNP and a set of genes to test if at least one is not null. ELL maintains high power and correct Type 1 error in the realistic setting of correlated gene expressions. In simulations, when a SNP is a trans eQTL for 10 to 80 of 10,000 tested genes, ELL has high power and outperforms other strategies (Generalized Higher Criticism, Generalized Berk-Jones, Score test, minimum p-value, FDR, CPMA) with, in many cases, relative power increases of 15 - 30% over the next best strategy while maintaining correct Type 1 error. We consider several scenarios of correlation between Z-scores for different genes. ELL is also computationally feasible such that evaluation against 100,000s of genetic variants can be done quite quickly.

We did trans eQTL mapping in the LG/J x SM/J advanced intercross line of mice (LG x SM AIL), which is a multi-generational outbred population dataset consisting of 523,028 SNPs and expression levels of 16,661 genes recorded for the hippocampus, pre-frontal cortex and striatum for 208 mice. We show that when applied to this data, ELL is fast -- it can be calculated for all 523,028 genetic variants in under an hour -- and powerful -- we are able to discover numerous trans eQTLs.

Our ELL statistic is a novel method that has the potential to significantly improve the efficacy of trans eQTL mapping studies.

## PgmNr 3038: Modeling imputation uncertainty in fine-mapping improves accuracy.

**Authors:**

C. Albers <sup>1</sup>; Z. Fan <sup>2</sup>; H. Finucane <sup>1</sup>

[View Session](#)[Add to Schedule](#)**Affiliations:**

1) Broad Institute, Cambridge, MA; 2) Department of Statistics and Data Science, Yale University, New Haven, CT

---

Identifying causal variants is a fundamental problem in analysis of genome-wide association study data. Existing computational methods for fine-mapping have led to important insights, but there are no scalable methods that explicitly model uncertainty due to genotype imputation.

Here, we first introduce a new simulation framework that allows us to assess the effects of ignoring imputation uncertainty and to evaluate potential solutions to this problem. To do this, we simulate human sequence data at large sample sizes using a coalescent model, and we use these sequences to simulate heritable phenotypes. We then mask variants to simulate the process of genotyping, and we re-impute the masked variants. Finally, we use our imputed genotypes and simulated phenotypes to perform fine-mapping. Our results show that ignoring imputation uncertainty can lead to high-confidence false positives.

Second, we introduce a new method for fine-mapping with imputed genotypes. Our method expands on the state-of-the-art 'Sum of Single Effects' model (Wang et al 2018). It uses the imputation as a prior for genotype values, which are treated as latent variables, to incorporate imputation uncertainty into the posterior inference. We use mean-field variational inference to approximate the joint posterior of genotypes and genetic effects. We demonstrate through simulations that explicitly modeling imputation uncertainty improves the accuracy of fine-mapping.

## PgmNr 3039: Revealing the complex genetic architecture of type 1 diabetes complications: Joint modeling of multivariate longitudinal and time-to-event traits.

### Authors:

M. Brossard <sup>1</sup>; A.D. Paterson <sup>2,3</sup>; O. Espin-Garcia <sup>1,3</sup>; R.V. Craiu <sup>4</sup>; S.B. Bull <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Lunenfeld-Tanenbaum Research Institute, Sinai Health System, Toronto, ON, Canada; 2) Hospital for Sick Children Research Institute, Toronto, ON, Canada; 3) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 4) Statistical Sciences, University of Toronto, Toronto, ON, Canada

---

Motivated by the complex genetic architecture of Type 1 diabetes complications (T1DC), where multiple SNPs are associated directly &/or indirectly with time to complications via intermediate longitudinal risk factors, we develop a joint model (JM) of multiple longitudinal & time-to-event traits. JM links (a) a multivariate linear mixed model for the observed longitudinal measures of each trait, accounting for measurement error, including a time trajectory, fixed SNP & random subject effects; and (b) a frailty survival model for time-to-event traits depending on the fitted longitudinal trajectories from (a), fixed SNP & random subject effects. The *overall* SNP association with a time-to-event trait can be decomposed into *direct* ( $\gamma$ ) & *indirect* ( $\beta$ ) effects on each longitudinal process. We classify a SNP association with a time-to-event trait as: *direct* if  $\gamma \neq 0$  &  $\beta = 0$ ; *indirect* if  $\gamma = 0$  &  $\beta \neq 0$ ; *direct & indirect* if  $\gamma \neq 0$  &  $\beta \neq 0$  and propose a multi-trait SNP test for global SNP association with at least one trait. Under a system model we formulate for the complex architecture of a T1DC study, we simulate datasets with phenotypes that involve 5 causal SNPs with various direct effects on two time-to-event traits (retinopathy, nephropathy) &/or indirect effects via two observed (glucose, blood pressure) and one simulated longitudinal traits. For each SNP, we fit (a) & (b) sequentially and estimate the covariance matrix via bootstrap. Compared to standard approaches (marginal (MM) and frailty survival models (FM) adjusted for observed longitudinal values instead of trajectories in (b)), SNP effect estimates from JM are more efficient, especially for SNPs associated with time-to-event outcomes with lower event rates, and JM detects direct or indirect SNP associations with highest power, even when the model is not fully specified. For direct & indirect SNP associations, MM & FM overestimate the direct SNP effect, due to their inability to separate correctly the direct & indirect effects from the overall effect when longitudinal traits are ignored from the model (for MM) or have measurement error (for FM). The proposed multi-trait SNP test can be more powerful for discovery of SNPs associated with multiple traits and has slightly inferior power than single parameter tests for other types of SNP associations. Application of JM methods provides improved SNP effect estimates & hypothesis tests that can reveal complex architecture of related traits.

## PgmNr 3040: Sparse empirical kinship matrices enable computationally efficient and accurate association tests in large samples.

### Authors:

M. Conomos<sup>1</sup>; T. Zhang<sup>1</sup>; S.M. Gogarten<sup>1</sup>; D. Jain<sup>1</sup>; C.P. McHugh<sup>1</sup>; Y. Hu<sup>2</sup>; A.P. Reiner<sup>3</sup>; K.M. Rice<sup>1</sup>; on behalf of NHLBI Trans-Omics for Precision Medicine (TOPMed) Analysis & Hematology Working Groups

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Washington, Seattle, WA.; 2) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Department of Epidemiology, University of Washington, Seattle, WA

Mixed models for genetic association testing have traditionally accounted for structure among samples by using an empirical genetic relationship matrix (GRM) that measures genetic covariance, genome-wide, from both ancestry and relatedness. However, fitting mixed models in samples with tens or hundreds of thousands of individuals can be a prohibitive computational burden. Here, we address this problem by using a sparse empirical kinship matrix (KM) and ancestry principal components in place of a GRM.

Standard forms of empirical GRMs and KMs estimated from genotype data are dense; i.e. have no entries equal to zero. To exploit the computational speedups that sparse matrices enable, we make an empirical KM sparse by clustering samples based on their pairwise kinship estimates, setting all inter-cluster estimates to zero; this can also be thought of as approximating low levels of relatedness as 'unrelated'. In today's large-scale population studies, where those in pedigrees are a small proportion of the overall sample, this approximation can be expected to be highly accurate, and the computational speedup substantial.

To illustrate the computational advantage and statistical impact of using sparse empirical KMs, we performed genetic association analyses using seven red blood cell traits and WGS data from TOPMed freeze 6. Between 17,469 and 48,858 samples were available for these traits. Using a 4th degree relatedness threshold (i.e. kinship > 0.022) and our proposed algorithm, 98.3% to 99.5% of entries in the sparse KM were set to zero, and the largest cluster ranged from 1667 to 2459 samples. Compared to using a GRM, using a sparse KM significantly improved computational performance; e.g. fitting the null models for these traits took just 0.5-6.2% of the CPU time and required 1.4-6.7% of the memory. Furthermore, differences in association p-values between the two approaches were small. For these traits, over 99.99% of tests differed in  $-\log_{10}(p)$  by less than 0.5; i.e. by an amount very unlikely to change the practical interpretation of results. With the level of sparsity attainable in population studies such as TOPMed, we also find that our approach performs favorably compared to SAIGE, another mixed model method designed for analysis of large samples. The use of sparse KMs is a promising and flexible approach to improve the computational efficiency of association testing in large population studies, without sacrificing accuracy.

## PgmNr 3041: Prediction-based functionally informed priors for fine-mapping.

### Authors:

R. Cui<sup>1</sup>; O. Weissbrod<sup>2</sup>; M. Kanai<sup>1,3,4</sup>; J. Ulirsch<sup>1,5</sup>; H. Finucane<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Inst, Cambridge, Massachusetts.; 2) Harvard School of Public Health, Epidemiology Department, Boston, MA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; 4) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA 02115, USA; 5) Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, USA

---

Using functional annotations to inform prior causal probabilities for fine-mapping can increase power to identify causal variants. However, taking advantage of a large number of functional annotations with existing methods requires very strong modeling assumptions. Here we introduce a new procedure for constructing Bayesian priors for fine-mapping from hundreds to thousands of functional annotations that is more flexible and potentially more robust to model misspecification than previous methods. Our procedure has two steps: first, we learn a function of the annotations that allows us to rank SNPs; second, we partition the SNPs to equally sized bins based on this ranking, and we compute one value per bin that we use as the prior probability of causality for SNPs in that bin.

To rank the SNPs, we use a prediction framework. First, we compute posterior inclusion probabilities (PIPs) by running a state-of-the-art fine-mapping algorithm such as SuSiE<sup>[1]</sup> with a uniform prior. Second, we select annotations to include in our procedure by marginally evaluating each one using stratified LD score regression<sup>[2]</sup> which leverages polygenic signal, and keeping those that are significantly associated with the trait in question. We then train a predictor of the PIPs using the selected annotations.

In a preliminary analysis of UK Biobank Height data, using baselineLDv2.2 annotations, we compared five methods for prediction in a leave-one-chromosome out framework: Ordinary Least Squares (OLS), Decision Tree (DT), Gradient Boosted Tree (GBT), Random Forest (RF), Logistic Regression (LR). To assess prediction accuracy, we first computed the mean target PIPs in each quintile of predicted PIP, and we then took a ratio of mean target PIP in the top and bottom quintile. For OLS, GBT, and LR we see ratios of 22.05, 22.48, and 17.68 respectively; DT and RF had ratios of 1.78 and 3.31 respectively. Using only top loci to run the model fitting step decreased this ratio slightly (OLS: 18.26).

To compute priors for fine-mapping that are robust to model misspecification, we rank the SNPs using a trained predictor, we partition the SNPs into equally sized bins, and we use the average target PIP within the bin as the prior for the SNPs on the fine-mapping chromosome. We use separate chromosomes for training the predictor, binning, and fine-mapping.

[1] Wang et. al. 2018 bioRxiv

[2] Finucane et. al. 2015 Nature Genetics

## PgmNr 3042: Polygenic prediction via Bayesian regression and continuous shrinkage priors.

### Authors:

T. Ge<sup>1,2</sup>; C.-Y. Chen<sup>1,2</sup>; Y. Ni<sup>3</sup>; Y.-C.A. Feng<sup>1,2</sup>; J.W. Smoller<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Statistics, Texas A&M University, College Station, TX

---

Polygenic risk scores (PRS), which summarize the effects of genome-wide genetic markers to measure the genetic liability to a trait or a disorder, have shown promise in predicting human complex traits and diseases, and may facilitate early detection, risk stratification, and prevention of common complex diseases in healthcare settings. Conventional methods for PRS construction, which rely on LD-pruning and P-value thresholding (P+T), discard information and limit prediction accuracy. More sophisticated Bayesian polygenic prediction methods (such as LDpred) often model SNP effect sizes as a discrete mixture of normal distributions, which imposes daunting computational challenges and may result in inaccurate adjustment for LD structure. We have developed PRS-CS, a polygenic prediction method that utilizes a high-dimensional Bayesian regression framework and places a conceptually novel class of priors --- continuous shrinkage (CS) priors --- on SNP effect sizes, which can accommodate varying genetic architectures, enable accurate and multivariate modeling of local LD patterns, and provide substantial computational advantages over discrete mixture priors. In addition, with our pre-calculated, population-specific LD references, PRS-CS infers posterior SNP effect sizes from GWAS summary statistics and does not require individual-level genotype data. We conducted extensive simulations across a wide range of genetic architectures, GWAS sample sizes and SNP heritability, using data from the UK Biobank. PRS-CS robustly improved the predictive performance of PRS over existing methods. We applied PRS-CS to predict six common complex diseases (breast cancer, coronary artery disease, depression, inflammatory bowel disease, rheumatoid arthritis, and type 2 diabetes mellitus) and six quantitative traits (height, body mass index, high-density lipoproteins, low-density lipoproteins, cholesterol, and triglycerides) in the Partners HealthCare Biobank. PRS-CS consistently outperformed all alternative methods examined, with an average improvement of 14% in prediction accuracy relative to LDpred and an average improvement of 55% relative to P+T. We envision that PRS-CS will demonstrate bigger advantages as the GWAS sample sizes continue to grow. We will further present new results on trans-ethnic polygenic modeling and prediction using GWAS summary statistics and LD reference panels from multiple populations, based on an extension of the modeling framework of PRS-CS.

## PgmNr 3043: Causal inference for highly pleiotropic biomarkers using Mendelian randomisation and Bayesian networks.

**Authors:**

R. Howey; H.J. Cordell

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom

---

Mendelian randomization (MR) is a popular tool for performing causal inference in genetic epidemiology, but it can have limitations for evaluating simultaneous causal relationships in complex data sets. We consider a scenario that may arise when analysing biomarker data as generated from modern “omics” technologies, whereby multiple genetic variants are associated with multiple potential intermediate variables (biomarkers) that mediate an outcome of interest. This highly pleiotropic scenario violates one of the required assumptions of MR. However, Bayesian network analysis (BN) offers an alternative approach. In BN, the conditional dependencies and independencies of variables are described by a graphical model (a directed acyclic graph) and its accompanying joint probability, which may be estimated using individual-level data.

We perform computer simulations to investigate the utility of BN in this situation. We simulate data using previously estimated real effect sizes of 150 genetic variants on 12 biomarkers which are assigned to have either an effect on the outcome, no effect or a reverse effect from the outcome to the biomarker. As well as BN and MR we evaluate several other recently-proposed causal inference methods: multivariable MR based on Bayesian model averaging (MR-BMA), a multi-SNP mediation intersection-union test (SMUT) and a latent causal variable (LCV) test. Our results showed that BN outperformed all other methods in terms of high power to detect an effect in the correct direction, while maintaining low type I error. As expected, MR had high power but also very high type I error due to pleiotropy. MR-BMA performed well at identifying which biomarkers were causal and had excellent type I error, but it is not designed to detect reverse effects.

We conclude that BN is a useful complementary approach to existing methods for performing causal inference in complex data sets such as those generated from modern “omics” technologies.

## PgmNr 3044: A resource-efficient tool for mixed model association analysis of large-scale data.

### Authors:

L. Jiang<sup>1</sup>; Z. Zheng<sup>1,2</sup>; T. Qi<sup>1</sup>; K. Kemper<sup>1</sup>; N. Wray<sup>1,3</sup>; P. Visscher<sup>1,3</sup>; J. Yang<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia; 2) Institute for Advanced Research, Wenzhou Medical University, Wenzhou, Zhejiang 325027, China; 3) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia

---

The genome-wide association study (GWAS) has been widely used as an experimental design to detect associations between genetic variants and a phenotype. Two major confounding factors, population stratification and relatedness, could potentially lead to inflated GWAS test-statistics and thereby spurious associations. Mixed linear model (MLM)-based approaches can be used to account for sample structure. However, genome-wide association (GWA) analyses in biobank samples such as the UK Biobank (UKB) often exceed the capability of most existing MLM-based tools especially if the number of traits is large. Here, we developed an MLM-based tool (called fastGWA) that controls for population stratification by principal components and relatedness by a sparse genetic relationship matrix for GWA analyses of biobank-scale data. We demonstrated by extensive simulations that fastGWA is reliable, robust and highly resource-efficient. We then applied fastGWA to 3,613 traits on 456,422 array-genotyped and imputed individuals and 2,090 traits on 46,191 whole-exome-sequenced (WES) individuals in the UKB.



## PgmNr 3045: Fine-scale estimates of regional polygenicity provide insights into the genetic architecture of complex traits.

### Authors:

R. Johnson <sup>1</sup>; K. Burch <sup>2</sup>; K. Hou <sup>1,3</sup>; B. Pasaniuc <sup>2,4,5,6</sup>; S. Sankararaman <sup>1,2,4,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Computer Science, University of California, Los Angeles, Los Angeles, CA, USA; 2) Bioinformatics Interdepartmental Program, University of California, Los Angeles, Los Angeles, CA, USA; 3) College of Computer Science and Technology, Zhejiang University, Hangzhou, Zhejiang, China; 4) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; 5) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA; 6) Department of Computational Medicine, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

---

Investigating how many genetic variants contribute to phenotypes and how they are distributed throughout the genome is a fundamental question in quantitative genetics. While previous studies have used partitioned heritability within regions of the genome to gain insights into genetic architecture, estimates of heritability alone do not provide information about the number of SNPs with nonzero effect sizes (causal SNPs).

We propose an approach to infer the proportion of causal SNPs and the expected per-causal SNP effect size at each region. Our approach, BEAVR (Bayesian EstimAtion of Variants in Regions), estimates the proportion of causal SNPs in a region from GWAS data while accounting for linkage disequilibrium. We introduce a new inference algorithm that leverages the intuition that most SNPs are not causal to perform efficient inference that scales linearly with the number of SNPs.

We divide the genome into 470 non-overlapping 6MB regions (~1K SNP per region) and apply our method to estimate the regional proportion of causal SNPs across each region for 22 complex traits from the UK Biobank. We find that each 6MB region contains an average of 19 (SD: 16.8) causal SNPs aggregated across all traits. First, we show that heritability is proportional to the number of causal SNPs both at the genome-wide scale and within most regions. On average across all regions and traits, each causal SNP contributes 3.9E-05 of heritability. Second, we identify 42 regions that deviate from this pattern, where the increased heritability is attributed to increased effect sizes as opposed to a higher number of causal SNPs. To understand the factors that affect the distribution of causal SNPs, we regressed the number of causal SNPs against the number of genes per region. We observe a linear relationship between the number of causal SNPs and the number of genes for most traits ( $r^2 = 0.02$  across 13 traits,  $P = 1.0E-15$ ). After controlling for heritability and the total number of SNPs in a region, this relationship remains significant for height, BMI, and asthma ( $P = 1.6E-03$ ;  $4.2E-02$ ;  $1.8E-02$ ). Finally, we model the genome-wide effect size distribution by aggregating our estimates of regional heritability and polygenicity to define a distribution parameterized by a mixture of variance components. Compared to effect size distributions described only by genome-wide parameters, this

distribution described by regional parameters contains 3.6x more SNPs within the tails of the distribution.

## PgmNr 3046: Partitioning genetic correlation by annotation using LD score regression.

### Authors:

D. Kassler <sup>1,2,3</sup>; Y. Reshef <sup>5</sup>; A. Price <sup>2,3,4</sup>; H. Finucane <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biomedical Informatics, Harvard University, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 4) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA; 5) Harvard Medical School, Boston, MA

---

LD score regression (LDSC) is a method for detecting confounding and estimating genetic correlation and components of heritability from GWAS summary statistics. Here we extend LDSC to estimate genetic correlation within annotations and analyze the properties of this method.

Using genotype data from the UK Biobank, we simulate phenotypes and summary statistics for genetically correlated traits with annotation-specific enrichment of genetic covariances. By simulating under a range of parameter values for overall trait heritability and enrichment of heritability and genetic covariance within annotations (minor allele frequency bin, coding, conserved, and DHS), we describe the conditions under which LDSC produces reliable and numerically stable estimates of annotation-level genetic correlation. For traits with low annotation-specific heritability, it is often impossible to estimate genetic correlation within an annotation due to numerical instability of heritability estimates within some LD blocks, even for traits for which genome-wide genetic correlation can be estimated. For example, we simulated pairs of correlated traits with and without enrichment of correlation in a MAF 1%-5% annotation, with genome-wide genetic correlation 0.6, heritability 0.5, and no enrichment of heritability in the annotation, at a sample size of 50K. We were able to estimate the genome-wide genetic correlation using LDSC in 98% of these simulations, but could not estimate the presence or absence of enrichment of genetic correlation within the annotation in 19.1% of null simulations (no enrichment) and 25.0% of non-null simulations (annotation-specific correlation 0.9).

For certain pairs of traits which permit estimates, we also present preliminary results from real data.

## PgmNr 3047: Is the first-order variance approximation in Mendelian randomization OK?

### Authors:

H. Kim <sup>1</sup>; K. Kim <sup>2</sup>; B. Han <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Sciences, Seoul National University, Seoul, South Korea; 2) Department of Medical Sciences, University of Ulsan, College of Medicine, Seoul, South Korea

---

Mendelian randomization (MR) is a method of using genetic variants as instrumental variables (IV) to estimate causal effect between exposure and outcome. Recently, summary statistics released from genome-wide association studies with large sample size facilitated MR by providing exposure effect sizes of multiple genetic variants. This type of MR analysis utilizing external dataset for exposure effect sizes is called two-sample MR design. The causal effect between exposure and outcome is commonly obtained by combining the per-variant ratio estimates via inverse-variance weighted method. In this procedure, conventionally, the standard error of the ratio estimate is approximated by the first-order term from the delta method. However, as indicated in [Thomas et al., 2007], this approximation can lead to underestimation of variance. Here, using simulations, we show how much an impact this approximation has on the false positive rate and power of MR. Our simulations showed that when the IVs were selected with stringent p-value threshold ( $5 \times 10^{-8}$ ), the errors in variance due to the first-order approximation was about 3%. Interestingly, this value appeared to be independent of the effect size ratio between the exposure effect and causal effect, and of the sample size ratio between the two datasets in the two-sample design. However, when the IVs were selected by less stringent threshold, the error in variance increased. In contrast, an analytical approximation including up to the second order term in variance was nearly as accurate as the empirical estimation. In sum, these results showed that there is no reason not to use the more accurate analytical approximation for the variance estimation of ratio estimates.

## PgmNr 3048: Identifying causal variants by fine mapping across multiple studies.

### Authors:

N. LaPierre<sup>1</sup>; K. Collins<sup>1</sup>; E. Eskin<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Computer Science, UCLA, Los Angeles, California.; 2) Computational Medicine, UCLA, Los Angeles, California.; 3) Human Genetics, UCLA, Los Angeles, California.

---

Genome-Wide Association Studies (GWAS) have identified numerous loci associated with a variety of common traits. However, within each of these loci, many SNPs are associated with the trait of interest, and most are likely to be non-causal variants that are in LD with a limited number of causal SNPs. Identifying these causal variants is crucial for elucidating the genetic architecture of complex traits, but it is too costly to test all significantly-associated variants with biological assays.

The problem of statistical “fine mapping” addresses this challenge by identifying a small set of variants that, with strong statistical evidence, contains all of the causal variants. Various methods have emerged in recent years to perform fine mapping, and have successfully been used to help identify causal variants. However, there is great interest in multiple study settings, such as trans-ethnic fine mapping, because they maximize sample size and therefore statistical power. The existing methods do not generalize well to this setting, because they do not take into account the fact that effect sizes of causal variants in different studies are expected to be correlated. Additionally, computational efficiency is critical due to the necessity of inverting large matrices that contain many SNPs from multiple studies.

In this work, we present a method that performs statistical fine mapping efficiently in the multiple-study setting. We generalize a model introduced by CAVIAR and extended by CAVIAR-BF and FINEMAP, in which association statistics are modeled from a multivariate normal distribution and a Bayesian framework is applied to find minimum-sized putative sets of causal SNPs with high posterior probability. The covariance matrix for causal SNPs in all studies models the correlation between the causal SNP effect sizes in different studies according to the random effects model. This matrix must be inverted and its determinant must be taken to compute the multivariate normal likelihood function. However, we show that its structure forms a repetitive “block” structure that can be leveraged to compute the inversion and determinant efficiently.

Our method provides an efficient, statistically sound framework for performing fine mapping while leveraging information from multiple studies, and therefore shows promise for improved discovery of causal variants across populations.

## PgmNr 3049: Fine-tuning polygenic risk score models using GWAS summary statistics.

### Authors:

Q. Lu <sup>1,2</sup>; Z. Zhao <sup>1</sup>; Y. Yi <sup>3</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin.; 2) Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin.; 3) Eli Lilly and Company, Indianapolis, Indiana.

---

Polygenic risk scores (PRSs) have wide applications in human genetics research. It is straightforward to construct a PRS using variant effect estimates from genome-wide association studies (GWAS). More sophisticated methods, including Bayesian frameworks (e.g. LDpred) and penalized regression models (e.g. LASSOSUM), have also been developed. These methods use GWAS association statistics as input to re-estimate variant effects and improve PRS performance. Notably, from the simplest approach to highly sophisticated methods, nearly all PRS models include tuning parameters. These parameters, such as the p-value cutoff in simple PRS and the causal variant proportion in LDpred, add flexibility to PRS models and improve the predictive performance when properly selected. Although fine-tuning methods are well established in the statistical learning literature, most methods (e.g. cross-validation) require individual-level data and do not apply to PRS methods. Consequently, tuning parameters in PRS need to be selected by maximizing the predictive performance on a validation dataset that is independent from both the input GWAS and the testing dataset in downstream application. However, such data rarely exist in practice, creating a significant gap between PRS methodology and application. Here, we introduce PUMAS (Parameter-tuning Using Marginal Association Statistics), a novel method to fine-tune PRS models. PUMAS has two key steps. First, we sample marginal association statistics for a subset of individuals based on the complete GWAS summary data. Using this, we can generate summary statistics for a training set (e.g. 75% of the samples) and a testing set (e.g. 25% of the samples) without actually partitioning the samples. Second, we propose an approach to evaluate the predictive performance of PRS using summary statistics from the validation set. Both steps in PUMAS are statistically rigorous, computationally efficient, and highly novel. Combining these two steps, PUMAS can conduct a variety of model tuning procedures including cross-validation for PRS models that use GWAS summary data as input. Through extensive simulations, we demonstrate that the performance of PUMAS is nearly as good as cross-validation based on individual-level data. Further, we apply our method to publicly available GWAS summary statistics and provide an atlas of optimized PRSs for thousands of diseases and traits. We believe our method will greatly benefit risk prediction applications.

## PgmNr 3050: Multi-ancestry meta-analysis of genome-wide association studies.

### Authors:

A.R. Martin<sup>1,2,3</sup>; P. Turley<sup>1,2,3</sup>; R.K. Walters<sup>1,2,3</sup>; H. Li<sup>4</sup>; M. Kanai<sup>1,2,3,5</sup>; C.E. Carey<sup>1,2,3</sup>; D. Palmer<sup>1,2,3</sup>; K.J. Karczewski<sup>1,2,3</sup>; M. Zacher<sup>6</sup>; J. Koskela<sup>7</sup>; G. Wojcik<sup>8</sup>; M. Akiyama<sup>9,10</sup>; C.R. Gignoux<sup>11</sup>; E.E. Kenny<sup>12</sup>; Y. Okada<sup>5,13</sup>; Y. Kamatani<sup>9,14</sup>; D. Cesarini<sup>15,16</sup>; D. Benjamin<sup>16,17,18</sup>; S. Ripatti<sup>7</sup>; A. Palotie<sup>7</sup>; M.J. Daly<sup>1,2,3,7</sup>; B.M. Neale<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; 2) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; 3) Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA; 4) Department of Biostatistics, Harvard Chan School of Public Health; 5) Department of Statistical Genetics, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan; 6) Department of Sociology, Harvard University, Cambridge, MA; 7) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki 00014, Finland; 8) Department of Genetics, Stanford University, Stanford, CA 94305, USA; 9) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 10) Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Fukuoka 812-8582, Japan; 11) Colorado Center for Personalized Medicine, University of Colorado, Anschutz Medical Campus, Aurora, CO 80045, USA; 12) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA; 13) Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Suita, Japan; 14) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 15) Department of Economics and Center for Experimental Social Science, New York University, New York, NY, USA; 16) National Bureau of Economic Research, 1050 Massachusetts Ave, Cambridge, MA 02138, USA; 17) Behavioral and Health Genomics Center, Center for Economic and Social Research, University of Southern California, 635 Downey Way, Los Angeles, CA 90089, USA; 18) Department of Economics, University of Southern California, 635 Downey Way, Los Angeles, CA 90089, USA

---

The vast majority of genome-wide association studies (GWAS) are performed in individuals of European descent, limiting their applicability to other populations. However, increasingly large genetic datasets of diverse individuals are beginning to emerge. These data present key methodological challenges regarding how best to estimate effect sizes for each population. We have developed the multi-ancestry meta-analysis (MAMA) method, which produces unbiased and precise effect size estimates for each population. This method models summary statistics together with LD patterns within and across all available population pairs by extending two related methods, LD score regression and multi-trait analysis of GWAS (MTAG). In this approach, more information is shared across populations for genomics regions where LD patterns are similar and when estimation error is low, and less information is shared otherwise; we leverage these relationships to boost signal where local genomic effects are consistent across populations. We have computed LD scores from 15,708 whole genomes in gnomAD to use with MAMA or for other purposes. Our approach is unique among existing trans-ethnic meta-analysis approaches in its direct modeling of LD, production of easily interpretable summary statistics, ability to consider two or more cohorts or populations, and

generation of summary statistics for each population. We apply this method to summary statistics for quantitative and disease traits from large-scale biobanks, including the UK Biobank, FinnGen, BioBank Japan, and China Kadoorie Biobank. Unlike several existing methods, MAMA is unbiased and has correct statistical coverage; among methods that meet these criteria, MAMA is the most precise, has the highest power, and is computationally efficient. Thus, MAMA is an effective method for jointly analyzing multi-ethnic GWAS data, providing an approach towards the goal of ensuring that genetics does not exacerbate disparities for those already most underserved.



## PgmNr 3051: Polygenic phenotype analysis by estimating risk scores and associated pathways.

### Authors:

R. Ota; S. Morishita

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo

---

A number of causal genomic mutations associated with diseases have been reported. These mutations on genes can alter the functions of the genes and may also affect the subsequent pathways. For studying polygenic diseases, genome-wide association study (GWAS) has been largely contributing to identifying single nucleotide polymorphisms (SNPs); however, SNPs for polygenic diseases often have low odds ratio (1.1-1.3) and only explain a small proportion of heritability, which is called 'missing heritability.'

Recently, Khera *et al.* has reported that one can identify SNP sets with large odds ratio ( $>3.0$ ) by using LDpred (Vilhjalmsson *et al.*, 2015), which calculates polygenic risk scores via estimation of effect size of each SNP under a linear additive model. However, it remains largely unexplored how to interpret an enormous number of relevant SNPs output by LDpred (e.g. 6.6 million associated SNPs are claimed to be associated to coronary artery disease, CAD, Khera *et al.*, 2018). With numerous candidate SNPs, it is difficult to understand mechanisms (e.g. relevant pathways) of polygenic diseases.

Although a number of pathways associated with many diseases have been reported by GWAS, most of previous pathway analyses assume that only a small number of genes have larger effects than the others in contrast to numerous SNPs with small effect sizes in polygenic diseases. To settle this problem, we propose an extension of the theory of polygenic risk score, aiming to identify polygenic-disease-related pathways and functions of mutations by associating a set of mutations with genes which have a large effect size on the relevant pathway.

We introduce the notion of effect sizes of SNVs in each pathway whose sums are assumed to be effect sizes of the SNVs, thereby enabling us to evaluate the contribution of each pathway to the focal phenotype while considering the situation where a SNV is involved in several pathways. We evaluated the effect sizes of SNVs in each pathway from summary statistics of GWAS using Bayesian statistics. We then introduce indicators showing how large each pathway contributes to the phenotype. One can identify with the indicators which pathway is potentially causal.

This proposed new method allowed us to interpret the result of polygenic risk score even in the presence of a large number of associated SNPs and is therefore useful in an era that thousands or millions of exome sequencing data or whole genome sequencing data are available.

## PgmNr 3052: Case-case GWAS of two different disorders using case-control summary statistics.

**Authors:**

W.J. Peyrot; A.L. Price

[View Session](#) [Add to Schedule](#)**Affiliation:** Harvard T.H. Chan School of Public Health, Boston, USA

---

Comparing cases of different disorders can help elucidate biological differences between disorders (PGC 2018 Cell), but currently requires individual-level data from cases of both disorders. We developed a new method (CC-GWAS) to test for differences in allele frequency among cases of two different disorders using summary statistics from the respective case-control GWAS. CC-GWAS (i) maximizes power by employing sample-size dependent weights for each case-control GWAS that minimize the expected genome-wide squared difference between estimated and true effect sizes, accounting for sample overlap, and (ii) controls type I error by incorporating theoretical weights based on infinite sample size. CC-GWAS can be extended to incorporate case-case summary statistics when available. We confirmed via simulations and analytical computations that CC-GWAS attains accurate control of type I error and higher power than other approaches.

We applied CC-GWAS to summary statistics from case-control GWAS of Crohn's disease (CD;  $N=6k$  cases, 15k controls, 55 independent GWAS loci) and ulcerative colitis (UC; 7k cases, 20k controls, 38 independent GWAS loci) with an estimated genetic correlation of 0.58. CC-GWAS identified 20 independent genome-wide significant loci with different allele frequencies in CD cases vs. UC cases, including 7 novel loci that were independent from the 55 CD and 38 UC loci. For replication, we applied CC-GWAS to independent CD and UC case-control ImmunoChip summary statistics. Summary statistics were available for 13 of the 20 loci, including 2 of 7 novel loci. All 13 loci were significantly replicated ( $P<0.05/13$ ) with the same direction of effect.

We also applied CC-GWAS to summary statistics from case-control GWAS of schizophrenia (SCZ; 41k cases, 65k controls, 97 independent GWAS loci) and major depressive disorder (MDD; 171k cases, 329k controls, 45 independent GWAS loci) with an estimated genetic correlation of 0.29. CC-GWAS identified 72 independent genome-wide significant loci with different allele frequencies in SCZ cases vs. MDD cases, including 6 novel loci independent from the 97 SCZ and 45 MDD loci.

In conclusion, CC-GWAS robustly identifies novel loci with different allele frequencies among cases of different disorders using case-control summary statistics. Thus, CC-GWAS can help elucidate biological differences between disorders, which may facilitate improved clinical diagnoses and treatment.

## PgmNr 3053: Testing epistasis between eQTLs using whole-genome sequencing and RNA-seq data in COPDGene.

### Authors:

D. Qiao<sup>1</sup>; P.J. Castaldi<sup>1,2</sup>; C.P. Hersh<sup>1,3</sup>; A. Saferali<sup>1</sup>; M.M. Parker<sup>1,4</sup>; D. Prokopenko<sup>1,5</sup>; S. Won<sup>6</sup>; S. Lee<sup>1</sup>; E.K. Siverman<sup>1,3</sup>; C. Lange<sup>7</sup>; M.H. Cho<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Channing Division of Network Medicine, Department of Medicine, Brigham & Women's Hospital, Boston, Massachusetts.; 2) Division of General Internal Medicine and Primary Care, Brigham & Women's Hospital, Boston, Massachusetts; 3) Division of Pulmonary and Critical Care Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts; 4) Alnylam Pharmaceuticals, Cambridge, Massachusetts; 5) Massachusetts General Hospital, Boston, Massachusetts; 6) Department of Public Health Sciences, Seoul National University, Seoul, Korea; 7) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts

**Introduction:** Epistasis has long been hypothesized to be present in complex human traits, but identifying these effects has been challenging. Major limitations to date include lack of comprehensive sequence data, insufficient power, and control of false positives in statistical testing. With the availability of large-scale whole-genome sequencing (WGS) data and RNA-Seq data in COPDGene, we are interested in identifying SNP-SNP interactions that are associated with gene expression levels. Gene-expression is an ideal phenotype to study epistasis since there are known molecular interactions and large effect sizes have been shown for marginal expression quantitative trait locus (eQTL) studies.

**Methods:** To explore the properties of interaction tests, we used simulated WGS data and RNA-Seq data, and real data from the COPDGene study to perform simulations under several scenarios: 1) Two SNPs with or without linkage disequilibrium (LD), where 0, 1, or 2 of them is causal; 2) One of the two SNPs is interacting with a third SNP (heteroscedasticity); 3) both SNPs are in LD with the same causal SNP (confounding); and 4) both SNPs are in LD with a different causal SNP. We used the Wald test for interactions, the Wald test with heteroscedasticity-consistent sandwich estimator, the 4-df test from Cordell's model, and the 4-df test with sandwich estimator, with different transformations of the RNA-Seq count data. We also looked at the type I error rates while conditioning on one causal SNP under scenarios 3 and 4.

**Results:** In a simulation study of 1,000 subjects, we found that tests using sandwich estimators require parametric bootstrapping to maintain the type I error rate. Tests using log transformation maintain the type I error rate, with little reduction in power. Methods without sandwich estimators do not guard against heteroscedasticity or confounding. Tests using sandwich estimator guard against heteroscedasticity, and could control the type I error rate after conditioning on one causal SNP. The Wald test has similar power to the 4-df test in most scenarios.

**Conclusions:** We recommend a general strategy to identify epistasis on gene expression, including 1) log transformation; 2) screening using the Wald test; 3) additional screening using the Wald test with sandwich estimators with parametric bootstrapping; and 4) conditional analysis to remove false positives due to confounding. We are currently testing this method using COPDGene data to identify epistasis eQTLs.



## PgmNr 3054: Estimating the autocorrelation of causal minor allele effect sizes as a function of genomic distance.

### Authors:

A. Schoech<sup>1,2</sup>; O. Weissbrod<sup>1,2</sup>; L. O'Connor<sup>1,2</sup>; N. Patterson<sup>2</sup>; Y. Reshef<sup>1,2</sup>; A. Price<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA

---

Most models of complex trait genetic architecture implicitly assume that signed causal trait effect sizes of nearby SNPs (defined with respect to the minor allele) are uncorrelated, i.e. the fact that a causal SNP lies close to a trait increasing SNP does not change its probability of being itself trait increasing rather than trait decreasing. Here we investigate this assumption, and evaluate implications for defining and estimating SNP-heritability. We developed a method to estimate the genome-wide autocorrelation of causal minor allele effect sizes as a function of genomic distance. The idea behind the inference method is that GWAS summary statistics of nearby SNPs can be correlated both due to correlated causal effects as well as LD tagging. Our method regresses products of GWAS summary statistics on distance dependent LD scores, accounting for LD tagging and yielding estimates of causal effect correlations. We determined that the method produces unbiased estimates in simulations, when model assumptions hold.

We applied the method to 18 UK Biobank traits, analyzing 114K UK-ancestry individuals and 18 million imputed SNPs (MAF>0.1%) and meta-analyzing results across traits. Our estimate of the correlation of causal minor allele effect sizes of SNP pairs <1000bp apart was  $-0.14 \pm 0.04$ , with negative estimates for 15 of 18 traits; we are currently incorporating explicit modeling of LD-dependent causal architectures into these estimates. Our result is supported by findings of Brown et al. (2016 Genetics), who observed many examples of pairs of adjacent linked SNPs with strong opposite signed effects, a phenomenon the authors term linkage masking. Our initial analysis implies that linkage masking is widespread and broadly impacts the architecture of complex traits. A plausible explanation for this widespread effect is natural selection: large-effect loci are often under strong selection, quickly removing isolated variants from these sites. In contrast, linked variant pairs with opposite signed effects neutralize each other and can therefore remain polymorphic. Our results suggest that GWAS analyses may commonly underestimate true causal effects, since linked SNPs with opposing effects cancel each other out. Furthermore, we show that the assumptions underlying commonly used definitions of SNP-heritability do not hold when causal SNP effects are correlated, and propose an alternative definition.

## PgmNr 3055: Perturbation-based clustering linear combination approach for analyzing multiple phenotypes.

### Authors:

Q. Sha; S. Yan; S. Zhang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Mathematical Sciences, Michigan Technological University, Houghton, Michigan

Although the conventional genome-wide association studies (GWASs) focus on a single phenotype, there is an increasing interest in joint analysis of multiple phenotypes because cohort studies usually contain multiple phenotypes and jointly analyzing multiple phenotypes may increase statistical power. In addition, pleiotropy, the phenomenon of one genetic locus influencing several traits, has become a widely researched field in human genetics. Association studies based on multiple phenotypes can also provide great opportunities to systematically assess the genetic overlap between multiple phenotypes and diseases.

Recently, many statistical methods have been developed for joint analysis of multiple phenotypes, including the clustering linear combination (CLC) method [Bioinformatics, 2018]. In the CLC method, one first clusters individual statistics from the association tests for each phenotype into positively correlated clusters, then combine the individual statistics linearly within each cluster and combine the between-cluster terms in a quadratic form. It was theoretically proved that if the individual statistics can be clustered correctly, CLC is the most powerful test among all tests with certain quadratic forms. Due to the unknown number of clusters for a given data, the test statistic of CLC is the minimum p-value among all p-values of the test statistics obtained from each possible number of clusters. Therefore, a simulation procedure must be used to evaluate the p-value of the final test statistic. This makes CLC computationally demanding. In this presentation, we used a recently developed perturbation clustering approach to determine the number of clusters in the CLC method. It was shown that perturbation clustering approach is able to find the correct clusters when the data consists of distinct classes. Using the unique clusters identified by perturbation, the test statistic of CLC has an asymptotic distribution, which is very computationally efficient and makes perturbation-based CLC (PCLC) applicable for GWASs.

We used simulation studies as well as an application to a real data set to evaluate the performance of PCLC. Simulation results showed that PCLC is either the most powerful test or has similar power to the most powerful test among the tests we compared. By using perturbation, PCLC increases power comparing with CLC. In the real data analysis of COPDGene, PCLC identified more significant SNPs related to COPD than other tests.

## PgmNr 3056: Instrumenting the transcriptome at scale: eQTL colocalization across the UK Biobank for target identification.

### Authors:

K.B. Sieber <sup>1</sup>; M. Hurle <sup>1</sup>; G. Dall-Olio <sup>2</sup>; V.D. Kumar <sup>1</sup>; L. Yerges-Armstrong <sup>1</sup>; R. Scott <sup>3</sup>; T. Johnson <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, GlaxoSmithKline, Collegeville, Pennsylvania.; 2) DCS, GlaxoSmithKline, Stevenage, United Kingdom.; 3) Human Genetics, GlaxoSmithKline, Stevenage, United Kingdom.

---

A key limitation in the translation of genome-wide association studies into the identification of the next generation of therapeutic targets is the difficulty in identifying potentially causal genes underlying genetic associations. This is made more difficult by the fact that the majority of GWAS associations are driven by non-coding variants. Incorporation of expression quantitative trait loci (eQTLs) is a promising approach to prioritize potentially causal genes where there is evidence that the same variant(s) affect both gene expression level and the disease phenotype. Many studies have focused on reciprocal single-variant lookups between gene expression and disease results, which are prone to spurious association findings making Bayesian colocalization approaches a particularly attractive alternative.

Here, we demonstrate scalability and test utility in target identification by systematically applying colocalization of all GTEx and Blueprint eQTLs with 34,200 genome-wide significant associations from GWAS of 2,120 phenotypes from UK Biobank (~2.6 billion tests in total). By evaluating success of historic drug development pipelines, we find that colocalization of an eQTL-phenotype pair is associated with a higher rate of successful drug development and conclude that this supports the ability of colocalization approaches to identify causal associations at scale. By taking a comprehensive approach to colocalization, we are now able to implement colocalization in a PheWAS-like manner and have the capacity to instrument up to 18,720 protein coding genes in at least 51 tissues, offering an exciting opportunity for therapeutic target discovery and validation.

## PgmNr 3057: Fast multivariate estimation for heterogeneous traits.

**Authors:**

B.W. Ting; Y.-H. Zhou; F.A. Wright

[View Session](#) [Add to Schedule](#)

**Affiliation:** North Carolina State University, Raleigh, NC.

---

As has been commonly observed, decreasing sequencing costs and increasing computational power in recent years have provided an increasing availability of genomic data. Such increased availability of data naturally invites greater demand for methodological sophistication in modeling, testing, and inference. However, at the same time, striving for such methodological improvements can face material headwind in the form of computational challenges and/or lack of clear statistical frameworks.

Thus, we look to meet this demand by presenting some novel methods for genomics and precision medicine, but can also be applied in other settings. High priority concerns include computation time and statistical efficiency.

We introduce a Two-Stage Composite Likelihood approach for multivariate probit estimation for responses with binary components. This approach is designed to be fast, lending itself well to settings such as Genome Wide Association Studies (GWAS) for detecting Single Nucleotide Polymorphism (SNP) associations with mean changes in phenotype.

For the multivariate probit, we are chiefly concerned with settings where the correlation parameters are unconstrained (aside from variances constrained at 1)–and among these in particular, where the coefficient parameters are unconstrained, as well. Historically, multivariate probit estimation has been slow and/or computationally intensive. As the number of components and predictors increases, the number of coefficient and correlation parameters tends to grow quickly, adding to the difficulty of numerical optimization in finding estimates. This can be especially challenging in the genomics space, due to the number of predictors involved.

Next, we extend this approach to incorporate heterogeneous multivariate responses, i.e. where the response can include both binary and continuous components. For example, in genomics and/or precision medicine settings, these could be phenotypic traits and/or disease statuses. Now instead of having only bivariate probits as associated likelihoods, there can also be bivariate normal densities for continuous-continuous pairs, as well as associated likelihoods for binary-continuous pairs.

Further considerations could include heteroskedasticity in the GWAS setting, i.e. nonconstant variance across allele counts. There could be possible tie-ins here with gene-gene and/or gene-environment interactions. SNPs associated with variance may be indicative of the presence of interactive effects.



## PgmNr 3058: Approximately LD independent regions for biobank scale genetic analyses.

### Authors:

J.C. Ulirsch<sup>1,2</sup>; M. Kanai<sup>1,3,4</sup>; L. Abbott<sup>1</sup>; C. Seed<sup>1</sup>; J. Bloom<sup>1</sup>; H.K. Finucane<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Broad Institute, Cambridge, MA.; 2) Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA.; 3) Program in Bioinformatics and Integrative Genomics, Harvard Medical School, Boston, MA.; 4) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA.

A number of statistical methods have been developed to improve on the biological return of genome-wide association studies (GWASs) that have identified thousands of genomic loci linked to complex traits and diseases. Many of these methods, including genetic fine-mapping (e.g. fgwas, FINEMAP, and SuSiE), estimators of local heritability and genetic correlation (HESS and p-HESS), and signed estimators of polygenic TF binding disruption (SLDP) are performed at or have key procedures that are performed at a per-locus level. For these methods to produce valid estimates and/ or inference, loci are assumed to be independent (*i.e.* there is negligible linkage disequilibrium (LD) between SNPs in different loci). Typically, to identify approximately LD independent (ALDI) regions, **(A)** a small window (0.5 - 3 Mb or cM) is taken around a SNP of interest or **(B)** the “anti-diagonal” LD vector is minimized at the boundaries, often using a 1000G reference panel (Berisa and Pickrell 2016).

Here, we compute in-sample dosage LD from 361,144 densely genotyped, unrelated individuals in the UK Biobank, and find considerable LD among regions identified as approximately independent using the Berisa and Pickrell method in 1000G. We then introduce a new method for identifying ALDI regions that minimizes different functions of the full “off-diagonal” LD band at region boundaries and apply it to our UK Biobank LD matrix. We show that our method **(1)** allows for application-specific parameter tuning to balance approximate independence with region length, **(2)** produces boundaries correlated with relevant features such as recombination hotspots, **(3)** validates previously reported (Price et al. 2008) regions of long range (3 Mb - 10Mb) LD, and **(4)** highlights the impact of newly arisen, low allele frequency (0.1 - 1%) variants on boundary identification.

In simulations, we show that methods aimed at gaining insight into genetic architecture and underlying trait biology, particularly genetic fine-mapping, can be less biased when using our definitions of ALDI regions compared to other widely used definitions. In up to 361,144 individuals from the UK Biobank, we performed genetic fine-mapping on over 30 complex traits and diseases using FINEMAP and SuSiE. We identify thousands of likely causal genetic variants with high posterior probability (> 0.90) and assess the impact of ALDI region choice. Our method will have wide application to a number of commonly used statistical genetics methods.

## PgmNr 3059: Evaluation of modern approaches for the complex trait prediction using genetic data.

### Authors:

M. Zhang <sup>1</sup>; J.M. Granka <sup>1</sup>; L.T. Nguyen <sup>1</sup>; T. Arbel <sup>1</sup>; J. Han <sup>1</sup>; K. Noto <sup>1</sup>; E.L. Hong <sup>1</sup>; K.G. Chahine <sup>2</sup>; C.A. Ball <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ancestry.com, San Francisco, California.; 2) Ancestry.com, Lehi, Utah.

---

Genome-wide association studies (GWAS) have been widely used to discover the genetic basis of complex phenotypes. As such, GWAS provide starting points and potential opportunities for researchers to develop methods to predict complex traits. As a natural extension to GWAS, the polygenic risk score (PRS) is one of most popular methods for complex traits prediction in the genetics community. The main motivations behind PRS are 1) only GWAS summary statistics are needed in the PRS calculation; 2) multiple SNPs works together to predict a given trait. At the same time, other PRS extensions and

alternatives have also been developed in the context of greater precision in prediction. To explore the advantages and disadvantages among various prediction methods under different genetic architectures, we leverage genotype data and phenotype data collected from surveys from a subset of over one million AncestryDNA customers who have consented to research. First, we perform genome-wide association analyses on various traits with different architectures, and select promising markers for prediction from these GWAS using various forms of linkage disequilibrium (LD) pruning and significance thresholding. Then, we assess the prediction ability of standard PRS and other models. Our study shows that no specific method dominates across all traits, as each method has its merits in terms of model complexity and prediction accuracy under certain genetic architectures. While more in-depth research is still required, this work provides a sound foundation in the comparison of various statistical techniques for prediction, as well as in our attempts to understand and effectively interpret variation in the human genome.

# PgmNr 3060: UK Biobank whole exome sequence binary phenome analysis with robust region-based rare variant test.

## Authors:

Z. Zhao <sup>1</sup>; W. Bi <sup>1</sup>; W. Zhou <sup>2,3</sup>; L.G. Fritsche <sup>1</sup>; S. Lee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, United States of America; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, United States of America; 3) Program in Medical and Population Genetics and Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, United States of America

---

In biobank data, most binary phenotypes have unbalanced case-control ratios, which can cause inflation of the type I error rates. Recently, saddlepoint approximation (SPA) based single variant test has been developed to provide an accurate and scalable method to test associations of such phenotypes. However, for region-based rare variant tests, the existing methods are either less accurate when case-control ratios are unbalanced or not scalable for large data analyses.

To address these problems, we propose robust SKAT/SKAT-O type region-based tests that adjust single variant score statistics using SPA and efficient resampling (ER), and aggregate the adjusted statistics. The accuracy of region-based tests largely depends on the accuracy of estimated variances of single variants. The SPA and ER help to precisely calculate the reference distribution of the single variant score statistics, thereby properly controlling for the type I error rates.

Through simulation studies, we show that the proposed method provides well-calibrated p-values. In contrast, the unadjusted approach has greatly inflated type I error rates ( $2.25 \times 10^{-4}$  when exome-wide  $\approx 2.5 \times 10^{-6}$ ) when the case-control ratio is 1:99. Additionally, the empirical power of the robust methods is almost the same as or larger than unadjusted methods across all case-control ratios. Furthermore, the computation cost of the robust methods is comparable to unadjusted tests, and hence can be applied to large biobank data.

We applied the robust region-based tests to UK Biobank whole exome sequence (WES) data analysis with 45,596 unrelated European samples and analyzed 791 binary phenotypes. We tested rare variants ( $MAF < 0.01$ ) of the nonsynonymous and splicing variants in the exon and neighboring regions of 18,360 genes. The analysis identified 117 gene $\times$ phenotype associations with p-values  $< 2.5 \times 10^{-6}$ , where 10 associations had p-values  $< 10^{-7}$ , including the associations between *JAK2* and myeloproliferative disease (p-value:  $8.92 \times 10^{-30}$ ; previously reported), *TNC* and large cell lymphoma ( $6.10 \times 10^{-9}$ ), and *HOXB13* and prostate cancer ( $3.00 \times 10^{-8}$ ; previously reported). Among the 10 gene $\times$ phenotype pairs, only 2 had a single rare-variant p-value  $< 5 \times 10^{-8}$ , suggesting that region-based approaches can be more powerful than single variant analyses. All ten associations remained significant after the conditional analysis, which indicates that these rare variant association signals are not driven by the nearby common variants.

## PgmNr 3061: Testing and controlling for horizontal pleiotropy with the probabilistic Mendelian randomization in transcriptome-wide association studies.

### Authors:

Zhou <sup>1</sup>; Z. Yuan <sup>1,2</sup>; H. Zhu <sup>1</sup>; P. Zeng <sup>3</sup>; S. Yang <sup>1</sup>; S. Sun <sup>1</sup>; C. Yang <sup>4</sup>; J. Liu <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biostatistics, University of Michigan, Ann Arbor, Michigan.; 2) Department of Biostatistics, School of Public Health, Shandong University, Jinan, China.; 3) Department of Epidemiology and Biostatistics, Xuzhou Medical University, Xuzhou, Jiangsu, China; 4) Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong, China; 5) Centre for Quantitative Medicine, Program in Health Services and Systems Research, Duke-NUS Medical School, Singapore

---

Integrating association results from both genome-wide association studies (GWASs) and expression quantitative trait locus (eQTL) mapping studies has the potential to shed light on the molecular mechanism underlying disease etiology. Several statistical methods have been recently developed to integrate GWASs with eQTL studies in the form of transcriptome-wide association studies (TWASs). These existing methods can all be viewed as two sample Mendelian randomization (MR) methods, which are also widely used in various GWASs for inferring the causal relationship among complex traits. Unfortunately, most existing TWAS and MR methods make an unrealistic modeling assumption that the instrumental variables do not exhibit horizontal pleiotropic effects. However, horizontal pleiotropic effects have been recently discovered to be wide spread across complex traits, and as we will show here, are also wide spread across gene expression traits. Therefore, allowing for no horizontal pleiotropic effects can be overall restrictive, and, as we will be show here, can lead to a substantial inflation of test statistics and subsequently false discoveries in TWAS applications. Here, we present a probabilistic MR method, which we refer to as PMR-Egger, for testing and controlling of horizontal pleiotropic effects in TWAS applications. PMR-Egger relies on a new MR likelihood framework that unifies many existing TWAS and MR methods, accommodates multiple correlated instruments, tests the causal effect of gene on trait in the presence of horizontal pleiotropy, and, with a newly developed parameter expansion version of the expectation maximization algorithm, is scalable to hundreds of thousands of individuals. With extensive simulations, we show that PMR-Egger provides calibrated type I error control for causal effect testing in the presence of horizontal pleiotropic effects, is reasonably robust for various types of horizontal pleiotropic effect misspecifications, is more powerful than existing MR approaches, and, as a by-product, can directly test for horizontal pleiotropy. We illustrate the benefits of our method in applications to 39 diseases and complex traits obtained from three GWASs including the UK Biobank and show how PMR-Egger can lead to new biological discoveries through integrative analysis.

## PgmNr 3062: The association between African ancestry and telomere length across the African diaspora: Evidence from the CAAPA study.

### Authors:

K.R. Iyer<sup>1</sup>; M.A. Taub<sup>2</sup>; M. Daya<sup>3</sup>; S. Chavan<sup>3</sup>; N. Rafaels<sup>3</sup>; M.P. Boorgula<sup>3</sup>; C. Ober<sup>5</sup>; J.G. Wilson<sup>6</sup>; J. Knight-Madden<sup>7</sup>; L. Caraballo<sup>8</sup>; M.G. Foreman<sup>9</sup>; C.O. Olopade<sup>10</sup>; M. Yazdanbakhsh<sup>11</sup>; A. Mayorga<sup>12</sup>; E.F. Herrera-Paz<sup>13</sup>; J. Marrugo<sup>14</sup>; L.B. Ware<sup>15</sup>; T. Hartert<sup>15</sup>; M.I. Araujo<sup>16</sup>; E. Bleecker<sup>17</sup>; D. Meyers<sup>17</sup>; K.C. Barnes<sup>3</sup>; T.H. Beaty<sup>1</sup>; R.A. Mathias<sup>1,4</sup>; on behalf of the Consortium on Asthma among African Populations of the Americas (CAAPA)

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, 21205, USA; 2) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, 21205, USA; 3) Department of Medicine, University of Colorado, Denver, 80045, USA; 4) Department of Medicine, Johns Hopkins University, Baltimore, 21224, USA; 5) Department of Human Genetics, University of Chicago, Chicago, 60637, USA; 6) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, 39216, USA; 7) Caribbean Institute for Health Research, The University of the West Indies, Kingston, 00007, Jamaica; 8) Institute for Immunological Research, Universidad de Cartagena, Cartagena, 130000, Colombia; 9) Pulmonary and Critical Care Medicine, Morehouse School of Medicine, Atlanta, 30310, USA; 10) Department of Medicine and Center for Global Health, University of Chicago, Chicago, 60637, USA; 11) Department of Parasitology, Leiden University Medical Center, Leiden, 02333, Netherlands; 12) Centro de Neumología y Alergias, San Pedro Sula, 21102, Honduras; 13) Facultad de Medicina, Universidad Católica de Honduras, San Pedro Sula, 21102, Honduras; 14) Instituto de Investigaciones Immunologicas, Universidad de Cartagena, Cartagena, 130000, Colombia; 15) Department of Medicine, Vanderbilt University, Nashville, 37232, USA; 16) Immunology Service, Universidade Federal da Bahia, Salvador, 401110170, Brazil; 17) Department of Medicine, University of Arizona College of Medicine, Tucson, 85724, USA

---

Over the past two decades, telomere length (TL) has emerged as a marker of biological aging. TL is highly heritable, reflects gender dysmorphism and is also influenced by race/ethnicity. Some studies report that people of African ancestry harbor longer TL than those of European ancestry, while other studies observe the opposite trend or no differences. However, to our knowledge, the association between admixture components and TL has not been explored. We examined the association between estimated ancestry proportions and TL by leveraging existing whole genome sequence (WGS) data from the NHLBI-supported multi-centered, case-control Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA).

Global estimates of ancestry were obtained using ADMIXTURE on 677 CAAPA participants including African-admixed individuals from North, Central, and South America, the Caribbean Islands and Yoruba-speaking individuals from Ibadan, Nigeria. TL on each individual was estimated using a bioinformatic approach (TelSeq). After exclusions based on QC, linear regression models were implemented on the analytic samples (N=673) to assess the relationship between TL, African

ancestry (%YRI) and Native American ancestry (%NAT) across the African diaspora. After adjusting for age and sex, each percent increase in YRI and NAT was associated with a 14bp ( $p_{\text{YRI}} = 2.64\text{e-}05$ ) and 30bp ( $p_{\text{NAT}} = 6.66\text{e-}05$ ) increase in TL, respectively. Additional adjustments for asthma case status did not change this association.

Our results suggest that YRI and NAT admixture components are significantly associated with TL. As CAAPA genomes were collected across 21 different sites, we are also examining the possibility of technical sources of variation that may have been inevitably introduced in our TL data and could confound our associations. Prior GWAS predominantly used European ancestry populations to identify 16 loci associated with TL resulting in poorly understood ancestry-specific genetic architecture behind these loci. We propose to replicate these implicated loci using a candidate gene study design while concurrently probing for novel variants affecting TL in individuals of African descent in the CAAPA study.

## **PgmNr 3063: Phasing aware modelling of allele-specific expression for X chromosome inactivation prediction elucidates the important contribution of XCI escape genes to gender-biased diseases.**

### **Authors:**

R. Sauteraud <sup>1</sup>; L. Carrel <sup>2</sup>; D. Liu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA, 17033, USA; 2) Department of Biochemistry and Molecular Biology, Penn State College of Medicine, Hershey, PA, 17033, USA

X chromosome inactivation (XCI) mechanism randomly silences one female X and only expresses the alleles on the active X. Yet, it is estimated that ~30% of the X-linked genes may escape at least partially in a subset of individuals. Escape genes result in unequal expression levels between sexes, which may be disease causative. It is thus of great interest to identify XCI escape genes.

We seek to develop methods to infer XCI escape genes from bulk RNA-seq due to its broad availability. We first estimate the sample skewing using the allelic specific expression (ASE) from a set of commonly inactivated genes, which reflects the fraction of cells where a particular X is active. Then, we will identify escape genes as the ones with statistically significantly more balanced ASE than sample skewing. Current methods to obtain ASE either use the most expressed SNP or sum up read counts on phased haplotypes. Summing reads across haplotypes can be more powerful as it aggregates information across different SNPs. Yet, it may result in biased estimates of sample skewing when the haplotype phase is incorrect. To solve this problem, we propose a new mixture model based approach that incorporates the likelihoods with and without phasing errors as different components. The new approach allows us to obtain consistent estimates of the sample skewing and hence reliable inference of the escape genes even in the presence of phasing errors.

In simulations, we show that our model accurately corrects phasing errors to provide optimal read counts for each gene, and increase the power of identifying escape genes by 30%. We apply the proposed method to 217 females of the GEUVADIS project and obtained updated annotations for XCI escape status. Based on these annotations, we examined the contribution of X-linked genes to the disease heritability in the UK Biobank dataset. We show that escape and variable escape genes explain the largest proportion of X heritability, which is in large part attributable to X genes with Y homology. We also investigated the role of each XCI state in gender-biased diseases and found that while XY homologous gene pairs have a larger overall effect size, enrichment for variable escape genes is significantly increased in female-biased diseases. Our results, for the first time, quantitate the importance of variable escape genes for the etiology of gender-biased disease and our pipeline allows analysis of larger datasets for a broad range of phenotypes.

A- A+

## **PgmNr 3064: Multi-disease associated risk locus in *IL6* represses the anti-inflammatory gene *GPNMB* through chromatin looping and recruiting MEF2-HDAC complex.**

### **Authors:**

A.H. Sawalha <sup>1,2,3</sup>; X. Kong <sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Medicine, University of Pittsburgh, PA; 3) Department of Medicine, University of Michigan, Ann Arbor, MI; 4) Division of Rheumatology, Zhongshan Hospital, Fudan University, Shanghai, China

---

We have previously revealed a genetic association between Takayasu arteritis and a non-coding genetic variant in an enhancer region within *IL6* (rs2069837 A/G). The risk allele in this variant (allele A) has a protective effect against chronic viral infection and cancer. Using a combination of experimental and bioinformatics tools, we identified the monocyte/macrophage anti-inflammatory gene *GPNMB*, ~520kb away, as a target gene regulated by rs2069837. We revealed preferential recruitment of myocyte enhancer factor 2-histone deacetylase (MEF2-HDAC) repressive complex to the Takayasu arteritis risk allele. Further, we demonstrated suppression of *GPNMB* expression in monocyte-derived macrophages from healthy individuals with the AA compared to AG genotype, which was reversed by histone deacetylase inhibition. Our data show that the A allele in rs2069837 represses the expression of *GPNMB* by recruiting MEF2-HDAC complex, enabled through a long-range intra-chromatin looping. Suppression of this anti-inflammatory gene might mediate increased susceptibility to Takayasu arteritis and enhance protective immune responses in chronic infection and cancer. Our data highlight long-range chromatin interactions in functional genomic studies.



## PgmNr 3065: Preconception TGM1 carrier screening for autosomal recessive congenital ichthyosis: A new variant.

### Authors:

N. Abdelmoula <sup>1</sup>; W. Smaoui <sup>2</sup>; K.H. Trigui <sup>2</sup>; F. Abid <sup>2</sup>; S. Kammoun <sup>2</sup>; S. Aloulou <sup>2</sup>; M.H. Ben Ayed <sup>2</sup>; A. Sbai <sup>2</sup>; H. Yahia <sup>2</sup>; K. Chaker <sup>2</sup>; H. Fourati <sup>2</sup>; S. Fellah <sup>2</sup>; S. Ben Ali <sup>2</sup>; A. Yaich <sup>2</sup>; A. Kaabi <sup>2</sup>; H. Chaari <sup>2</sup>; W. Kammoun <sup>2</sup>; S. Kammoun <sup>2</sup>; O. Kaabi <sup>1</sup>; R. Louati <sup>1</sup>; N. Abdellaoui <sup>1</sup>; W. Ltaif <sup>2</sup>; M. Ben Aziza <sup>2</sup>; J. Jabr <sup>2</sup>; J. Dammak <sup>2</sup>; S. M'rabet <sup>2</sup>; H. Ben Ameer <sup>2</sup>; B. Abdelmoula <sup>2</sup>; M.N. Mhiri <sup>2</sup>; N. Gouiaa <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Lab Histology UR17ES36 MESRS, Univ Medicine, Sfax, Tunisia; 2) UR17ES36 MESRS, Univ Medicine, Sfax, Tunisia

---

Preconception care is the provision of biomedical, behavioural and social health interventions to women and couples before conception occurs. In this context, a first degree consanguineous couple has been seen at our genetic counselling following the birth of a Collodion male baby and asking for preconception testing before any other pregnancy project. The newborn was the second child of a 31 and 27 years old male and female, who have already a first healthy son. He was delivered by cesarean section at 38 WG. At birth he had dry and thick skin with deep fissures as well as generalized erythematous desquamation. He survived for 5 days at the neonatal intensive care unit with clinical diagnosis of Harlequin ichthyosis baby. Unfortunately, no DNA banking was made. Molecular investigation of TGM1 gene was done as a first step using direct sequencing of the entire coding parts and splice sites of the gene. While no pathogenic mutations were detected in the father's TGM1 gene, a heterozygous unclassified variant was detected in the mother: c.281G-A (p.Gly94Asp). The pathogenicity of this variant is very unclear, because it is a non-conservative change of a weakly conserved amino acid. The inheritance of congenital ichthyosis is autosomal recessive, and affected babies are usually homozygous for the mutation. So, molecular results of TGM1 gene screening don't allow explanation of the ARCI condition in the diseased child of the counselees. Mutation screening must be extended to the other genes related to ARCI to look for a composite status of mutations.

# PgmNr 3066: MHC risk haplotype sequencing and allele-specific genome editing by CRISPR/CAS9 system reveal *CCHCR1* as a susceptibility gene for alopecia areata.

## Authors:

N. Yoshihara <sup>1</sup>; A. Takagi <sup>1</sup>; E. Komiyama <sup>1</sup>; T. Mabuchi <sup>2</sup>; A. Otomo <sup>3</sup>; M. Ohtsuka <sup>3</sup>; A. Oka <sup>4</sup>; I. Shigaku <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Dermatology and Allergology, and Atopy (Allergy) Research Center, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo, Tokyo, Japan; 2) Department of Dermatology, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa, Japan; 3) Department of Molecular Life Sciences, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa, Japan; 4) The Institute of Medical Sciences, Tokai University, 143 Shimokasuya, Isehara, Kanagawa, Japan

## Introduction & Objectives:

Previous genome-wide association study about Alopecia areata (AA) results have implicated a number of immune and non-immune loci in the aetiology of AA. However, no variants among those have provided experimental evidence for biological functions between alleles and AA pathogenesis. Therefore, we attempted to pinpoint a susceptibility variant within the major histocompatibility complex (MHC) and to confirm the susceptibility variant.

## Materials & Methods:

We performed association and haplotype analysis for the MHC region to identify risk haplotypes using AA patients and controls. Next, we sequenced the risk and control haplotypes to identify AA-susceptibility variants. Next, we engineered mice carrying the human risk allele using allele-specific genome editing with the CRISPR/Cas9 system to reproduce AA phenotype.

## Results:

We sequenced risk and non-risk MHC chromosomal segments. Among the large number of variants there was only one nonsynonymous variant identical in the AA risk haplotype. The variant is located in the coiled-coil alpha-helical rod protein 1 (*CCHCR1*) gene. Next, mice were generated with the risk allele concordant with the variant of *CCHCR1* in humans, then we established mouse strains with risk alleles (AA mice). Half of AA mice displayed patched hair loss until 10 months after birth, thus successfully phenocopying the hair loss phenotype. Over time, the initial area of hair loss expanded in the majority of the AA mice, though constant and recovered hair loss was observed in some of those mice. Their surface displayed black spots, while the hairs appeared to be broken and tapering, similar to the conditions seen in humans with AA and specific for AA-associated hair loss.

## Conclusions:

First, the discovered susceptibility variant is a rare example of a non-mendelian common disease variant that can be phenocopied in a mouse model. Second, our alopecia model mouse model provides a novel mechanism for the observed hair loss and potential avenue for developing a drug for AA patients.

## PgmNr 3067: Regulation of impaired hepatic glucose homeostasis in a single extra copy of Down syndrome at critical region 1-4.

### Authors:

G. Chau <sup>1,2</sup>; S. Dong <sup>2</sup>; B. Bui <sup>4</sup>; K. Baek <sup>3</sup>; S. Um <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Translational Metabolic disease, University of medicine and pharmacy, Hochiminh, VietNam; 2) Department of Molecular Cell Biology, Samsung Biomedical Research Institute, Korea; 3) Samsung Advanced Institute for Health Sciences and Technology, Samsung Medical Center, Sungkyunkwan University, Korea; 4) Centre for molecular biomedicine, University of medicine and pharmacy, Hochiminh, Vietnam

---

### OBJECTIVES:

During fasting, hepatic gluconeogenesis is induced to maintain energy homeostasis. Moreover, abnormal dysregulation of hepatic glucose production is commonly observed in type 2 diabetes. However, the signaling components controlling hepatic glucose production to maintain normal glucose levels are not fully understood. Here, we examined the physiological role of Down syndrome critical region 1-4 (DSCR1-4), an endogenous calcineurin signaling inhibitor in the liver that mediates metabolic adaptation to fasting.

### METHODS:

We assessed the effect of cyclosporine A, an inhibitor of calcineurin signaling on gluconeogenic gene expression in primary hepatocytes. DSCR1-4 expression was examined in diet- and genetically-induced mouse models of obesity. We also investigated the metabolic phenotype of a single extra copy of DSCR1-4 in transgenic mice and how DSCR1-4 regulates glucose homeostasis in the liver.

### RESULTS:

Treatment with cyclosporin A increased hepatic glucose production and gluconeogenic gene expression. The expression of DSCR1-4 was induced by refeeding and overexpressed in obese mouse livers. Moreover, transgenic mice with a single extra copy of DSCR1-4 exhibited pyruvate intolerance and impaired glucose homeostasis. Mechanistically, DSCR1-4 overexpression increased phosphorylation of the cAMP response element-binding protein, which led to elevated expression levels of gluconeogenic genes and, thus, enhanced hepatic glucose production during fasting.

### CONCLUSION:

A single extra copy of DSCR1-4 results in dysregulated hepatic glucose homeostasis and pyruvate intolerance. Our findings suggest that nutrient-sensitive DSCR1-4 is a novel target for controlling hepatic gluconeogenesis in diabetes.

## PgmNr 3068: Common noncoding variant at PPP1R3B/LOC157273 promotes liver glycogen storage and metabolic syndrome, but protects against myocardial infarction.

### Authors:

N.D. Palmer <sup>1</sup>; B. Kahali <sup>3</sup>; Y. Chen <sup>2</sup>; M.F. Feitosa <sup>4</sup>; L.F. Bielak <sup>5</sup>; J.R. O'Connell <sup>6</sup>; S.K. Musani <sup>7</sup>; Y. Chen <sup>2</sup>; M.A. Province <sup>4</sup>; P.A. Peyser <sup>5</sup>; B. Halligan <sup>2</sup>; E.K. Speliotes <sup>2,8</sup>; on behalf of the Genetics of Obesity-associated Liver Disease (GOLD) Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina.; 2) Department of Internal Medicine, University of Michigan, Ann Arbor, MI; 3) Centre for Brain Research, Indian Institute of Science, Bangalore, India; 4) Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO; 5) School of Public Health, Department of Epidemiology, University of Michigan, Ann Arbor, MI; 6) Department of Endocrinology, Diabetes, and Nutrition, University of Maryland-Baltimore, Baltimore, MD; 7) Department of Medicine, University of Mississippi Medical Center, Jackson, MS; 8) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI

---

Human glycogen storage diseases are rare. Using multiple population-based cohorts we show that carriers of SNP rs4841132-A (frequency 2-19%) have increased liver attenuation, liver damage, serum lactate and triglyceride levels similar to what is seen in individuals with liver glycogen storage diseases. Rs4841132 is in the second exon of a liver-specific long noncoding RNA (*LOC157273*) and 190kb upstream of *PPP1R3B* which has been shown to regulate glycogen synthesis. Using allelic imbalance analysis, we demonstrate that rs4841132-A increases *PPP1R3B* expression through a cis genetic effect (allele on the same chromosome as the variant). We used the CRISPR/Cas9 system to produce a 105bp deletion of exon 2 of *LOC157273*, including the rs4841132 SNP. This deletion leads to increased levels of PPP1R3B protein and increased levels of glycogen. We also show that PPP1R3B overexpression increased glycogen in HuH-7 cells. Based on EHR data, rs4841132-A associates with all components of the metabolic syndrome: increased waist-to-hip ratio, hypertension, insulin resistance, serum triglycerides and decreased total serum high-density lipoprotein cholesterol. However, rs4841132-A decreases total cholesterol and myocardial infarction risk. A metabolic pattern for rs4841132-A includes increased serum glycine, lactate, triglycerides, and decreased serum acetoacetate, and beta hydroxybutyrate. These results show that rs4841132-A promotes hepatic glycogen storage disease by disrupting *LOC157273* allowing constitutive increase in PPP1R3B protein levels that promotes liver glycogen accumulation and promotes development of metabolic syndrome but lowers serum low-density lipoprotein cholesterol and myocardial infarction risk. These results suggest that elevated hepatic glycogen is one cause of metabolic syndrome that does not invariably promote myocardial infarction.

## PgmNr 3069: Enhancer-based immune dysregulation revealed by integrated analysis of T1D-associated genetic variants.

### Authors:

S. Kim; A.D. Hudgins; J. Yang; Y. Suh

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics, Albert Einstein College of Medicine, BRONX, NY.

---

Type 1 diabetes (T1D) is a polygenic autoimmune disease that results in the loss of functional pancreatic  $\beta$ -cells. Although chromosome 6 human leukocyte antigen (HLA) locus is a well-established T1D risk region, the mechanisms of the non-HLA genetic factors to the risk of T1D are still unclear. GWAS have identified over 200 T1D genetic variants, most of which lie in the non-coding regions of the genome, suggesting that gene regulatory changes substantially contribute to inter-individual differences in susceptibility to T1D. However, finding regulatory variants and their target genes are still challenging. To identify regulatory variants with a functional role in T1D risk, here we performed an integrated functional annotation analysis using various public databases. We first defined 1,817 candidate T1D SNPs, utilizing the GWAS catalog and 1000 Genomes Project European population data and then performed comprehensive annotation analyses using; 1) RoadMAP, ENCODE, and RegulomeDB for epigenome data; 2) GTEx for tissue-specific gene expression and eQTL data; 3) lncRNASNP2 for long non-coding RNA data. Integrating our results, we found 159 genes associated with the candidate T1D SNPs, and gene ontology analysis revealed an enrichment of the processes related to immune response and antigen presentation. We identified 26 high-probability causal SNPs, all of which were residing in enhancer regions. Furthermore, we found complex regulatory circuits at HLA (6p21) and non-HLA (16p11.2) loci. The HLA locus includes *rs886424* and *rs3129716* as well as 25 associated genes, including 6 HLA genes. The non-HLA locus contains *rs4788084*, *rs743590*, *rs762633*, and *rs62031562* as well as 19 downstream genes, including 3 immune-related genes (*IL27*, *SH2B1*, and *NUPR1*). Besides, among the 53 normal human tissues of GTEx, the HLA genes were highly expressed in Epstein-Barr virus (EBV)-transformed lymphocytes, lung, and spleen. From an independent cell type-specific enrichment analysis SNPsea using the candidate T1D SNPs, we found that the highly expressed genes were significantly enriched in CD4<sup>+</sup> T cells and CD8<sup>+</sup> T cells. These results indicate that enhancer-based immune dysregulation is likely to contribute to T1D pathogenesis. And these results suggest that our prioritized causal variants may help to diagnose T1D and can be therapeutic targets for patients.

## **PgmNr 3070: Evaluating the transdifferentiation of mesenchymal stem cells into insulin-secreting cells for transplantation to restore insulin secretion and regulation in companion animals.**

### **Authors:**

A. Wysong <sup>1</sup>; L. Stehno-Bittel <sup>2</sup>; L. Ott <sup>2</sup>; F. Karanu <sup>2</sup>; N. Kibiryeve <sup>3</sup>; M. Filla <sup>1</sup>; D. Bittel <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Kansas City University, Kansas City, MO.; 2) Likarda, LLC, Kansas City, MO; 3) Children's Mercy Hospital, Kansas City, MO

---

Type I diabetes is characterized by the autoimmune destruction of insulin-secreting  $\beta$ -cell in the pancreas. Current therapies for treating Type I diabetes are inadequate. We are developing protocols to derive insulin-producing cells directly from adipose stem cells without de-differentiation into embryonic-like cells. Preclinical studies are underway in companion animals to demonstrate these bioengineered insulin-producing cells can compensate for inadequate  $\beta$ -cell function prior to human application. Here, we analyze the differentiation process using biomarkers that represent key milestones as mesenchymal stem cells (MSC) differentiate into islet cells. MSCs are harvested from adipose tissue surrounding the reproductive organs. The MSCs are exposed to a 3-stage protocol incorporating growth factors that encourage differentiation into insulin-secreting cells. Using qRT-PCR, biomarkers including Sox17, Gata-6, PDX-1, and insulin, were used to assess the differentiation cycle. The cycle was divided into 3 stages based on the time point of added differentiation factors that initiate the differentiation of the MSCs into insulin-producing cells. Preliminary results indicate presence of Gata-6 and Sox17 in all stages of the differentiation. PDX1 was detected in the final stage of the differentiation, suggesting that the cells were progressing toward insulin-producing cells. These studies open the door to using engineered stem cells to produce an alternative therapy for restoration of appropriate glucose metabolism.

## PgmNr 3071: Chromatin accessibility and gene expression changes during adipocyte differentiation identifies regulatory elements at cardiometabolic trait GWAS loci.

### Authors:

H.J. Perrin<sup>1</sup>; K.W. Currin<sup>1</sup>; C.K. Raulerson<sup>1</sup>; S. Vadlamudi<sup>1</sup>; J.P. Davis<sup>1</sup>; M. Wabitsch<sup>2</sup>; P. Pajukanta<sup>3</sup>; M. Laakso<sup>4</sup>; K.L. Mohlke<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.; 2) University of Ulm, Ulm, Germany; 3) University of California, Los Angeles, Los Angeles, CA; 4) University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland

GWAS have identified thousands of cardiometabolic trait loci, but most lie in noncoding regions with mechanisms that remain poorly understood. At these loci, chromatin accessibility (CA) and gene expression in relevant cells and contexts can guide identification of regulatory elements and mechanisms. To identify elements that influence adipocyte differentiation, we performed ATAC-seq and RNA-seq in SGBS human adipocytes at days 0, 4, and 14 of differentiation (D0, D4, D14). We evaluated differential CA between timepoints (DESeq2, log2 fold change (LFC)>1, FDR<5%) and identified elements more accessible in adipocytes compared to preadipocytes (D4>D0=30,038 elements, D14>D0=20,158 elements). 90.6% of CA elements more accessible at D14 compared to D0 were also more accessible at D4 compared to D0, suggesting that CA changes occurred early and the majority remained stable. Of 17,250 expressed genes, we evaluated differential gene expression between timepoints (DESeq2, LFC>1, FDR<5%) and identified genes more expressed in adipocytes compared to preadipocytes (D4>D0=1,012 genes, D14>D0=1015 genes). 51.2% of genes more expressed at D14 compared to D0 were also more expressed at D4 compared to D0, suggesting that gene expression continues to change between timepoints. We evaluated differentially expressed genes in proximity (10,000 bases) to differential CA elements and identified 512 (50.4%) of genes more expressed at D14 compared to D0 near CA elements more accessible at D14 compared to D0. To identify differential regulatory elements at cardiometabolic GWAS loci, we examined loci with differential CA and a colocalized adipose eQTL. Of 14,938 candidate variants ( $r^2>0.8$ ), 370 variants at 130 loci were located in an element of differential CA. For example, at a GWAS locus for phospholipids colocalized with an adipose eQTL for *SCD* ( $P=8.4E-23$ ), we observed an adipocyte CA element (D4>D0 LFC=4.2, D14>D0 LFC=2.7) containing rs603424. *SCD* was also more expressed in adipocytes (D4>D0 LFC=5.4, D14>D0 LFC=7.5). This variant exhibited stronger allelic differences in transcriptional reporter activity in adipocytes (FC=3.1,  $P=0.003$ ) than in preadipocytes (FC=1.4,  $P=0.09$ ). These results demonstrate that epigenomic annotation in different cell contexts can guide discovery of regulatory variants at GWAS loci. Identification of CA elements and gene expression changes in other disease-relevant contexts could aid identification of additional regulatory mechanisms.



## PgmNr 3072: Adipose eQTL meta-analysis of 1,426 individuals identifies eQTLs for 512 genes colocalized with GWAS signals for cardiometabolic traits.

### Authors:

S.M. Brotman<sup>1</sup>; L. Guan<sup>2</sup>; J.S. El-Sayed Moustafa<sup>3</sup>; A.U. Jackson<sup>4</sup>; C.K. Raulerson<sup>1</sup>; A. Ko<sup>5</sup>; M.R. Erdos<sup>6</sup>; N. Narisu<sup>6</sup>; M. Laakso<sup>7</sup>; M. Boehnke<sup>4</sup>; P. Pajukanta<sup>8</sup>; F.S. Collins<sup>6</sup>; K.L. Mohlke<sup>1</sup>; L.J. Scott<sup>4</sup>; K.S. Small<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 2) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, USA; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London SE1 7EH, UK; 4) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI, USA; 5) Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 6) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA; 7) Institute of Clinical Medicine, Kuopio University Hospital, University of Eastern Finland, Kuopio, Finland; 8) Department of Human Genetics and Institute for Precision Health, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

---

Genome-wide association studies (GWAS) have discovered hundreds of loci associated with complex cardiometabolic traits; however, identifying causative genes at these loci has proven more difficult. Expression quantitative trait loci (eQTL) in relevant tissues, such as adipose, may help identify candidate genes for GWAS loci, if a GWAS signal and eQTL signal are colocalized. We conducted an eQTL meta-analysis with adipose tissue RNA-sequencing-based eQTL data from 3 studies: METSIM (n = 426 males), FUSION (n = 151 males; 131 females), and TwinsUK (n = 718 females). All studies retained genes with >5 counts in >25% of samples, analyzed normalized counts per million, adjusted for BMI and PEER factors, and tested variants with a MAF>0.01 in cis-eQTL (within  $\pm 1$  Mb from TSS) analyses. In a preliminary sample-size weighted meta-analysis using METAL, we identified 21,545 genes with at least 1 significant eQTL variant (FDR< 1%, equivalent  $p$ -value<4.9e-04). We tested the adipose eQTLs for colocalization with 3,038 GWAS signals from 70 cardiometabolic traits (clumped by linkage disequilibrium [LD]  $r^2 \geq 0.8$ ). GWAS and eQTL signals were considered colocalized if the lead variants were identical or in strong LD ( $r^2 \geq 0.8$ ) with each other. We found eQTLs for 512 genes that were colocalized with GWAS signals for one or more of 35 traits. Among these, we identified 48 eQTL signals colocalized with type 2 diabetes (T2D) GWAS loci (16.9% of T2D loci tested), including colocalization of recently-identified T2D loci with expression level of *CAMK2G*, *AOC1*, *DCAF16*, *CALR*, *PHBP9*, *MRPS30*, *NUAK2*, *FAM154A*, and *CITED2*. For example, *NUAK2* ( $r^2 \geq 0.91$ ; eQTL  $p$ -value 1.36e-11), also known as *SNARK*, is a kinase that exhibits decreased adipose tissue expression in obesity and influences inflammation and lipid deposition in liver and muscle. Further analysis of additional eQTL signals for each gene will likely identify additional colocalized eQTL signals. These preliminary results demonstrate that a larger eQTL sample size from meta-analysis improves our ability to detect adipose eQTLs and GWAS candidate genes.



## PgmNr 3073: Exploring the utility of activity QTLs in identifying genes mediating complex traits.

### Authors:

J. Hoskins<sup>1</sup>; J. Shi<sup>2</sup>; A. Jermusyk<sup>1</sup>; J. Zhong<sup>1</sup>; K. Conelly<sup>1</sup>; L. Amudadottir<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Translational Genomics, DCEG, National Cancer Institute, Gaithersburg, MD; 2) Biostatistics Branch, DCEG, National Cancer Institute, Rockville, MD

---

Genome-wide association studies (GWAS) have uncovered a wealth of genetic signals affecting complex traits. Most GWAS variants are non-coding, suggesting effects on gene expression without indicating which gene(s) are affected. Expression quantitative trait loci (eQTL) analysis is the gold standard for identifying the gene(s) mediating the effects of GWAS variants. However, multiple studies comparing GWAS and eQTL signals have found co-localization of typically ~50% of GWAS signals with any-tissue eQTLs and a much smaller proportion among single-tissue eQTLs. As a proof-of-concept we have explored the use of activity QTLs (aQTLs) in identifying genes mediating BMI GWAS signals using the Eurobats adipose eQTL dataset. Activities of expression regulators were inferred from the expression of downstream target genes using VIPER. Analyses of 2,920 regulator genes within 1Mb of 41,103 BMI GWAS variants identified 565 eGenes with 1,405 *cis*-eQTLs and 72 aGenes with 335 *cis*-aQTLs at FDR<0.05. The *cis*-eQTL associations were typically stronger than the corresponding aQTLs, but there were some interesting exceptions. For example, the BMI GWAS signal at 1p36.12 had *EPHB2* *cis*-aQTL *P*-values up to three orders of magnitude better than the corresponding *cis*-eQTLs. We speculated that aQTLs may hold more advantage in *trans* analyses, particularly when looking at master regulators (MRs) that integrate upstream genetic and environmental information into a gene regulatory program instantiating a cell state of interest. MRs of the BMI-associated adipose cell state were inferred by cross-validated Random Forest feature selection based on the adipose samples' regulator activities. This yielded 100 BMI MRs that were validated in GTEx subcutaneous adipose. *Trans* analyses between BMI GWAS variants and these BMI MRs identified 6,287 and 8,996 significant *trans*-eQTLs and *trans*-aQTLs, respectively, and showed all 100 MRs were *trans*-eGenes and aGenes at FDR<0.05. Among these *trans*-QTLs, 60% showed stronger aQTL versus eQTL association. Strikingly, the 1p36.12 BMI signal had significant *trans*-aQTLs with 67/100 BMI MRs while only 8 had significant *trans*-eQTLs. Furthermore, the *EPHB2* *cis*-aQTL signal co-localized with the MR *trans*-aQTLs, suggesting this BMI GWAS signal may be mediated through effects on *EPHB2* activity that subsequently affect BMI MR activities. In conclusion, aQTL analysis may provide a useful method to supplement eQTLs in linking GWAS signals to their mediating genes.

## PgmNr 3074: Functional characterization of *NFIA-AS2* in human obesity.

### Authors:

K. Bandesh; Y.L. Muller; M. Traurig; P. Piaggi; P. Kumar; L.J. Baier

[View Session](#) [Add to Schedule](#)

**Affiliation:** Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 445 N 5th Street, Phoenix, AZ, 85004, USA.

---

Obesity which is estimated by body mass index (BMI), primes several chronic diseases and is a global health concern. Its common forms are highly polygenic and environmentally influenced; so, studies in populations with high prevalence of obesity, and minimal genetic admixture and environmental variability may aid in the identification of genetic markers. Pima Indians have these features, and many Pimas from a community in Arizona participated in longitudinal exams providing height and weight measures to calculate BMI. In this study, we used genome-wide imputed data (whole genome sequence data from Pimas as reference and genotypes in 7,701 Pimas from a custom array as scaffold) to identify all common SNPs that nominally associate ( $p < 10^{-4}$ ) with maximum BMI recorded after 18 years of age in 6,800 Pimas. Such 6,983 SNPs were meta-analyzed with reported BMI associations in GIANT ( $N = 253,288$ ) and GIANT-UKBB ( $N = 795,640$ ) datasets. On top of the known BMI genes- *FTO*, *TMEM18*, *AKAP6*, *NLRC3*, *KCNQ1* and *GRP*, we identified strong association with variants residing in an uncharacterized long non-coding RNA gene- *NFIA-AS2* ( $p = 10^{-??}$ ). This locus has never been implicated in obesity and its variants had fourfold higher frequency in Pima Indians as compared to Caucasians ( $MAF = 0.16$  and  $0.04$ , resp.) with a large effect in Pimas e.g., each copy of minor allele for exonic variant rs1777538 resulted in a decrease of  $0.8 \text{ kg/m}^2$  in BMI. This SNP was observed *in silico* to be associated with *NFIA-AS2* expression in human subcutaneous adipose tissue ( $p = 0.01$ ). Thus, we studied expression of 12 genes flanking *NFIA-AS2* in adipose tissue biopsies of 210 non-diabetic Pimas and found a strong inverse correlation of mRNA levels of *NFIA* with BMI ( $p = 0.004$ ). *NFIA* (Nuclear Factor IA) is required for differentiation of brown adipocytes. Analysis of *NFIA-AS2*, *NFIA* and neighboring genes in human brown pre-adipocyte cell-line PAZ6 revealed significant changes in *NFIA* as well as *NFIA-AS2* expression upon differentiation ( $p_{(\text{adipocyte stage})} = 0.0021$ ;  $0.00022$  resp.). Through cellular fractionation of PAZ6 cells, we learnt that *NFIA-AS2* localizes to the nucleus which implies that it regulates gene expression via epigenetic processes. We are now studying how perturbation of lncRNA expression affects brown adipocyte differentiation and *NFIA* levels. Alleles of variant rs1777538 were predicted to produce a local change in secondary structure of *NFIA-AS2* lncRNA; experiments to validate this change are ongoing.

## PgmNr 3075: In vivo models for Qatari-specific classical homocystinuria as platforms for development of novel treatment.

**Authors:**

G. Nasrallah

[View Session](#) [Add to Schedule](#)**Affiliation:** Biomedical Science, College of Health Sciences, Qatar University, Doha, Qatar

---

Homocystinuria is a rare inborn error of methionine metabolism caused by cystathionine  $\beta$ -synthase (CBS) deficiency. The prevalence of homocystinuria in Qatar is 1:1,800 births, mainly due to a founder Qatari missense mutation, c.1006C>T; p.R336C (p.Arg336Cys). We characterized the structure-function relationship of the p.R336C-mutant protein and investigated the effect of different chemical chaperones to restore p.R336C-CBS activity using three models: in silico,  $\Delta$ CBS yeast, and CRISPR/Cas9 p.R336C knock-in HEK293T and HepG2 cell lines. Protein modeling suggested that the p.R336C induces severe conformational and structural changes, perhaps influencing CBS activity. Wild-type CBS, but not the p.R336C mutant, was able to restore the yeast growth in  $\Delta$ CBS-deficient yeast in a complementation assay. The p.R336C knock-in HEK293T and HepG2 cells decreased the level of CBS expression and reduced its structural stability; however, treatment of the p.R336C knock-in HEK293T cells with betaine, a chemical chaperone, restored the stability and tetrameric conformation of CBS, but not its activity. Collectively, these results indicate that the p.R336C mutation has a deleterious effect on CBS structure, stability, and activity, and using the chemical chaperones approach for treatment could be ineffective in restoring p.R336C CBS activity.

## PgmNr 3076: Multi-omics evaluation of a patient-derived HNF-1A MODY iPSC disease model.

### Authors:

A. Wesolowska-Andersen <sup>1</sup>; R. Rejnholdt Jensen <sup>2</sup>; A. Grotz <sup>3</sup>; V. Nylander <sup>3</sup>; M. Jansen <sup>1</sup>; M. Hansson <sup>5</sup>; M.I. McCarthy <sup>1,3,4</sup>; A.L. Gloyn <sup>1,3,4</sup>; C. Honore <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; 2) Dept of Stem Cell Biology, Novo Nordisk A/S, Måløv, Denmark; 3) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 4) NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK; 5) Dept of Stem Cell Research, Novo Nordisk A/S, Måløv, Denmark

Maturity Onset Diabetes of the Young (MODY) due to *HNF1A* mutations is the most common cause of monogenic diabetes. It results from heterozygous, loss-of-function mutations in the transcription factor HNF-1A, which plays a critical role in pancreatic development. Mice with heterozygous *Hnf1a* mutations do not recapitulate the diabetic phenotype, therefore human iPSC-derived cells are more suitable for disease modelling.

We established a patient-derived *HNF1A* Pro291fsinsC/+ iPSC model, and using CRISPR-Cas9 genome-editing corrected the mutation to establish isogenic controls. We differentiated multiple (n=6) iPSC clones towards endocrine lineage in triplicates, and performed RNA-seq (bulk and single-cell), ATAC-seq and methylation assays at definitive endoderm (DE), pancreatic endoderm (PE) and beta-like cells (BLC) stages.

No expression of the MODY allele in the RNA-seq data confirmed *HNF1A* haploinsufficiency in the patient lines. We found that *HNF1A* expression was restored at both transcript and protein level in the corrected lines. Genes differentially expressed at BLC stage were significantly enriched in targets of HNF-1A identified in an independent *HNF1A* knock-down experiment in EndoC-βH1 cell line (p=4.4e-132). Consistent with reported effects of HNF-1A on insulin secretion, we observed downregulation of *INS* (p=7.2e-03) and other insulin secretion genes in the MODY clones. At the PE stage, corrected clones showed higher expression of *SOX9* (p=7.1e-04) and *PTF1A* (p=1.5e-03), known regulators of pancreatic progenitors, suggesting that HNF-1A is involved in establishment of this cell population. These findings were replicated in the single-cell RNA-seq data, which additionally highlighted differences in cell populations between the MODY and corrected clones. We also identified 7811 sites at PE and 1710 sites at BLC with significant changes in open chromatin, with >90% of the sites being more accessible in the corrected clones. These sites were enriched for HNF-1A binding motifs, suggesting its direct role in establishing open chromatin.

In summary, the patient-derived HNF1A-MODY iPSC model provides a valuable resource for studying of the transcriptional and epigenetic rewiring during pancreatic development caused by the Pro291fsinsC/+ mutation, and yields insights into the cellular and molecular phenotypes caused by HNF-1A deficiency.

This work has received support from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking

(StemBANCC grant°115439).

## PgmNr 3077: Subcutaneous adipose tissue eQTLs identify hundreds of candidate genes at cardiometabolic GWAS loci.

### Authors:

C. Raulerson<sup>1</sup>; A. Ko<sup>2,3</sup>; J. Kidd<sup>4</sup>; K. Currin<sup>1</sup>; S. Brotman<sup>1</sup>; M. Cannon<sup>1</sup>; Y. Wu<sup>1</sup>; C. Spracklen<sup>1</sup>; A. Jackson<sup>5</sup>; H. Stringham<sup>5</sup>; R. Welch<sup>5</sup>; C. Fuchsberger<sup>6</sup>; A. Locke<sup>7</sup>; N. Narisu<sup>8</sup>; A. Lusi<sup>2</sup>; M. Civelek<sup>9</sup>; T. Furey<sup>1,10</sup>; J. Kuusisto<sup>11</sup>; F. Collins<sup>8</sup>; M. Boehnke<sup>5</sup>; L. Scott<sup>5</sup>; D. Lin<sup>4</sup>; M. Love<sup>1,4</sup>; M. Laakso<sup>11</sup>; P. Pajukanta<sup>2,12</sup>; K. Mohlke<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 3) Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; 4) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 5) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI; 6) Center for Biomedicine, European Academy of Bolzano/Bozen, University of Lübeck, Bolzano/Bozen, Italy; 7) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO; 8) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 9) Department of Biomedical Engineering, University of Virginia, Charlottesville, VA; 10) Department of Biology, University of North Carolina, Chapel Hill, NC; 11) Institute of Clinical Medicine, Kuopio University Hospital, University of Eastern Finland, Kuopio, Finland; 12) Institute for Precision Health, David Geffen School of Medicine at UCLA, Los Angeles, CA

Genome-wide association studies (GWAS) have identified thousands of genetic signals associated with cardiometabolic traits including type 2 diabetes (T2D), circulating lipid levels, and body fat distribution. However, the underlying genes remain unknown. To identify variants that may influence the expression of nearby genes, we used genotypes and RNA-sequencing of subcutaneous adipose tissue from 434 Finns from the METabolic Syndrome in Men (METSIM) study. We tested for association of 7.8M variants with expression levels of 21,735 genes using FastQTL and identified 9,687 genes associated with  $\geq 1$  variant within  $\pm 1$  Mb of the TSS (eQTLs, FDR<1%). To identify additional signals, we conditioned on the variant most strongly associated with each gene (eSNP) and identified 2,785 genes with secondary eQTLs. Compared to primary eQTL signals, secondary signals were located further from transcription start sites, had smaller effect sizes, and were less enriched in adipose tissue regulatory elements. We used linkage disequilibrium (LD) between the GWAS and eQTL variants and conditional analysis to assess colocalization. Among 2,843 cardiometabolic GWAS signals, 262 colocalize with primary or secondary eQTLs for 318 transcripts. Of cardiometabolic traits examined, waist-hip ratio adjusted for BMI (WHRadjBMI) and circulating lipid traits showed the highest percentage of colocalized eQTLs (12% and 11%, respectively). Among alleles associated with increased cardiometabolic GWAS risk, approximately half (47%) were associated with increased gene expression level. For 21 (66.7%) GWAS signals colocalized as secondary eQTLs, no primary eQTL was colocalized with the GWAS variant, thus no colocalization would have been identified without the secondary eQTL analysis. For example, the primary eQTL signal for *DGKQ* does not colocalize with the WHRadjBMI GWAS variant, rs11724804 (LD  $r^2=0.26$ ), but the secondary eQTL variant, rs13101828 ( $P_{\text{initial}}=2.9 \times 10^{-8}$ ), does colocalize ( $r^2=1.0$ ;  $P_{\text{conditional}}=0.67$ ). Silencing *DGKQ* in hepatic cells has been shown to alter mTOR signaling, and misregulation of the mTOR pathway has been shown to promote

adipogenesis and lipogenesis in adipose tissue. Taken together, these results demonstrate that GWAS-adipose eQTL colocalization can identify candidate genes for cardiometabolic trait loci.

## PgmNr 3078: Analysis of alterations in gene expression related to chronic obstructive pulmonary disease.

**Authors:**

H.M. Choi; D.Y. Kim; W.J. Kim; J.H. Kim; S.H. Hong; S.S. Choi; S.H. Lee; M.S. Lee

[View Session](#) [Add to Schedule](#)

**Affiliation:** Kangwon National University, Chuncheon-si, Kangwon-do, Korea, Republic of

---

We identified COPD-related differentially expressed genes (DEGs) using 189 samples accompanying either adenocarcinoma (AC) or squamous cell carcinoma (SC), comprising 91 normal and 98 COPD samples. DEGs were obtained from the intersection of two DEG sets separately identified for AC and SC to exclude the influence of different cancer backgrounds co-occurring with COPD. We also measured patient samples named group 'I', which were unable to be determined as normal or COPD based on alterations in gene expression. The Gene Ontology (GO) analysis revealed significant alterations in the expression of genes categorized with the 'cell adhesion', 'inflammatory response', and 'mitochondrial functions', i.e., well-known functions related to COPD, in samples from patients with COPD. Multi-omics data were subsequently integrated to decipher the upstream regulatory changes linked to the gene expression alterations in COPD. COPD-associated expression quantitative trait loci (eQTLs) were located at the upstream regulatory regions of 96 DEGs. Additionally, 45 previously identified COPD-related miRNAs were predicted to target 66 of the DEGs. The eQTLs and miRNAs might affect the expression of 'respiratory electron transport chain' genes and 'cell proliferation' genes, respectively, while both eQTLs and miRNAs might affect the expression of 'apoptosis' genes. We think that our present study will contribute to our understanding of the molecular etiology of COPD accompanying lung cancer.



## PgmNr 3079: Aberrant biogenesis and trafficking of secretory proteins is a common pathogenetic mechanism of autosomal dominant tubulointerstitial kidney disease (ADTKD).

### Authors:

M. Zivna<sup>1</sup>; H. Hartmannova<sup>1</sup>; K. Hodanova<sup>1</sup>; P. Vyletal<sup>1</sup>; V. Baresova<sup>1</sup>; H. Hulkova<sup>1</sup>; J. Sovova<sup>1</sup>; I. Jedlickova<sup>1</sup>; L. Piherova<sup>1</sup>; A.J. Bleyer<sup>1,2</sup>; K. Kidd<sup>2</sup>; S. Kmoch<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Research Unit for Rare Diseases, Department of Pediatrics and Adolescent Medicine, First Faculty of Medicine, Charles University, Prague, Czech Republic; 2) Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC, USA

---

Autosomal dominant tubulo-interstitial kidney disease (ADTKD) encompasses a group of genetic disorders characterized by renal tubular and interstitial abnormalities, leading to slow progressive loss of kidney function requiring dialysis and kidney transplantation. This condition has been identified in almost 1000 families world-wide. Genetic studies revealed causal mutations in *UMOD*, *REN*, *MUC1* and *SEC61A1* encoding uromodulin, renin, mucin-1 and translocon subunit SEC61A, respectively.

Uromodulin, renin and mucin-1 are glycoproteins that are prominently expressed in kidney.

Uromodulin and mucin-1 are targeted to tubular cell membrane and released into urine, whereas renin is actively secreted into circulation. Biosynthesis of these three proteins is dependent on proper function of the translocon.

Disease-causing mutation of uromodulin, renin, mucin-1 and *SEC61A1* affect biogenesis and trafficking of the corresponding proteins. This leads to intracellular accumulation of mutated proteins along the secretory pathway and to various forms of cellular responses such as organelle stress or unfolded protein response in affected nephron structures. In the long term the sustained cellular responses initiate processes that lead to deterioration of kidney function and development of ADTKD.

## PgmNr 3080: Blood metabolome diversity in heterozygous adult carriers of pathogenic variants of newborn screening diseases.

### Authors:

Y. Yamaguchi-Kabata; S. Koshiba; I. Motoike; D. Saigusa; J. Inoue; M. Shirota; Y. Aoki; S. Tadaka; S. Kure; Y. Yamamoto; K. Kinoshita

[View Session](#) [Add to Schedule](#)

**Affiliation:** Tohoku Medical Megabank Organization, Tohoku University, Aoba-ku, Sendai, Japan

---

Although heterozygous carriers of causative variants for recessive disorders are not expected to develop the diseases, molecular phenotypes of the variant carriers have not been clarified well. Here we examine blood metabolite levels as a molecular phenotype in variant carriers in adults in general Japanese residents for 16 congenital metabolic disorders for primary targets of newborn screening in Japan. Utilizing whole genome reference panel of 3552 general Japanese individuals (3.5KJPNv2), we identified pathogenic variants for the 16 diseases through variant annotation and interpretation. The number of individuals having at least one of the selected pathogenic variants were 211 (5.9%) based on ACMG-AMP guideline (2015), 319 (9.0%) with additional ClinVar variants, and 605 (17.0%) with additional HGMD disease-causing variants. Then we analyzed metabolite levels of the variant carriers in the cohort participants. One example is phenylalanine hydroxylase (PAH), which is involved in the conversion of phenylalanine to tyrosine and functional defects cause hyperphenylalaninemia (HPA) and phenylketonuria (PKU). A missense variant, p.Arg413Pro existed at the allele frequency of 0.00099, which was reported as frequently found among PKU/HPA patients in Japan. Our analysis with blood metabolite showed that this missense variant, p.Arg413Pro increased the levels of phenylalanine in heterozygous individuals. Another example is methylenetetrahydrofolate reductase (MTHFR), which is involved in folate-mediated one-carbon metabolism. Deficiency of MTHFR increases homocysteine levels and causes hyperhomocystinemia with homocystinuria, mild hyperhomocystinemia, and is also involved in other diseases, such as vascular disease, neural tube defects, cancer, and developmental delay. In 3.5KJPNv2, a missense variant p.Ala222Val (rs1801133, c.665C>T) was found in 1674 individuals as heterozygous and 522 individuals as homozygous (allele frequency = 0.383) as well as low-frequency reported pathogenic variants. Our analysis with blood metabolite showed that this variant is associated with concentration of plasma formate. Our results suggest that common and rare variants in causative genes Mendelian diseases affect molecular phenotypes in a human body in heterozygous and these metabolome individualities with possible effects on disease susceptibilities can be monitored by metabolome analysis.

## PgmNr 3081: Functional assessment of 38 intronic and 19 exonic putative splice variants in *CFTR* to interpret pathogenicity and inform treatment.

### Authors:

A. Joynt; K. Paul; D. Osorio; M.A. Aksit; K. Raraigh; N. West; C. Merlo; N. Sharma; G.R. Cutting

[View Session](#) [Add to Schedule](#)

**Affiliation:** Johns Hopkins Medicine, Baltimore, MD

Genetic variants that cause missplicing are often thought to result in no protein production. In fact, some of these variants preserve a fraction of normal splicing thus producing reduced levels of functional protein. On the other hand, exonic variants that alter amino acids are thought to cause disease by reducing protein function, but may instead alter splicing. To identify *CFTR* variants in these two categories, we evaluated 57 variants that occurred in extended consensus splice signals or were predicted to introduce cryptic splicing by *in silico* algorithms (CryptSplice, NNSplice). Using expression mini-genes (EMGs) containing *CFTR* full-length or abridged introns, we evaluated the functional consequences of each variant (intronic=38, exonic=19).

Of the 38 intronic variants, 35 resulted in missplicing while 3 spliced normally. Importantly, 8 variants generated reduced levels of normally spliced transcript indicating that they should allow synthesis of full-length *CFTR*. Of the 19 exonic variants, 9 misspliced completely, 2 generated reduced levels and 8 generated wild-type levels of normally spliced transcript. We assessed protein by immunoblotting for 41 variants (intronic=34, exonic=7) and found 7 variants that produced no protein, 25 with shortened protein, and 9 that produced full-length protein.

We hypothesized that variants producing shortened or full-length protein would respond to FDA-approved *CFTR* modulators. We created cell lines expressing EMGs for 11 variants and found that 6 variants that allowed synthesis of reduced levels of full-length protein responded to modulators, while 5 variants that produced shortened protein did not. While improvement of modulators leaves the possibility of protein-level treatment of these variants, splice variants that produced no protein will require alternative treatments.

For two exonic variants, we evaluated the effect of the missense change using cells expressing *CFTR* cDNA. In both cases, we found that the amino acid substitution had little detriment to *CFTR* function and responded to modulators. This result indicates that treatments focused on correction of missplicing could allow for production of functional protein, but also reveals the potential for false categorization of exonic variants as modulator responsive. This study demonstrates the importance of characterizing the functional consequences of both intronic and exonic variants at the RNA level in order to select appropriate molecular treatments.

## PgmNr 3082: Functional characterization of tumor-associated, germline *TMEM127* variants reveals novel insights into membrane topology and protein trafficking.

### Authors:

S.K. Flores; Y. Deng; X. Zhang; I. Chu; Z. Cheng; Y. Qin; P.L.M. Dahia

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medicine, Division of Hematology/Oncology, UT Health San Antonio, San Antonio, Texas

---

TMEM127, identified as a tumor suppressor in highly hereditary neuroendocrine tumors known as pheochromocytomas (PHEOs), is a poorly characterized transmembrane (TM) protein with no substantial identity to any other protein and no identifiable functional domains other than its three putative TM domains. The function and regulation of TMEM127 remain poorly defined which limits prediction of pathogenicity of germline *TMEM127* variants. Characterizing structure-function features of TMEM127 is relevant for understanding how dysfunction of this tumor suppressor can lead to inherited tumors. Moreover, a better understanding of TMEM127 functional domains and critical residues will increase our ability to identify pathogenic variants which will improve interpretation of genetic screening results and help guide clinical decisions for patients and their families. In this study, we used tagged TMEM127 plasmid constructs based on patient-derived, tumor-associated, germline *TMEM127* variants (n=21; 16 missense and 5 truncating/frameshift/indel variants) and investigated their subcellular localization using confocal microscopy and steady-state levels using immunoblot analysis. Wild-type (WT) TMEM127 localized predominantly to endo-lysosomal vesicles (punctate appearance) with some plasma membrane localization. The localization of variant TMEM127 proteins separated into three distinct categories: punctate (similar to WT), diffuse (cytoplasmic), and plasma membrane. All diffuse proteins, and a few punctate proteins, decreased at a faster rate than the WT suggesting instability. Variants resulting in diffuse proteins occurred within TM domains and indicated that membrane binding ability was lost. This observation, supported by *in silico* analysis and additional *in vitro* assays, led us to conclude that TMEM127 is a four-TM protein, not a three-TM protein, as previously predicted, with a novel TM domain in the N-terminus. The variant that predominantly localized to the plasma membrane indicated that membrane binding ability was maintained but internalization capability was lost. By mutagenesis scanning, we identified an atypical, extended acidic dileucine motif in the C-terminus which is responsible for TMEM127 internalization through clathrin-mediated endocytosis. Our findings provide novel insights into structure-function features of TMEM127 which will allow for better understanding of its physiological role and improved prediction of pathogenic *TMEM127* variants.

## PgmNr 3083: Single-cell RNA sequencing identifies a novel profibrotic epithelial cell population in primary human lungs from patients with pulmonary fibrosis.

### Authors:

M. Chung <sup>1</sup>; A. Habermann <sup>2</sup>; A. Gutierrez <sup>1</sup>; L. Bui <sup>1</sup>; N. Winters <sup>2</sup>; C. Calvi <sup>2</sup>; L. Peter <sup>1</sup>; C. Taylor <sup>2</sup>; S. Yahn <sup>1</sup>; T. Blackwell <sup>2,3,4</sup>; J. Kropski <sup>2,3,4</sup>; N. Banovich <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Translational Genomics Research Institute, Phoenix, AZ; 2) Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Veterans Affairs Medical Center, Nashville, TN; 4) Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN

---

Pulmonary fibrosis (PF) is a chronic interstitial lung disease (ILD) in which lung epithelium is progressively replaced by fibrotic scarring tissue leading to respiratory failure. The most common and aggressive form of PF is idiopathic pulmonary fibrosis (IPF), yet IPF accounts for only 20% of all PF cases. Available therapeutic approaches to this irreversible disease only slow down the fibrotic progression and lung transplantation is often required for late stage patients. To date, the precise mechanisms that drive fibrosis remain incompletely understood. To better understand pulmonary remodeling during the fibrotic progression, we performed single-cell RNA sequencing (scRNA-seq) on lung tissues collected from 18 healthy lungs (declined donors) and 29 lung explants from patients with a diverse set of PF diagnosis to obtain a comprehensive census of cell identities. Jointly analyzing cells from both healthy and diseased samples reveals at least 32 diverse pulmonary cells covering heterogeneity of epithelial cells, mesenchymal cells, endothelial cells, and immune cells. For examples, *TP63*<sup>+</sup>, *KRT17*<sup>+</sup>, *KRT5*<sup>+</sup> basal cells, *FOXJ1*<sup>+</sup> ciliated cells, and *ABCA3*<sup>+</sup> alveolar type 2 cells were shown and consistent with previously addressed cell characteristics. Interestingly, we identified a novel epithelial cell population which co-expresses *TP63* and *KRT17*, but not *KRT5* (*KRT17*<sup>+</sup>*KRT5*<sup>-</sup>). Moreover, a large number of genes associated with profibrotic processes (e.g. *COL1A1*, *VIM*, and *FN1*) were detected. The *KRT17*<sup>+</sup>*KRT5*<sup>-</sup> cells were almost exclusively found in PF lungs. Finally, this population appears to be the unique population in the lung expressing integrin  $\alpha\text{v}\beta_6$ , a heterodimer that is being actively investigated as a target for antifibrotic therapies. Using RNA scope, we validated the expression and localized the *KRT17*<sup>+</sup>*KRT5*<sup>-</sup> cells to be next to collagen-enriched fibroblasts, generally apart from *SFTPC*<sup>+</sup> epithelial cells. We are currently carrying out in vitro studies to functionally assess the impact of *KRT17*<sup>+</sup>*KRT5*<sup>-</sup> cells, their interactions with fibroblasts, and their biological significance in lung disease.

## **PgmNr 3084: Single-cell RNA-sequencing of primary human lung reveals disease-specific cell types and cell type-specific molecular mechanisms associated with disease progression.**

### **Authors:**

S.L. Yahn<sup>1</sup>; A.C. Habermann<sup>2</sup>; A. Gutierrez<sup>1</sup>; L. Bui<sup>1</sup>; N.I. Winters<sup>2</sup>; C.L. Calvi<sup>2</sup>; L. Peter<sup>1</sup>; M. Chung<sup>1</sup>; C.J. Taylor<sup>2</sup>; T.S. Blackwell<sup>2,3,4</sup>; J.A. Kropski<sup>2,3,4</sup>; N.E. Banovich<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Translational Genomics Research Institute, Phoenix, AZ; 2) Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Veterans Affairs Medical Center, Nashville, TN; 4) Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN

---

Pulmonary fibrosis (PF) is a heterogeneous clinical syndrome that represents the end-stage of chronic interstitial lung diseases (ILD). The most common (20%) and severe form of ILD is idiopathic pulmonary fibrosis (IPF), which typically leads to respiratory failure within 5 years of diagnosis. While current therapies for IPF can slow disease progression, they have not been shown to improve survival or quality of life, and they are not effective for the 80% of ILD patients with other PF diagnoses. To better understand the cell types and molecular mechanisms involved in the initiation of ILD and its progression to PF, we performed single-cell RNA-sequencing (scRNA-seq) of 67 samples from explant lungs of patients with IPF (n=17) or from healthy controls (n=21), as well as from patients diagnosed with other forms of ILD (n=12) including sarcoidosis, nonspecific interstitial pneumonia (NSIP), and chronic hypersensitivity pneumonitis (cHP). To further harness the natural pathology of the disease, we sampled two sites from a majority of the diseased lungs: 1) more fibrotic (n=28) and 2) less fibrotic (n=18). Using graph-based clustering, we identified over 30 distinct clusters representing the major known subtypes in the lung and found an enrichment of fibroblasts, macrophages, and T-cells in diseased lungs compared to controls. Interestingly, we observed cell states specific to diseased lungs including a previously undescribed population of profibrotic epithelial cells and a distinct fibroblast population specific to IPF. Furthermore, we identified hundreds of differentially expressed genes in most cell types when comparing diseased to control lungs, or within diseased state (less fibrotic samples vs more fibrotic samples or controls), some of which exhibited previously uncharacterized cell type specificity. By combining these results with trajectory inference, we were able to identify genes that are potentially driving disease initiation and progression. Finally, we identified cell type-specific expression quantitative trait loci (eQTLs) in diseased and healthy samples. To our knowledge, this is the largest single cell transcriptomic analysis of primary human lung tissue from patients with multiple forms of ILD and the first scRNA-seq eQTL analysis of diseased samples from primary human tissue.

## **PgmNr 3085: *WNT1*, causing autosomal recessive osteogenesis imperfecta, is required to sustain functional skeletal stem cells.**

### **Authors:**

N. Haelterman; K. Lei; D. Park; B. Lee

[View Session](#) [Add to Schedule](#)

**Affiliation:** Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

---

Mutations in *WNT1* cause an autosomal dominant form of early onset osteoporosis as well as an autosomal recessive form of the brittle bone disorder Osteogenesis Imperfecta (OI), identifying the molecule as an important regulator of human bone homeostasis. Similarly, loss of *WNT1* leads to severe bone loss in mice, whereas increasing *WNT1*-expression in osteoblasts strongly increases bone density in mice. The identity of the cell types that are responding to *WNT1*, and the mechanism by which alterations in *WNT1* levels induce such significant changes in bone density remains unclear. Here, we tested the hypothesis that *WNT1* regulates the maintenance and/or function of the skeletal stem cells that build and sustain bone throughout life. To this end, we first compared the prevalence of Paired-related homeobox gene-1 (PRX1)-expressing progenitor cells between *WNT1* mutants and their wildtype littermates. PRX1 is expressed in a population of postnatal skeletal stem cells that resides in the sutures of the calvaria and is necessary for calvarial bone regeneration upon injury. Surprisingly, we found a significant increase in the population of PRX1<sup>+</sup> progenitor cells, residing in *WNT1* mutant calvarial sutures. In addition, this pool of progenitor cells persists in 4-month-old mutant mice, whereas they have largely disappeared in wildtype animals by this age. We next tested the ability of *WNT1* mutant PRX1<sup>+</sup> progenitor cells to regenerate calvarial injuries and found that although mutant cells are able to migrate to the defect, bone regeneration is significantly delayed upon loss of *WNT1*. We are currently assessing the expression profile as well as the differentiation potential of *WNT1* mutant PRX1<sup>+</sup> stem cells in order to clarify how *WNT1* regulates this pool of progenitor cells. In summary, we find that loss of *WNT1* increases the pool of PRX1<sup>+</sup> skeletal progenitor cells in the skull, but that these cells are unable to perform their function in bone repair.



## PgmNr 3086: Integrating genomic fine-mapping and Hi-C to dissect causal variants and regulatory target genes underlying osteoporosis GWAS association.

### Authors:

Chen; F. Jiang; R.H. Hao; S.S. Dong; M.R. Guo; T.L. Yang

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, China

---

Genome-wide association studies (GWASs) have successfully identified thousands of independent genetic variants conferring osteoporosis risk, while the underlying regulatory mechanisms are largely unknown. To fulfil this gap, we firstly performed genomic fine-mapping analysis implemented by FINEMAP using GWAS summary data on bone mineral density at 6 different sites (forearm, femoral neck, lumbar spine, heel, total body and total-body less head) and fracture from the GEFOS portal. We prioritized 21,188 high-confident causal variants at 1,523 conditionally independent signals ( $\log_{10}BF \geq 3$ ), of which only 1.5% ones are coding or splicing, indicating the prevalent regulatory roles for osteoporosis risk variants. We therefore leveraged paired high-throughput chromosome interaction data (Hi-C, capture Hi-C, HiChIP, ChIA-PET) and cis-eQTL data across three bone-related cell types (osteoblast, monocyte and lymphocytes) to identify the potential regulatory target genes. We performed Hi-C analysis in primary osteoblasts induced from human mesenchymal stem cells at 2-KB resolution. We identified 72 distal regulated gene targets with both cis-QTL and chromatin interaction evidence in the same cell type as well as 49 local cis-regulated genes (SNPs within target gene promoter). Notably, 75.8% of target genes could be regulated by distal causal SNPs, indicating the critical regulatory roles of chromatin looping in osteoporosis. We further employed the Coloc and PICCOLO method to assess whether GWAS signal and eQTL association shared the same causal variant, and detected strong colocalization evidence ( $PP4 > 0.7$ ) for 58 gene targets. These genes included some well-characterized bone-related genes (*TNFSF11*, *NFATC1*, *MAFB*) and several other genes whose KO in mouse model could lead to abnormal bone phenotypes (*ATPAF2*, *GPATCH1*), as well as 17 genes with nominal significant ( $P < 0.05$ ) expression difference between low and high BMD samples in either osteocytes (E-MEXP-1618), monocytes (GSE2208) or B lymphocyte (GSE7429). Moreover, the summary data-based Mendelian randomization (SMR) analysis suggested that 25/33 measurable genes had causal regulatory roles for osteoporosis in either osteoblast or lymphocytes ( $P_{SMR} < 0.05$ ,  $P_{HEIDI} > 0.05$ ). In summary, we systematically explored the regulatory effects linking osteoporosis risk casual variants to distal target genes and further evaluated their functions. Our analysis results might facilitate the future experimental follow-up.



## PgmNr 3087: Validation and classification of atypical splicing variants associate with osteogenesis imperfecta.

### Authors:

Y. Cao; L. Li; F. Zhao; B. Mao; Y. You; S. Li; T. Yang; X. Zhao

[View Session](#) [Add to Schedule](#)

**Affiliation:** Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Beijing, China

---

Osteogenesis Imperfecta (OI) is a rare inherited bone dysplasia, which is mainly caused by mutations in genes encoding type I collagen including *COL1A1* and *COL1A2*. Typical mutational spectrum of OI includes missense, nonsense, frameshift and splice site mutations. However, functional examination of whether the atypical splicing variants that locate beyond splice sites affect splicing in OI patients is limited. Here we collected blood samples from 34 patients carrying the variants that located close to splice donor/acceptor sites in either *COL1A1* or *COL1A2* from a cohort of 867 OI patients. Whole exome sequencing combined with Sanger sequencing were performed to detect and verify the variants, followed by minigene splicing assay to investigate the splicing effect. The results of minigene assay showed that 19 of 34 variants led to aberrant splicing effects, while 15 without any remarkable aberrant splicing effect. There were 21 variants located in introns and 17 of them led to aberrant splicing and 4 with no aberrant splicing. Three splicing effects can be categorized: (i) variants that led to exon skipping (n=9); (ii) truncated exon (n=2); and (iii) intron retention (n=4). The rest of 13 variants located in exons and most of them showed no effect on alternative splicing (n=11), which were confirmed to be missense mutation. While variant *COL1A2*: c.792G>A, which expected to be a synonymous mutation, was proved to cause exon skipping. Interestingly, there were some complicated cases (n=3) in which more than one mutant transcript were found caused by recognition of different splice sites, and this can be found in variants located in either introns (n=2) or exons (n=1). We further analysed the correlation between genotypes of aberrant splicing and their phenotypes, and found that most of the patients who have the aberrant splicing presented mild OI phenotype, especially in the patients with multiple mutant transcripts and those with intron retentions. While for the most prevalent effect, the exon skipping, phenotypes of those patients varied from mild to severe: 30% mild OI, 30% moderate OI and 40% severe OI. The mechanism on the development of aberrant splicing from these atypical splicing variants was proposed and summarized. This study expands our knowledge of atypical splicing variants, and emphasizes the importance to classify the splicing effect for variants near exon/intron boundaries.

## PgmNr 3088: Genetic and genomic variants in *EPHA4*-mediated dendritic spine morphogenesis pathway contribute to idiopathic scoliosis.

### Authors:

L. Wang<sup>1,2</sup>; S. Zhao<sup>1,2</sup>; X. Fan<sup>3</sup>; L. Li<sup>4</sup>; Y. Zhang<sup>1,2</sup>; S. Chen<sup>3</sup>; X. Zhao<sup>4</sup>; Z. Wu<sup>1,5,7</sup>; N. Wu<sup>1,5,6,7</sup>; G. Qiu<sup>1,5,6,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Orthopedic Surgery, Peking Union Medical College Hospital, Peking, Beijing, China; 2) Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China; 3) Department of Pediatric Endocrine and Metabolism, Maternal and Child Health Hospital of Guangxi Zhuang Autonomous Region, Nanning 530000, China; 4) Department of Medical Genetics, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China; 5) Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Beijing, China; 6) Medical Research Center of Orthopedics, Chinese Academy of Medical Sciences, Beijing 100730, China; 7) Department of Central Laboratory, Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, China

**Background:** Idiopathic scoliosis (IS) is one of the most common pediatric skeletal disease. Complex genetic factors have been implicated in its etiology. Recently, advances in exome sequencing (ES) and whole genome sequencing (WGS) have made it efficient to investigate disease-associated variants in complex traits.

**Methods:** A total of 15 trios and 242 singleton patients with IS recruited in DISCO (Deciphering Disorders Involving Scoliosis & Comorbidities) project were included. ES was performed on the 15 trios and 124 singletons, while WGS was performed on 103 singletons, totaling 257 cases (287 subjects). A previously reported case affected with Waardenburg syndrome was also recruited and followed up for scoliotic phenotype. Western blot was performed to investigate the function alterations of likely-pathogenic *EPHA4* mutation.

**Results:** Using a de novo strategy for the 15 trios after ES, a splicing variant c.1443+1G>C in *EPHA4* was prioritized as a candidate variant for IS. Minigene assay showed that this variant induced a new splicing site, which lead to a 34del frameshift change. Another *EPHA4* heterozygous missense variant (c.2546G>A, p.Cys849Tyr) was identified in a sporadic IS patient. Western blot showed that this variant could disrupt the phosphorylation function of the *EPHA4* protein. In addition, we followed up a previously reported patient affected with Waardenburg syndrome caused by a 4.46 Mb de novo deletion which encompasses *EPHA4*, and found that he newly developed a mild IS (Cobb angle 15°, located at lumbar vertebrae). Under the hypothesis that *EPHA4* could lead to IS through dendritic spine morphogenesis pathway, we analyzed five genes involved in this process. Three potential pathogenic genetic variants were identified, including a novel missense variant c.896G>A, p.Gly299Asp in *EFNB3*, a de novo start-loss variant c.1A>G, p.Met1\* in *NGEF*, and a novel missense variant c.596G>A, p.Ser199Asn in *NGEF*.

**Conclusion:** In this study, we identified five (5/257, 1.9%) potential pathogenic variants in the *EPHA4*-mediated dendritic spine morphogenesis pathway in our IS cohort and one deletion encompassing *EPHA4* in an isolated IS patient, demonstrating the important role of this pathway in the etiology of IS.



## PgmNr 3089: A patient with cardiospondylocarpofacial syndrome caused by dominant negative effect of a novel variant in *MAP3K7*.

### Authors:

K. Yanagi <sup>1</sup>; K. Saito <sup>2</sup>; T. Ishitani <sup>3</sup>; M. Takeshita <sup>1</sup>; N. Kobayashi <sup>1</sup>; K. Satou <sup>1</sup>; T. Tohma <sup>4</sup>; I. Inoue <sup>5</sup>; Y. Matsubara <sup>6</sup>; T. Kaname <sup>1</sup>; Initiative on Rare and Undiagnosed Diseases Beyond

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genome Medicine, National Center for Child Health and Development, 2-10-1 Setagaya, Tokyo, Japan; 2) Invertebrate Genetics Laboratory, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka, Japan; 3) Research Institute for Microbial Diseases, Osaka University, 2-2 Yamadagaoka, Suita, Osaka, Japan; 4) Wanpaku Clinic 1490 Tsukazan, Haebaru, Okinawa, Japan; 5) Division of Human Genetics, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka Japan; 6) National Center for Child Health and Development, 2-10-1 Setagaya, Tokyo, Japan

---

Cardiospondylocarpofacial syndrome (CSCF, OMIM#157800) is characterized by growth retardation, dysmorphic facial features, brachydactyly, vertebral synostosis, cardiac septal defects, and deafness. We encountered a patient with growth retardation, dysmorphic facial appearance, vertebral abnormalities and congenital heart defect, who was clinically diagnosed with CSCF. Here we reveal that the patient is caused by a loss-of-function type of variant in *MAP3K7* by whole exome sequencing (WES). Effect of the variant was analyzed using model animals.

The patient was a 4-year-old boy. The boy was born at 39 weeks of gestation with Apgar index of 3 (1 min) and 6 (5 min). After birth, the boy was soon administrated our hospital due to severe cyanosis. At 4 year-of-age, the patient showed severe growth retardation (short stature (-5 S.D.)), developmental delay, dysmorphic facial features with macrocephalus, long face, frontal bossing, epicanthal fold and mid-face hypoplasia, congenital heart defects with coarctation of aorta and atrial septal defect, scoliosis due to vertebral synostosis and joint laxity. He was clinically suspected to CSCF.

Written informed consent was obtained from his parents. WES analysis was performed and a novel heterozygous missense variant in the *MAP3K7* gene (NM\_145331.3(MAP3K7\_v001):c.574A>G, p.(Ser192Gly)) was identified in the patient. The variant was confirmed using Sanger sequencing. It was de novo.

We investigated the effect of the *MAP3K7* variant in Zebrafish and Drosophila. Inhibition of *MAP3K7* function using zeaenol or morpholino affected head and heart development in zebrafish. The *MAP3K7* gene variant was introduced into transgenic Drosophila overexpressed eiger (TNFalpha), which phenotype was reduced the eye-size. It is known that *MAP3K7* acts at downstream of a signaling pathway of TNFalpha. Then, the eiger effect was suppressed by introduction of the variant *MAP3K7*, suggesting that the variant showed dominant negative effect.

We concluded that CSCF was caused by loss-of-function of *MAP3K7* and the patient was caused by dominant negative effect of the variant in *MAP3K7*.

## PgmNr 3090: Functional effects of 13 rare *WNT10A* variants identified in individuals with isolated tooth agenesis.

### Authors:

A. Letra <sup>1</sup>; Y. Zeng <sup>1</sup>; E. Baugh <sup>2</sup>; J.T. Hecht <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) UTHealth School of Dentistry at Houston, Houston, Texas.; 2) New York University, New York, NY; 3) UTHealth McGovern Medical School, Houston TX

---

Tooth agenesis (TA) is a common craniofacial anomaly affecting ~200 million humans worldwide that results from failure or disturbances during initial stages of tooth development. Both syndromic and isolated forms have been described, the latter which may occur sporadically or segregating in families. Numerous gene variants have been identified as potentially pathogenic for TA and their relevance in tooth development has been supported by studies using cells or animal model systems. *WNT10A* has been suggested as a major gene for TA, and yet the biological effects of associated variants remain unknown. Further, phenotypes in animal models showing *Wnt10a* knockdown/out are contradictory. In this study, we evaluated the effects of 13 rare *WNT10A* missense variants previously associated with TA using computational modeling (using VIPUR) and functional validation using a patient-based *in-dish* model system. Plasmid constructs for wild-type and each mutant *WNT10A* were transfected into SHED (stem cells of exfoliated deciduous teeth) cells for evaluation of *WNT10A* gene and protein expression, activation of WNT pathway, and their effects on expression of additional tooth development genes. Experiments were performed in triplicate and repeated twice. Data analysis was performed using one-way ANOVA followed by post-hoc tests as applicable; P-values<0.05 were considered statistically significant. Results showed that *WNT10A* variants located in exons 3 and 4 appear to contribute to a decrease in gene transcriptional activity and in WNT pathway activity. Specifically, T357I-, R360C-, and R379C-mutant cells presented significantly decreased *WNT10A* protein expression, likely through a reduced ability to bind to FZD. *WNT10A*-mutant cells also showed perturbed expression of *PAX9*, *MSX1*, *AXIN2*, and *RUNX2*. RNA-seq of T357I-mutant cells in comparison to WT cells identified 38 differentially expressed genes (10 upregulated, 38 downregulated), belonging to pathways related to cell pattern specification/regionalization, and adaptive immune response. Our results suggest that distinct *WNT10A* variants may present different biological effects contributing to impaired gene/function leading to milder or more severe TA phenotypes. Further studies investigating the potential additive effects of more than one *WNT10A* variant are currently under way.

# PgmNr 3091: The functional variant of *NTN1* contributed to the risk of non-syndromic cleft lip with or without cleft palate.

## Authors:

D. Li; L. Ma; S. Lou; X. Yu; Y. Pan

[View Session](#) [Add to Schedule](#)

**Affiliation:** Jiangsu Key Laboratory of Oral Diseases, Nanjing Medical University, Nanjing, Jiangsu, China

**Objective:** Our previous genome-wide association study of non-syndromic cleft lip with or without cleft palate (NSCL/P) identified a susceptible variant (rs4791774) in the intronic region of *NTN1*. We hypothesized that the functional single nucleotide polymorphism (SNP) may be in linkage disequilibrium with the lead SNP rs4791774 and contribute to the susceptibility to NSCL/P through a specific biological mechanism.

**Methods:** Bioinformatic analysis was conducted and the potential functional SNP (rs4791331) was identified. A case-control study with 891 cases and 830 healthy controls was designed. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using logistic regression analyses. The allele-specific DNA-protein binding preference was predicted by the JASPAR database. Cell proliferation, apoptosis, the cell cycle, luciferase activity and *NTN1* mRNA expression were examined after transfection with the rs4791331 C/T vector in HEK-293 and HEPM cell lines. Forty-six lip tissues of NSCL/P patients were collected to detect *NTN1* expression. The relative expression of *Ntn1* in 10.5–14.5 day mouse embryos was downloaded from the Gene Expression Omnibus database. *ntn1a* knockout zebrafish models were generated by CRISPR/Cas9 and observed with micro-CT.

**Results:** rs4791331 was identified as a functional variant highly linked to rs4791774. In the case-control study, the T allele was associated with an increased risk of NSCL/P (OR=1.41, 95% CI=1.19–1.68), as well as the subgroups: cleft lip only (OR=1.46, 95% CI=1.14–1.87) and cleft lip and palate (OR=1.58, 95% CI=1.27–1.96). The T allele of rs4791331 exhibited anti-apoptotic effects (HEK-293:  $P=0.011$ , HEPM:  $P=0.007$ ) and promoted cell cycle progression at the G1/S transition in HEK-293 ( $P=0.001$ ) and HEPM ( $P=0.011$ ) cells. Dramatically decreased enhancer activity and reduced expression of *NTN1* ( $P<0.001$  in HEK-293 and HEPM cells) following transfection of plasmids containing the T allele were observed. Carriers of the CT/TT genotypes showed significantly lower expression of *NTN1* than CC carriers ( $P=0.004$ ). *Ntn1* was detected in the distal and proximal maxillary process, with an increased trend of expression in the 10.5–14.5 day mouse embryos. The *ntn1a*<sup>-/-</sup> zebrafish showed relatively wider intermaxillary fissures than the control zebrafish.

**Conclusion:** rs4791331 (C>T) resulted in disrupted motif binding and led to abnormal expression of *NTN1*, which may be involved in the development of NSCL/P.

## PgmNr 3092: Somatic mutations in clonally expanded memory T-lymphocytes in type 1 diabetes and celiac disease.

### Authors:

M. Al-Riyami <sup>1</sup>; L. Marchand <sup>1</sup>; X. Du <sup>1</sup>; J. Wyse <sup>2</sup>; B. Richards <sup>3</sup>; C. Polychronakos <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, McGill University Health Centre Research Institute, Montreal, Quebec, Canada; 2) 1The Endocrine Genetics Laboratory, Child Health and Human Development Program and Jewish General Hospital, McGill Department of Medicine, Division of Gastroenterology; 3) Jewish General Hospital, Departments of Medicine, Human Genetics, Epidemiology and Biostatistics,

---

Autoimmunity is known to depend on both inherited susceptibility and environmental factors. However, these alone may not explain all variation in disease risk, for example the discordance of early-onset autoimmunity in monozygotic twins living in a shared environment, or the discordance in inbred-strain animal models kept in a controlled environment. These could be explained by stochastic events and one plausible, never explored, such event is post-zygotic mutations (PZMs) in the expanding antigen-specific autoreactive T-cell lineages, by analogy to their implication in the expanding tumor lineages of cancer cells.

In this study, we tested the hypothesis that somatic mutations in autoreactive T-cells involved in type 1 diabetes (T1D) and celiac disease (CD) contribute to the autoimmune process. Lymphocytes obtained from newly diagnosed T1D patients, proliferating after *ex vivo* activation by pro-insulin, were tested for somatic copy-number aberrations (CNA) not present in the germline (non-proliferating cells) by CGH array.

We found that 66.6% (12/17) of our study subjects harbor somatic copy-number mutations in insulin-reactive CD4+ T-cell clones. Many of these CNAs recurrently affected the same gene(s) in different patients, a strong evidence for pathogenicity in the cancer paradigm. Some of these genes have known T-cell proliferation and differentiation regulatory functions such as *CASZ1* and *E2f1*. As negative controls, we used lymphocytes proliferating in normal host defense by reacting to tetanus toxoid. They also harbored somatic mutations but these were significantly smaller in size ( $p = 0.0018$ ). Next, we tested the hypothesis in CD, another autoimmune disease. T-cells of CD patients, sampled after a 3-day gluten challenge and expanded *ex vivo* with gluten-peptides, were also examined for somatic CNAs. We found that 70% (7/10) of tested patients harbored somatic mutations in their gluten-reactive T-cells with three recurrent genes *CHD7*, *C8orf33*, and *LOC339685*.

Our data provide evidence for a potential causal role for PZMs in autoreactive T-cells in the pathogenesis of autoimmunity and suggested a putative mechanistic overlap between autoimmunity and cancer. These findings offer novel avenues for decrypting autoimmunity and its future individualized treatment through targeting disease causing cells.



## PgmNr 3093: Autoimmune disease-associated MHC polymorphisms influence T cell receptor repertoire.

### Authors:

K. Ishigaki<sup>1,2,3</sup>; Y. Luo<sup>1,2,3</sup>; S. Raychaudhuri<sup>1,2,3,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Divisions of Genetics and Rheumatology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; 2) Center for Data Sciences, Harvard Medical School, Boston, MA 02115, USA; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 4) Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA; 5) Graduate School of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA; 6) Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK

T cell reactivity to pathogenic autoantigens is a fundamental feature of autoimmunity, and is driven by T cell receptors (TCRs). Complementary determining region 3 (CDR3) is located at the antigen contact region of TCRs, and CDR3 determines antigen specificity. During the developmental process in thymus, TCRs are selected based on its affinity to autoantigens presented on major histocompatibility complex (MHC). Autoimmune disease-promoting MHC polymorphisms are primarily located in the antigen-binding pockets, and such polymorphisms affect spectrum of peptides presented to T cells. In this study, we hypothesized that these MHC polymorphisms would affect thymic selection of T cells and modify TCR repertoire.

To test our hypothesis, we performed CDR3 quantitative trait loci (cdr3-QTL) analysis where we treated the ratio of each amino acid (AA) at each position of CDR3 (position 104-118) as a quantitative trait, and tested its association with the AA dosages at each position of MHC genes. We first analyzed a deep-sequenced TCR dataset of peripheral blood presented in Emerson R. et. al. *Nature Genetics* (n=666). Among 181 CDR3 AAs and 835 HLA AA pairs we tested, 70 CAD3 AA had significant associations with at least one MHC AAs ( $P < 3.3e-07 = 0.05/151135$  hypothesis tested). We next validated these cdr3-QTL associations using the RNA-seq data of CD4<sup>+</sup> T cells from the BLUEPRINT consortium (n=169). The associations among class II MHC regions in the first dataset, which reflect the biology of CD4<sup>+</sup> T cells, were successfully replicated in the same effect directions (binomial test  $P$  value=1.7e-15). Finally, we examined the colocalization between MHC AA positions with significant cdr3-QTL effects and those associated with the risk of two autoimmune diseases (rheumatoid arthritis (RA) and type I diabetes (T1D)). In total, we identified 48 CDR3 AA modifications were possibly influenced by disease-promoting MHC polymorphisms. For example, the presence of valine or leucine at *HLA-DRB1* position 11, which is strongly associated with RA risk (OR~3.7), also alters glutamate usage at position 112 of CDR3.

Our previous understandings of the biological consequences of autoimmune disease-promoting MHC polymorphisms were limited to their high binding affinity to the pathogenic autoantigens (e.g. citrullinated peptides in RA). Here, we presented novel evidences that such MHC polymorphisms modify TCR repertoire, which might lead to aberrant T cell reactivity to pathogenic autoantigens.



# PgmNr 3094: Strong discordance between expression-level and secreted protein-level QTLs in induced monocyte cytokines.

## Authors:

L. Jostins-Dean <sup>1,2</sup>; A. Mentzer <sup>3,4</sup>; S. Makino <sup>3</sup>; G. McVean <sup>2</sup>; A. Hill <sup>3,4</sup>; J. Knight <sup>3</sup>; B. Fairfax <sup>5</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Kennedy Institute of Rheumatology, University of Oxford, Oxford, United Kingdom; 2) Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK; 3) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; 4) Jenner Institute, University of Oxford, Oxford, UK; 5) Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

## Introduction

Research into genetic determinants of human immunity is focused on either proximal cellular traits (e.g. gene expression) or systemic traits (e.g. blood cell counts). This approach neglects the role of effector cell function and context-modulated intracellular communication crucial to immune system function. Cytokines are signaling molecules that are sentinel to innate and adaptive immune responses, thus identification of genetic determinants of release is of high relevance to disease susceptibility. Here, we explore the impact of genetic variation on cytokine secretion during monocyte response to stimulation.

## Methods

Data were generated from monocytes exposed to LPS (20 ng/ml, 2 or 24 hours), or IFN $\gamma$  (20ng/mg for 24 hours) across 513 individuals. Gene expression and genotyping for this cohort has been previously described. 50 cytokines, chemokines and other inflammatory mediators (hereafter called cytokines) were quantified in collected supernatant using Luminex-based technology. After batch quantile normalization, and covariate regression, 23 high consistency cytokines were analysed. Genetic association analysis was performed using a MANOVA of genotype versus all 23 phenotypes.

## Results

Four genetic variants were associated with cytokine secretion profiles ( $p < 5e-8$ ). Two were *cis* (PDGF-BB and IL-1RA), and formed eQTLs to the relevant RNA. Two were in *trans* (*CCR1*, associated with multiple cytokines including MIP-1 $\beta$ ; *IFNB1*, impacting IP-10). Both PDGF-BB and *CCR1* are associated with inflammatory bowel disease. The PDGF-BB variant is only resolvable post-stimulation, suggesting cell-state specificity. The *CCR1* variant is associated with circulating MIP-1 $\beta$  but interestingly impacts multiple induced cytokines, indicating the strong systemic QTL may not be causal.

13/31 peak cytokine eQTLs across the three conditions have a nominal effect ( $p < 0.05$ ) on protein secretion. In many cases however, the RNA level post-stimulation is not correlated with protein levels. For example, a post-IFN $\gamma$ -stimulation eQTL explaining 8.8% of *TNF* RNA ( $p = 9.3^{-19}$ ) has no effect on measured TNF $\alpha$  ( $p = 0.78$ ).

## Discussion

This study demonstrates the importance of studying the correct “layer” of biological activity and the frequent mismatch between intermediate phenotypes, illustrative of biological buffering. Here we show that studying cytokine secretion from a single cell type can reveal phenotypes of disease-associated variants not noted in either eQTL or systemic QTLs.

## PgmNr 3095: Identification of novel cryptic pathogenic variants in two independent patients with deficiency of adenosine deaminase 2.

### Authors:

O. Schnappauf; N. Sampaio Moura; Q. Zhou; N. Deutch; I. Aksentijevich; D. Kastner

[View Session](#) [Add to Schedule](#)

**Affiliation:** Inflammatory Disease Section, NHGRI (National Human Genome Research Institute), Bethesda, Maryland.

---

Deficiency of adenosine deaminase 2 (DADA2) is an autosomal recessive disorder that manifests with fever, rash, hypocellular bone marrow, early-onset vasculitis, and propensity to ischemic and hemorrhagic stroke. This study aimed to identify novel pathogenic variants in *ADA2* through the application of a variety of molecular and biochemical methods.

The first index patient is a 5-year-old female who was born to a consanguineous Pakistani family and presents with features consistent with Diamond-Blackfan anemia (DBA). Sequencing analysis did not identify pathogenic variants in DBA associated genes and a CMA did not detect duplications or deletions. Several regions of homozygosity were identified in the proband, one of which comprises *ADA2*. Sanger sequencing of *ADA2* did not identify a pathogenic variant. MLPA analysis identified a homozygous duplication comprising exon 7 of *ADA2* in the proband, her affected sister, her affected father, and a heterozygous duplication in the mother. QRT-PCR experiments confirmed a loss of RNA-expression in the homozygous individuals. In agreement with mother's carrier status, she showed 50% reduction in *ADA2* RNA. Reduced *ADA2* enzyme activity corroborated the qRT-PCR findings. Long range PCR and long read sequencing were performed to localize the breakpoints of the duplication.

The second index patient is a 17-year-old female with a history of ischemic strokes, livedo and vasculitis. She was found to be heterozygous for a known pathogenic variant in *ADA2* (c.1358A>G, p.Y453C). It was inherited from the father and is present in her three unaffected siblings. *ADA2* enzyme activity assay revealed that the proband and two of her siblings have absent *ADA2* activity. Her parents, and a third sibling, exhibit *ADA2* carrier activity. Haplotype analyses found that the three children with low *ADA2* activity inherited the same maternal allele while the fourth sibling received the other maternal allele. WGS of mother and three affected children identified a novel canonical splice site variant (*ADA2*: c.-47+2T>C) present in all four individuals. Sanger sequencing confirmed segregation with disease. RNA analysis showed that the affected individuals express 50% of normal *ADA2* mRNA and cDNA sequencing confirmed that only the allele carrying the p.Y453C variant is expressed.

The identification of novel cryptic mutations in *ADA2* not detected by Sanger sequencing, suggests the incorporation of additional methods in the diagnosis of DADA2.

## PgmNr 3096: Using in vivo eQTL interactions to resolve the regulatory drivers of sepsis endotypes.

### Authors:

K.L. Burnham<sup>1</sup>; C. Goh<sup>2</sup>; A. Kwok<sup>2</sup>; Y. Mi<sup>2</sup>; T.C. Mills<sup>2</sup>; J. Radhakrishnan<sup>2</sup>; A. Rautanen<sup>2</sup>; P. Hutton<sup>3</sup>; N. Soranzo<sup>1</sup>; A.C. Gordon<sup>4</sup>; A.S.V. Hill<sup>2</sup>; C.J. Hinds<sup>5</sup>; S. Raychaudhuri<sup>6,7,8</sup>; E.E. Davenport<sup>1</sup>; J.C. Knight<sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Sanger Institute, Hinxton, United Kingdom; 2) Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Adult Intensive Care Unit, John Radcliffe Hospital, Oxford, UK; 4) Section of Anaesthetics, Pain Medicine and Intensive Care, Imperial College, London, UK; 5) William Harvey Research Institute, Barts and The London School of Medicine, Queen Mary University, London, UK; 6) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, USA; 7) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, USA; 8) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK

Sepsis is defined as organ dysfunction caused by a dysregulated host response to infection. Extensive variation in response across individuals has limited attempts to develop targeted treatments. We have previously identified two disease endotypes (Sepsis Response Signatures [SRS]) from transcriptomic data, which are associated with differential early mortality and response to hydrocortisone treatment in a clinical trial setting, but the regulatory drivers of these endotypes are unclear.

Expression quantitative trait locus (eQTL) mapping can identify genetic variants contributing to heterogeneity in gene expression. Specific eQTL may also be modulated by external factors, resulting in a genotype-by-environment interaction. We hypothesised that identification of a set of eQTL modulated by an environmental variable (e.g. causative infection type or SRS group) in the context of sepsis would reveal key regulatory factors and pathways through which the variable impacts the host response.

We recruited 550 adult patients with sepsis due to community acquired pneumonia or fecal peritonitis. We profiled gene expression (microarray, Illumina HT12-v4) in 730 peripheral blood leukocyte samples collected within five days of ICU admission and generated genome-wide genotyping data (Illumina OmniExpress, CoreExome, Global Screening Array).

We used a linear mixed model to identify *cis* eQTL genome-wide, and subsequently tested significant loci for interactions with SRS endotype and source of sepsis. Preliminary analysis identified eQTL for >2000 genes (Bonferroni corrected  $p < 5.2e^{-9}$ ), which were enriched for genes differentially expressed in the sepsis response and SNPs previously implicated in autoimmune and inflammatory disease. We observed 279 interactions (FDR < 0.05) with SRS, and 30 with infection site (lung/abdominal). These interaction eQTL involve a number of disease-relevant genes, including the transcription factor *IRF5* (SRS) and the viral response gene *OAS1* (site of infection). SRS interaction eSNPs were also enriched for GWAS risk variants for autoimmune disease; however, the majority of interaction eGenes were not differentially expressed between SRS groups, suggesting that this approach provides additional information to further resolve disease endotypes.

Identification of these eQTL interactions will help elucidate the mechanisms driving inter-individual

variation in the response to infection, and contribute to the development of personalised medicine in sepsis.

## PgmNr 3097: Genetic regulation of immune cell activation in subjects with immune-related disease.

### Authors:

N. Patsopoulos<sup>1 2 3</sup>; Z. Fang<sup>1 2 3</sup>; D. Hui<sup>1 2 3</sup>; I. Vlachos<sup>4</sup>; H. Klein<sup>5</sup>; V. Kuchroo<sup>1 2 3</sup>; H. Weiner<sup>1 2</sup>; P. De Jager<sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neurology, Brigham & Women's Hospital, Boston, Massachusetts.; 2) Harvard Medical School; 3) Broad Institute of Harvard and MIT; 4) Beth Israel Hospital, Boston, MA; 5) University of Columbia, New York

---

**Background:** Expression quantitative trait loci (eQTL) data have emerged as a source to characterize the function of disease associated genetic variants (DAGVs). Cell type specific data have recently allowed refined mapping of several disease associated loci, that were previously unattainable with tissue level results. Additionally, activation of cells and data derived from patients, instead of healthy controls, could allow further refinement of genetic functional mapping. An important question in the quest to functionally characterize DAGVs is whether data derived from subjects with the same disease and/or relevant activated cell types necessary.

**Methods:** We generated genetic and transcriptomic data for three immune cells - naïve CD4+ T cells (nT), memory CD4+ T cells (mT), and classical monocytes (Mo)- from multiple sclerosis (MS) subjects, untreated (UNT; n=55) or treated with glatiramer acetate (GA; n=84), and healthy controls (HC; n=38). We generated data at baseline (unstimulated) and after cell-specific activation (stimulated). We performed *cis* (*cis*-eQTL) and *trans* (*trans*-eQTL) analyses. We quantified the similarity of the genetic transcript expression regulation per cell type and condition within each group. Finally, we developed an algorithm to quantify the probability that the observed eQTL can be explained by the MS samples only.

**Results and Discussion:** We identified 600, 450, and 704 transcripts whose expression was controlled by nearby genetic variants (eGenes) in nT, mT, and Mo in the UNT MS group, respectively. The genetic control of transcript expression was more similar between the nT and mT ( $\rho=0.73$ ;  $p\text{-value}<10^{-50}$ ) compared to Mo ( $\rho<0.51$ ;  $p\text{-value}<10^{-50}$  in both). Activation resulted in decrease of the observed eGenes in all cell types with most prominent changes in Mo (1.5-fold in nT and mT, and 6-fold in Mo). A similar pattern was observed in the GA group, although more eGenes were identified, due to the larger sample size. Post-activation, the eGenes of the UNT and GA group were consistently similar across all three cell types ( $\rho\sim 0.85$ ;  $p\text{-value}<10^{-50}$  in all), implying that GA modulated the genetic control of subset of transcripts in different way compared to the UNT group. A similar pattern was observed with the identified *trans*-eQTL results. A small percent of identified eQTLs, either *cis*- or *trans*-, were only present in MS subjects suggesting genetic regulation of gene expression in the presence of disease-related context.

## PgmNr 3098: Post GWAS functional evaluation of SLC15A4 single nucleotide variants for Systemic Lupus Erythematosus (SLE) susceptibility.

### Authors:

G.P. Maiti; J.E. Molineros; B. Singh; J. Kaplan; S.K. Nath

[View Session](#) [Add to Schedule](#)

**Affiliation:** Oklahoma Medical Research Foundation, Oklahoma City, OKLAHOMA, 825 NE 13th St, OK 73104

**Background:** Systemic Lupus erythematosus (SLE) is a complex autoimmune disease with strong genetic basis. Recent studies have identified SLC15A4 gene as an important contributing factor for SLE pathogenesis. Several GWASs identified multiple disease associated SNPs within SLC15A4. However, the actual causal SNP(s) and contributing allele of SLC15A4 associated with SLE pathogenesis were largely unknown.

**Objective:** To characterize the candidate functional SNPs to define how they associate with disease susceptibility.

**Method:** Our meta-analysis with 6 independent cohort identified the significant SNPs in SLC15A4 gene region. Our *in-silico* bioinformatics and epigenetic analyses predicted the potential regulatory effects of SNPs on gene expression. Allele specific enhancer activity was measured in HEK, Jurkat and THP1 cell lines using luciferase reporter assay (LRA). Allele specific binding of different nuclear proteins were identified using DNA pull down assay followed by Electrophoretic Mobility Shift Assay (EMSA), Mass Spectrometry and western blot.

**Result:** Meta-analysis of SLC15A4 identified 653 informative SNPs among which 102 SNPs were qualified genome wide significance level associated with SLE. Further bioinformatics and epigenetic analysis prioritized 6 intronic SNPs which might have potential regulatory effects on gene expression. LRA identified significant allele specific different enhancer activities of 3 SNPs in HEK cell, 2 SNPs each in Jurkat and THP1. Interestingly the rs35907548 was showing highly significant difference ( $p < 0.005$ ) of allele specific enhancer activity across 3 cell lines. DNA pull down assay followed by EMSA and Mass Spectrometry identified cleaved PARP1 as an allele specific differential binding partner in rs35907548 region. We observed significant allele-specific bindings with several histones (H3K27Ac, H3K4Me3, and H3K4Me1). Our eQTL analysis demonstrated rs35907548 allele specific expression correlated with SLC15A4, GLT1D1 and TMEM132C.

**Conclusion:** The SNP rs35907548 of SLC15A4 is a potential functional variant predicted to alter the expression of SLC15A4, GLT1D1 and TMEM132C through its allele specific differential enhancer activity. Allele specific histone modification and differential binding of regulatory proteins may influence the expression of SLC15A4, GLT1D1 and TMEM132C, which might be leading to lupus.

## PgmNr 3099: Investigating the synaptic organisation of costimulatory IBD risk variants.

### Authors:

J. Hamp; P. Cespedes; S. Valvo; M. Dustin; L. Jostins-Dean

[View Session](#) [Add to Schedule](#)

**Affiliation:** The Kennedy Institute of Rheumatology, Oxford, United Kingdom

---

### Introduction

Inflammatory bowel disease (IBD) remains a highly prevalent disorder with a poorly understood aetiology and no cure. Genome-wide association studies (GWAS) have successfully identified 241 risk loci. These show a significant enrichment for eQTLs in CD4+ T cells, and contain numerous costimulatory molecules. These immunomodulators help form the immune synapse (IS), the interface between T cells and antigen presenting cells that helps regulate T cell activation. The organisation of immune synapses in IBD has not yet been explored. Here we investigate costimulatory molecule IBD risk genes within the IS of IBD patients as a first step in characterising a possible disease-related phenotype.

### Methods

Blood samples were obtained from consenting IBD patients attending the John Radcliffe Hospital (Oxford, UK). CD4+ Memory and Naïve T cells were isolated and stimulated using either anti-CD3 + anti-CD28 or isotype controls immobilised on 96-well plates. After 48hrs, cells were transferred onto glass-supported lipid bilayers (SLB) holding fluorescently tagged Anti-CD3, CD80, ICAM-1 and the costimulatory ligand of interest. After 20mins, SLBs were imaged on an Olympus IX83 inverted microscope and analysed using the Fiji software. Cells were also analysed by flow cytometry to confirm the expression of the costimulatory molecules.

### Results

Initial experiments on CD4+ T cells from healthy donors showed all GWAS-implicated costimulatory molecules could be detected upon activation. 24hrs and 48hrs post-activation were identified as shared time points for high expression of the costimulatory molecules compared to unstimulated controls.

Post-identification of key timepoints, we analysed the localisation of synaptic ICOS. ICOS signalling is a potential IBD risk factor as both ICOS and ICOSLG lie within IBD risk loci. CD4+ T cells were activated for 24hrs and introduced to a SLB containing ICOSLG, from which statistically different features of the IS could be discerned. The median fluorescent intensity (MFI) of Anti-CD3 was lower in stimulated samples; in contrast ICOSLG MFI was significantly higher. Additionally, the synaptic area for both Anti-CD3 and ICOSLG showed significant differences between stimulated and unstimulated groups.

This data shows our approach has the ability to quantify costimulatory molecule localisation within the IS. Our next steps are to implement this approach to all costimulatory molecules of interest for our IBD patient cohort.



## PgmNr 3100: RNA expression of *HLA-DRB* and *HLA-DQ* genes in healthy individuals differs between MHC class II haplotypes.

### Authors:

M. Houtman<sup>1</sup>; E. Hesselberg<sup>1</sup>; K. Tandré<sup>2</sup>; M.L. Eloranta<sup>2</sup>; L. Rönnblom<sup>2</sup>; L. Klareskog<sup>1</sup>; V. Malmström<sup>1</sup>; L. Padyukov<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Rheumatology, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden; 2) Department of Medical Sciences, Rheumatology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

---

**Background:** *HLA-DRB1* alleles have been associated with several autoimmune diseases. In anti-citrullinated protein antibody positive rheumatoid arthritis (ACPA-positive RA), *HLA-DRB1* shared epitope (SE) alleles are the major genetic risk factor. In contrast, the *HLA-DRB1*\*13:01 allele provides protection against ACPA-positive RA while the *HLA-DRB1*\*15:01 allele is not associated with RA. We investigated transcriptomic profiles of peripheral blood mononuclear cells (PBMCs) and isolated CD4+ T cells, CD8+ T cells, and CD14+ monocytes from healthy individuals carrying *HLA-DRB1* SE-positive and SE-negative alleles. **Methods:** Total RNA was extracted from PBMCs of genetically pre-selected healthy female individuals ( $n = 29$ ) and from CD4+ T cells ( $n = 26$ ), CD8+ T cells ( $n = 21$ ), and CD14+ monocytes ( $n = 10$ ) that were isolated from PBMCs via positive selection using microbeads. Sequencing libraries were generated using the Illumina TruSeq Stranded Total RNA kit and sequenced on an Illumina HiSeq 2500 platform, yielding about 40 million paired-end reads per sample. Reads were aligned to the standard reference using STAR. For the major histocompatibility complex (MHC) region, reads were mapped to available MHC reference haplotypes and *AltHapAlignR* was used to estimate gene expression. Differential gene expression (*HLA-DRB1* SE-positive versus SE-negative alleles (*HLA-DRB1*\*13:01 and *HLA-DRB1*\*15:01)) analyses were conducted using DESeq2. **Results:** We found no genes outside the MHC region that were differentially expressed between healthy individuals carrying *HLA-DRB1* SE-positive and SE-negative alleles in blood cells. In opposite, we found that *HLA-DRB* and *HLA-DQ* are differentially expressed between healthy individuals carrying *HLA-DRB1* SE-positive and SE-negative alleles. Interestingly, we found that *HLA-DRB1* is higher expressed in all studied cell types from individuals with *HLA-DRB1*\*04:01 alleles compared to individuals with *HLA-DRB1*\*15:01 alleles. **Conclusion:** Our data demonstrate that *HLA-DRB* and *HLA-DQ* genes are differentially expressed in healthy individuals carrying *HLA-DRB1* SE-positive and SE-negative alleles. This provides new insights into the molecular mechanisms of the development of autoimmunity.

## PgmNr 3101: *DLGAP4*: A novel candidate gene for autism spectrum disorders.

### Authors:

J. Mazzeu <sup>1,2</sup>; H. Dornelles-Wawruk <sup>2</sup>; M.M. Mehrjouy <sup>3</sup>; M. Bak <sup>3</sup>; L. Nazaryan-Petersen <sup>3</sup>; M. Santos <sup>1</sup>; M. Picanço <sup>1</sup>; N. Tommerup <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculdade de Medicina, Universidade de Brasilia, Brasilia, Brazil; 2) Programa de Pós-graduação em Ciências da Saúde, Universidade de Brasilia, Brazil; 3) Department of Cellular and Molecular Medicine, University of Copenhagen, Copenhagen, Denmark

---

Autism spectrum disorders are a group of neurodevelopmental disorders characterized by impaired verbal/nonverbal communication, stereotyped behavior and abnormal social skills. Several chromosomal imbalances and mutations have been demonstrated to contribute in the etiology of autism. However, for most cases the cause remains elusive. We studied monozygotic twin sisters with autism spectrum disorder carrying a balanced translocation 46,XX,t(9;20)(q22;q13.3),9qh+,9qh+. Chromosome microarray analysis did not disclose any imbalances. In order to precisely map the translocation breakpoints, we performed mate-pair sequencing. The breakpoint on chromosome 9 maps to 9q21.12 with no genes mapped nearby. The breakpoint on chromosome 20 disrupts *DLGAP4* between exons 4 and 5. *DLGAP4* belongs to *DLGAP* gene family that includes 5 genes. DLGAPs are highly expressed in the postsynaptic density and are involved in neuronal signaling through glutamate receptors. Mutations in *DLGAP1-4* have been described in several neurodevelopmental disorders such as obsessive compulsive disorder, parkinson's disease and schizophrenia. *DLGAP2* has been proposed as a candidate for autism spectrum disorder. *DLGAP4* has been previously linked with early-onset cerebellar ataxia but not autism, though Sapap4 (Dlga4) deficient mice present autistic-like behavior. Therefore, there is strong evidence that dysregulation of *DLGAP4* might cause neurodevelopmental disorders. Hence we propose that *DLGAP4* is a strong candidate for autism spectrum disorders.

## PgmNr 3102: New cases with “CNV-mutator” phenotype characterized by concurrent *de novo* NAHR-mediated deletion events.

### Authors:

B. Yuan <sup>1</sup>; C. Grochowski <sup>1</sup>; C.M.B. Carvalho <sup>1</sup>; S. Gu <sup>1</sup>; J. Posey <sup>1</sup>; M. Withers <sup>1</sup>; C. Shaw <sup>1</sup>; S.W. Cheung <sup>1</sup>; W. Bi <sup>1</sup>; J.R. Lupski <sup>1,2,3,4</sup>; P. Liu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX 77030, USA; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 3) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 4) Texas Children's Hospital, Houston, TX 77030, USA

---

**Background:** The CNV-mutator phenotype is described by multiple *de novo* copy number variants (MdnCNV) arising independently at multiple loci throughout the human genome. The MdnCNV genomic phenotype is ultra-rare, and reported cases are predominantly copy number gains resulting from tandem duplications or complex genomic rearrangement (CGR) restricted to the perizygotic period of early embryogenesis. A Replication-based repair mechanism (RBM) involving template switching (TS), such as FoSTeS/MMBIR, is proposed to cause MdnCNVs, as suggested by mutational signatures at the breakpoint junctions. It is hypothesized that deficient protein contributing to faithful DNA replication and repair machinery is expressed in maternal germ cells, carried over to the fertilized zygote, and functioning at the early cleavage stage before post-maternal clearance. However, an inheritable genetic defect causing this phenomenon is not-yet identified.

**Methods:** New samples with CNV profiles consistent with the MdnCNV phenomenon were ascertained from individuals subjected to chromosome microarray analysis (CMA) at Baylor Genetics (BG). DNA sequencing further characterized the MdnCNVs.

**Results:** We identified one new MdnCNV sample. Eight *de novo* duplications, ranging from 901 Kb to 1.032 Mb, dispersed on eight chromosomes including 4, 5, 6, 10, 12, 13, 14 and 21, along with an 8 Mb copy number neutral region with absence of heterozygosity (AOH) on 15q14q21.1. Interestingly, among 65,000 CMA samples at BG, we identified two cases with two concurrent NAHR-mediated recurrent deletion CNVs ranging from 227 Kb to 3.062 Mb (15q13.3 and 17q21.31 in one sample, and 15q24 and 16p11.2 in the other sample). These two cases with recurrent deletion CNVs are reminiscent of the MdnCNV phenomenon yet the “genomic character” is distinct from the previous predominant observation of gains.

**Conclusion:** We provide additional evidence for the MdnCNV genomic phenomenon. The MdnCNV duplications coupling AOH are consistent with CGRs due to potential RBM malfunctioning during replicative repair of single-ended, double-stranded DNA (seDNA). The two MdnCNV cases with multi-locus *de novo* NAHR-mediated deletions suggest contribution of a different mutational mechanism. Could these be caused by defects in repair of double strand breaks initiated during chromosome pairing and meiotic homologous recombination?

## PgmNr 3103: Disruptive *KMT5B* variation alters growth and adhesion properties in an *in vitro* model.

### Authors:

R. Wickramasekara; S. Anderson; H.A.F. Stessman

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pharmacology & Neuroscience, Creighton University School of Medicine, Omaha, NE.

---

Advances in sequencing technology have revealed extensive overlap in risk genes among complex, yet seemingly unrelated, human disorders. Recently, publications have highlighted *de novo* disruptive mutations in *KMT5B* as potentially-pathogenic in neurodevelopmental disorders (NDDs); variants in *KMT5B* have also been implicated in metastatic human breast cancer. As a chromatin-modifying enzyme, lysine methyl transferase 5B (*KMT5B*) is responsible for dimethylating the N-terminal lysine 20 residue of histone protein 4 (i.e., H4K20me2). Published data suggest that H4K20me2 is crucial in chromatin compaction, regulation of gene expression, cell cycle progression, maintenance of stem cell quiescence and in DNA double-stranded break repair. However, what role disruptive mutations in *KMT5B* have in NDD and/or cancer etiology is currently unknown. This study is aimed at understanding the biology underlying *KMT5B* mutations with the hypothesis that early-truncating, disruptive *KMT5B* mutations in human cell line models will result in reduced H4K20me2 and cell growth defects that may be initially characterized in this model system. CRISPR/Cas9 engineering was used to induce *KMT5B* mutations into the human myeloid cell line, K562. Multiple isogenic clones were derived, and three lines (SCC1, SCC2, and SCC3) containing unique heterozygous frameshifting mutations in *KMT5B* were chosen for further characterization. Although all three lines contained similar *KMT5B* variation (p.Ser69Profs\*11, p.Met72Thrfs\*3, and p.Ala74Glufs\*5, respectively), only SCC1 and 2 showed significant decreases in proliferation ( $p < 0.001$ , Student's t-test). Further, SCC2 showed reduced levels of H4K20me2, defects in G1-S phase transition, and an accumulation of double-stranded breaks in G1 and G2/M ( $p < 0.05$ , Student's t-test). RNA-sequencing of all three lines highlighted a conserved upregulation of *p21* (*CDKN1A*;  $p < 0.01$ , Student's t-test) and significant pathway changes in cell adhesion, cell lineage, proteoglycans, and Rap1 signaling compared to the sham control line ( $p < 0.05$ , FDR-corrected Fisher's exact test). Ongoing ChIP-sequencing experiments aim to better define putative *KMT5B* targets associated with the gene expression changes observed. Preliminary data indicate that disruptive mutations in *KMT5B* affect pathways common to tumorigenesis and neurogenesis—proliferation, cell cycle dysregulation, migration, and adhesion—suggesting that this gene may be important for the etiology of multiple disorders.

## PgmNr 3104: Loss-of-function mutation in inositol monophosphatase 1 (IMPA1) in intellectual disability patients impairs neurogenesis but not gliogenesis.

### Authors:

T. Figueiredo <sup>1,2</sup>; A.P.D. Mendes <sup>2</sup>; G.S. Kobayashi <sup>1</sup>; D.P. Moreira <sup>1</sup>; D. Oliveira <sup>1</sup>; E. Goulart <sup>1</sup>; F. Kok <sup>1</sup>; M.C. Marchetto <sup>2</sup>; R. Santos <sup>2,3</sup>; F.H. Gage <sup>2</sup>; M. Zatz <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genome and Stem Cell Research Center, Biosciences Institute, University of São Paulo, 106 Rua do Matão, 05508-090, São Paulo, Brazil; 2) Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA.; 3) Laboratory of Dynamic of Neuronal Structure in Health and Disease, Institute of Psychiatry and Neuroscience of Paris (UMR\_S1266 INSERM, University Paris Descartes), 102 rue de la Santé, 75014 Paris, France.

---

The IMPA1 enzyme is responsible for the generation of free inositol, and is therefore a modulator of signal transduction via the production of the second messengers, inositol triphosphate and diacylglycerol in brain as well as in numerous other tissues. In 2016, we described a novel homozygous 5-bp duplication leading to a frameshift and a premature stop codon in the inositol monophosphatase 1 (*IMPA1*) in nine individuals with severe intellectual disability and disruptive behavior. Currently, this mutation co-segregated in 59 genotyped family members from Northeastern Brazil. In this study, we derived iPSC lines from 3 patients with *IMPA1* mutation and 2 controls to expand the studies to understand the mechanisms by which impairment of *IMPA1* lead to development of severe ID and disruptive behavior. The iPSCs from our cohort were differentiated into hippocampal dentate gyrus-like neurons (hDG neurons), pan-neuronal cells and astrocytes and we investigated the effect of the mutation on signaling pathway networks, cellular processes, morphology and electrophysiological properties. The transcriptome analysis of hippocampal dentate gyrus-like neuronal precursors cells (hDG NPCs) and hDG neurons revealed significant alterations in genes involved in multiple developmental processes including embryonic development and morphogenesis, brain development, neurogenesis, neuron differentiation and gliogenesis. Moreover, we observed that neural progenitor cells have defects in proliferation and survival. These defects were associated with cell cycle arrest and increased apoptosis specifically in NPCs and not in iPSCs or GPCs (glial progenitor cells). Importantly, we showed that *IMPA1* mutation affects the ability of progenitor stem cells to differentiate into neurons but not into astrocytes and that inhibition of Rho-associated kinases (ROCK) rescued neurogenesis. Funding: CEPID-FAPESP, FAPESP (Process Numbers: 2017/19877-0 and 2016/09618-5), INCT, CNPq.

## PgmNr 3105: Bi-allelic variants in *AP1G1* and *SYNRG* cause autosomal recessive intellectual disability with developmental delay in humans.

### Authors:

M. Usmani <sup>1</sup>; F. Rashed <sup>2</sup>; M. Zaman <sup>3</sup>; M. Contreras <sup>4</sup>; E. Richard <sup>1</sup>; M. Shahzad <sup>2</sup>; H. Bokhoven <sup>5</sup>; T. Blanpied <sup>4</sup>; Z. Ahmed <sup>1</sup>; S.H. Riazuddin <sup>2,3</sup>; S. Riazuddin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Otorhinolaryngology, School of Medicine, University of Maryland, Baltimore, MD; 2) National Centre of Excellence in Molecular Biology, University of The Punjab, Lahore, Pakistan; 3) Allama Iqbal Medical College, University of Health Sciences, Lahore, Pakistan; 4) Department of Physiology, School of Medicine, University of Maryland, Baltimore, MD; 5) Department of Human Genetics, Donders Institute for Brain Cognition and Behavior, Radboud University Medical Center, Nijmegen, The Netherlands

---

Adaptor protein (AP) complexes mediate clathrin-coated vesicle assembly, intracellular sorting, and vesicular trafficking of membrane proteins in the cell. Various subunits of AP complexes are involved in severe human inherited disorders like MEDNIK syndrome, developmental and epileptic encephalopathy, and intellectual disability (ID). Of the five adaptor protein complexes (AP-1 to -5), AP-1 mediates cargo sorting between endosomes, lysosomes and trans-Golgi network, as well as in synaptic recycling in neurons. Here, we report two novel predicted homozygous pathogenic variants, including a missense variant c.1096 A>G (p.Met366Val) in *AP1G1* and a c.3211\_12delAA (p.Lys1071Glufs\*7) frameshift variant in its interacting partner *AP1GBP1* (*SYNRG*) segregating with mild to moderate ID, developmental delay, pectus excavatum, B/L toe micronychia, hypotonia, and epilepsy in two large consanguineous Pakistani families. *AP1G1* is part of the adaptor related complex 1, crucial for vesicular transportation and formation. However, the function of *SYNRG* is yet to be defined. By co-immunoprecipitation (co-IP) and immunofluorescent studies, we confirmed localization of both *AP1G1* and *SYNRG* proteins in specific vesicle bodies, and interactions with each other as well as with other AP-1 adaptors. Although the protein targeting or stability was not impacted by the disease-causing variants, however, co-IP studies revealed significant impact on the protein-protein interactions. Furthermore, *AP1G1* harboring p.Met366Val variant significantly delayed transferrin recycling through recycling endosomes, while the *SYNRG* with p.Lys1071Glufs\*7 frameshift mutation delayed both static early and late endosomes recycling. In zebrafish, knocking-out *Ap1g1* leads to severe morphological defect and lethality by 6 days post fertilization (dpf). Using CRISPR/Cas9 technology, we introduced truncation in *Synrg* similar to human variant. In contrast to the depletion of *Ap1g1*, truncation of *Synrg* does not affect survival of fish but decreased size of corpus callosum was observed as early as 3dpf. Taken together, these results confirm the function of both *AP1G1* and *SYNRG* in endocytic recycling and their role in normal brain development, as well as expand the repertoire of genes associated with ID and developmental delay in humans.



## PgmNr 3106: A novel homozygous variant in *TRAPPC2L*, a core subunit of the human TRAPP tethering factor complexes involved in membrane trafficking, results in a neurodevelopmental disorder.

### Authors:

M. Ganapathi <sup>1</sup>; P. Ahimaz <sup>2</sup>; M. Milev <sup>3</sup>; V. Okur <sup>2</sup>; N. Al-Deri <sup>3</sup>; M. Sacher <sup>3</sup>; W.K. Chung <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pathology & Cell Biology, University Medical Center, New York, NY, 10032, USA; 2) Department of Pediatrics, Columbia University Medical Center, New York, NY, 10032, USA; 3) Concordia University, Department of Biology, 7141 Sherbrooke Street West, Montreal, Quebec, Canada H4B 1R6; 4) Department of Medicine, Columbia University Medical Center, New York, NY, 10032, USA

Membrane trafficking shuttles proteins to the various intracellular compartments using membrane bound vesicles and requires numerous proteins including tethering factors. TRAPP (TRANsport Protein Particle) tethering factors are highly conserved protein complexes in eukaryotes which regulate fusion of vesicles to membranes. In humans, two TRAPP complexes, II and III, have been described. Each has a core of common subunits (*TRAPP-C1*, *C2*, ***C2L***, *C3*, *C4*, *C5* & *C6*) as well as complex-specific proteins.

Recently, 2 unrelated individuals with a homozygous p.Asp37Tyr variant in the ***TRAPPC2L*** core subunit were described with neurodevelopmental delay, fever associated encephalopathy & rhabdomyolysis, followed by developmental arrest, epilepsy and tetraplegia (J Med Genet. 2018 Nov;55(11):753-764). Using genome sequencing, we identified a novel *TRAPPC2L* homozygous missense variant (**p.Ala2Gly**) which segregated in 3 affected individuals in a sibship of 5, from a single Ashkenazi Jewish (AJ) family. This rare variant is only seen in the AJ cohort (AF-0.0002243) in GnomAD and is predicted to be deleterious (CADD 28.3, REVEL 0.68, MPC 0.46, MetaSVM damaging 0.58, GERP 3.78). The affected siblings have a history of global developmental delay, intellectual disability, severe speech delay and acquired microcephaly. Developmental regression after illness, rhabdomyolysis and seizures were not consistently noted.

To assess the effect of the *TRAPPC2L* p.Ala2Gly variant we performed functional studies, using yeast and human cells. Yeast two hybrid analysis showed that this variant blocks the interaction with *TRAPPC6a* and the TRAPP III-specific *TRAPPC12* subunit whereas the p.Asp37Tyr variant disrupts the interaction with TRAPP II-specific *TRAPPC10* subunit. In a yeast system with growth dependent on a functional form of *TRAPPC2L*, the p.Ala2Gly variant was unable to rescue the growth defect at 37°C. Analysis of membrane trafficking between the endoplasmic reticulum and the Golgi in live cells derived from the affected individual showed a delay, which was rescued by expression of wild type *TRAPPC2L*. *TRAPPC2L* is postulated to bind the complex close to the *TRAPPC3-TRAPPC6* interface. We found that the p.Ala2Gly, but not the p.Asp37Tyr variant, reduced the interaction of *TRAPPC2L* with the *TRAPPC3-TRAPPC6a* heterodimer *in vitro*. Collectively, these data support the pathogenicity of the *TRAPPC2L* p.Ala2Gly variant that may be due to a reduced association of *TRAPPC2L* with *TRAPPC6a* and or *TRAPPC12*.

## PgmNr 3107: JAK/STAT inhibitor therapy rescues the lipodystrophic and autoimmune phenotype in *Clec16a* KO mice.

### Authors:

R. Pandey <sup>1</sup>; H. Hain <sup>1</sup>; B. Strenkowski <sup>1</sup>; M. Bakay <sup>1</sup>; H. Hakonarson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.; 2) Department of Pediatrics, the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

*CLEC16A* is implicated in multiple autoimmune diseases. We generated *Clec16a* inducible knockout mice to address the role of *CLEC16A* loss of function in autoimmunity. *Clec16a* KO mice exhibited increased food consumption and without evidence of hyperglycemia, loss of adipose tissue, severe weight loss, elevated Immunoglobulins and significantly reduced circulating insulin levels. Metabolic analysis revealed disturbances in lipid profile measures. White adipose tissue decreased concomitantly with enhanced inflammatory response and energy wasting. The loss of *CLEC16A* leads to a vicious cycle of autophagic impairment and endoplasmic reticulum (ER) stress, which contributes to excessive lipolysis and lipotoxicity resulting in activation of JAK/STAT, mTOR, P38 and JNK and release of multiple proinflammatory mediators under compromised mitophagy environment. Treatment with a JAK/STAT inhibitor (tofacitinib) partially rescued the lipodystrophic phenotype of KO mice and improved survival by alleviating ER stress and JAK-STAT mediated cytokine signaling. These results establish a link between *CLEC16A*, ER stress, lipid metabolism and the immune system. *CLEC16A* is regulator of autophagy and mitophagy. ER stress is known to activate lipolytic cascade. Aberrant or excess cytokine production plays a key role in driving autoimmune and autoinflammatory disorders. Collectively, our data indicate that loss of *CLEC16A* induces ER stress, dysregulated autophagy and mitophagy, which in turn activates lipolytic cascade resulting in excessive adipocyte lipolysis which generates lipid mediators and triggers inflammation via activation of JNK/NF- $\kappa$ B signaling pathway. Thus, *CLEC16A* exerts its effect on a wide variety of immune cells through modulation of ER stress, SOCS expression and regulation of cytokine signaling, suggesting that perturbations in the molecular link between *CLEC16A*, ER stress, mitophagy, lipophagy, and SOCS1 may underlie inflammatory and autoimmune disorders. In patients with autoimmune disease, harboring variants that result in *CLEC16A* hypofunction, drugs with modulatory effects on ER stress/mitophagy/autophagy/SOCS1-JAK-STAT signaling could compensate for the attenuated *CLEC16A* activity and be developed for targeted interventions. Our *Clec16a* KO highlights multifaceted roles of *Clec16a* in normal physiology, including a novel target for weight regulation, as well as mutation-mediated pathophysiology.



## PgmNr 3108: Identification of novel genetic mechanisms involved in development and progression of non-alcoholic fatty liver disease.

### Authors:

H.-J. Kim <sup>1,2</sup>; S. Gustafsson <sup>3</sup>; P. Saliba-Gustafsson <sup>1,2,4</sup>; C. Park <sup>1,2</sup>; E. Ingelsson <sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, California, USA; 2) Stanford Diabetes Research Center, Stanford University, Stanford, California, USA; 3) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 4) Cardiovascular Medicine Unit, Department of Medicine, Center for Molecular Medicine at BioClinicum, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden.; 5) Stanford Cardiovascular Institute, Stanford University, Stanford, California, USA

---

Non-alcoholic fatty liver disease (NAFLD) is the most prevalent form of chronic liver disease, characterized by excess fat accumulation in the liver without a history of excessive alcohol consumption. Patients with NAFLD often progress to an advanced stage of non-alcoholic steatohepatitis (NASH) and ultimately irreversible end-stage liver disease. To date, there are no available pharmacological treatments against NAFLD and NASH.

The overall aim of this study was to investigate genetic mechanisms regulating lipid accumulation and lipotoxicity in hepatocytes, which will help us understand the cause of NAFLD and NASH. We treated human hepatocytes (HepG2) with the two most abundant dietary plasma fatty acids; a saturated fatty acid, palmitic acid (PA), and a monounsaturated fatty acid, oleic acid (OA); as well as TNF to study their impact on steatosis, apoptosis and inflammation, as well as gene expression, with the goal of developing an *in vitro* model for NAFLD development and progression.

First, we showed that OA markedly induced lipid accumulation, as well as mRNA expression of the lipid droplet binding protein, *Plin2*. In contrast, PA had little effect on lipid accumulation, but significantly activated the Caspase 3/7 apoptotic and NF- $\kappa$ B p65 inflammatory pathways. Second, using RNA-seq analysis, we studied differential expression (DE) by our perturbations. In addition to genes from lipid metabolism pathways, OA reduced expression of several fibrosis-related genes (*COL3A1* and *COL1A2* being the top DE genes;  $\log_2$ FC, -5.8 and -4.9;  $P_{adj}$ , 4.9E-14 and 1.4E-9, respectively); while the top DE gene after PA treatment was carnitine palmitoyltransferase 1A (*CPT1A*), a key enzyme in fatty acid oxidation ( $\log_2$ FC, 1.2;  $P_{adj}$ , 1.1E-15). As expected, top DE genes after TNF treatment were leucine-rich alpha-2-glycoprotein 1 (*LRG1*) and interleukin 32 (*IL32*) ( $\log_2$ FC, 1.1 and 1.6;  $P_{adj}$ , 1.27E-13 and 1.46E-08, respectively), which are involved in immune response.

Our preliminary results indicate that perturbations with PA, OA and TNF induce different effects on the hepatocyte lipid accumulation and inflammation that closely resembles development and progression of NAFLD and NASH. Next, we plan to perform further molecular profiling in our system and integrate findings from human genetics to advance knowledge about NAFLD and NASH.

## PgmNr 3109: Variants associated with liver chromatin accessibility suggest mechanisms for loci associated with lipid levels and gene expression.

### Authors:

K.W. Currin<sup>1</sup>; M.R. Erdos<sup>2</sup>; N. Narisu<sup>2</sup>; V. Rai<sup>3</sup>; K.A. Broadaway<sup>1</sup>; J.R. Idol<sup>2</sup>; T. Yan<sup>2</sup>; J.P. Didion<sup>2</sup>; P. Orchard<sup>3</sup>; R.D. Albanus<sup>3</sup>; X. Yin<sup>4</sup>; A.S. Chaudhry<sup>5</sup>; A.S. Etheridge<sup>6</sup>; P.J. Gallins<sup>7</sup>; Y. Zhou<sup>7</sup>; F. Wright<sup>7</sup>; F. Innocenti<sup>6,8</sup>; E.G. Schuetz<sup>5</sup>; L.J. Scott<sup>4</sup>; S.C.J. Parker<sup>3,9</sup>; K.L. Mohlke<sup>1</sup>; F.S. Collins<sup>2</sup>; NIH Intramural Sequencing Core

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI; 4) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI; 5) Department of Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN; 6) Eshelman School of Pharmacy and Center for Pharmacogenomics and Individualized Therapy, University of North Carolina, Chapel Hill, NC; 7) Bioinformatics Research Center, North Carolina State University, Raleigh, NC; 8) Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC; 9) Department of Human Genetics, University of Michigan, Ann Arbor, MI

Genetic variants that alter chromatin accessibility in disease-relevant tissues are candidate functional regulatory variants and provide mechanistic insight at GWAS loci. We mapped chromatin accessibility in frozen liver tissue from 20 individuals from the St Jude liver bank and used MACS2 to identify 287,589 consensus accessible chromatin regions (peaks, FDR<5%, median width=538 bp) that were present in at least 2 individuals. We mapped chromatin accessibility quantitative trait loci (cQTL) with RASQUAL using variants (imputation  $r^2 > 0.3$  and MAF>.10) within 25 kb of consensus peak centers, correcting for library size and 4 ATAC-seq principal components. We identified cQTL for 1,011 peaks (genome-wide FDR<5% after correcting for local peak variant density), corresponding to 1,001 unique variants. Consistent with previous cQTL studies, we found that cQTL lead variants are often located close to the associated peaks; 429 (43%) cQTL variants are within the associated peak, and the median distance from lead variants to peak centers is 1,628 bp. To identify variants that alter gene expression by modulating chromatin accessibility, we identified cQTL variants in high linkage disequilibrium (LD) with lead variants from a new liver eQTL meta-analysis (n=1183). 120 (12%) cQTL lead variants were in high LD ( $r^2 > 0.8$ ) with a lead liver eQTL variant. 5 of these cQTL were in high LD ( $r^2 > 0.8$ ) with liver eQTL that colocalized with GWAS loci for blood lipid levels (*PRMT6*, *SORT1/CELSR2/PSRC1*, *ANGPTL1/C1orf220/RALGPS2*, *LITAF*, and *DNAH11*). At the well-characterized *SORT1* GWAS locus for LDL and total cholesterol, the lead cQTL variant (rs12740374,  $p < 8.5 \times 10^{-11}$ ) is also the lead eQTL variant for *SORT1*, *PSRC1*, and *CELSR2*. Allele rs12740374-T, which has been shown to generate a CEBP binding site and increase *SORT1* expression in liver cells (Musunuru *et al.*, 2010), is associated with increased chromatin accessibility of a peak overlapping the 3' end of *CELSR2* and increased expression level of all three genes, suggesting that this variant increases gene expression by increasing chromatin accessibility and binding of a CEBP transcription factor. Despite modest sample size, we identified genetic variants that may alter gene expression and metabolic

traits by modulating chromatin accessibility. These results suggest that mapping liver cQTL in larger sample sizes will be a powerful method to identify candidate functional variants and molecular mechanisms at metabolic trait GWAS loci.

## PgmNr 3110: Colocalizing gene-lifestyle interaction associations with molecular signatures of smoking and alcohol consumption.

### Authors:

T.D. Majarian <sup>1</sup>; A. Bentley <sup>2</sup>; C.C. Gu <sup>3</sup>; P.S. de Vries <sup>4</sup>; A.K. Manning <sup>5,6</sup>; CHARGE Gene-Lifestyle Interactions Working Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts.; 2) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Division of Biostatistics, Washington University School of Medicine, St. Louis, MO; 4) Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA; 5) Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA; 6) Harvard Medical School, Cambridge, MA

---

Cigarette smoking and alcohol consumption modify genetic effects on lipid and blood pressure levels, yet the underlying mechanisms remain largely unexplained. We leveraged trans-ancestry GWAS summary statistics from the Gene-Lifestyle Interactions working group (GLIWG) within the Cohorts for Heart and Aging Research in Genomic Epidemiology consortium with data from external publications to colocalize genetic associations with molecular signatures of lifestyle exposures.

GWAS summary statistics, generated within the GLIWG, for three lipid (total triglycerides, HDL and LDL cholesterol) and four blood pressure traits (systolic [SBP], diastolic [DBP], pulse [PP], and mean arterial [MAP] pressure) with two lifestyle factors (cigarette smoking and alcohol consumption) as interaction terms in over 600,000 individuals of diverse ancestry were used. Associations were selected by interaction  $P$ -value  $< 5E-8$  per trait-exposure pair.

Molecular signatures of traits and exposures were taken from the literature focusing on epigenome-wide association studies (EWAS), methylation quantitative trait loci (mQTL), expression quantitative trait methylation (eQTM), and expression quantitative trait loci (eQTL) in adipose, liver, whole blood, and pancreatic tissue. Tissues were chosen based on differential heritability enrichment between exposed and unexposed individuals.

Colocalization was established by intersecting associations with molecular signatures relevant to each trait-exposure pair. Two criteria defined colocalization: at least one association (1) within 500kb of an EWAS/eQTM site and an eQTL of the eQTM target or (2) is an mQTL of an EWAS/eQTM site.

Two loci showed colocalization in analyses of HDL and alcohol: C1QTNF4 and SIRT2. rs10769254 ( $P=1.4E-8$ ) was an eQTL for C1QTNF4 ( $P=3.2E-7$ ) and 250Kb upstream of cg27051260, an EWAS site for alcohol consumption ( $P=1E-4$ ) and an eQTM of C1QTNF4. rs7121347 ( $P=2.4E-12$ ) was an eQTL for SIRT2 ( $P=2.3E-10$ ) and 185Kb upstream of cg24064506, an EWAS site for alcohol consumption ( $P=1.7E-5$ ) and an eQTM of SIRT2. 51 variants with  $P<5E-8$  were within the SIRT2 locus. These

variants show low LD with multiple known independent signals for HDL within the region. Further, SIDT2 has been implicated in lipid storage in mouse knockouts.

EWAS/eQTM in alcohol consumption provided context for observed genetic interaction. Our colocalization analysis suggests a genetic role in the epigenetic mechanism of alcohol consumption's effect on lipid levels.

## PgmNr 3111: Capture C-based variant to gene mapping for GWAS-implicated total cholesterol loci points to a degree of common genomic etiology between adipocytes and hepatocytes.

### Authors:

M.E. Johnson <sup>1</sup>; A. Chesi <sup>1</sup>; E. Manduchi <sup>1,3</sup>; Y. Wagley <sup>4</sup>; M.E. Leonard <sup>1</sup>; K.M. Hodge <sup>1</sup>; M. Caliskan <sup>5</sup>; K. Gawronski <sup>5</sup>; K. Hankenson <sup>4</sup>; C.D. Brown <sup>5</sup>; A.D. Wells <sup>1,2</sup>; S.F.A. Grant <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) University of Pennsylvania, Philadelphia, PA, USA; 3) University of Pennsylvania, Department of Biostatistics, Epidemiology & Informatics, Philadelphia, PA, USA; 4) University of Michigan, Ann Arbor, MI, USA.; 5) University of Pennsylvania, Department of Genetics, Philadelphia, PA, USA

---

Genome-wide association studies (GWAS) of total cholesterol levels have identified hundreds of associated loci; however, localization of causal variants at these loci, plus the corresponding effector genes, remain to be fully determined. It has been shown that liver and adipose exchange very low density lipoproteins and free fatty acids, via the vascular system, to form a common fatty acid pool for both storing and distributing energy to other tissues and organs. This suggests that adipocytes and hepatocytes represent key cell types in which to conduct variant to gene mapping for GWAS-implicated total cholesterol loci. We elected to leverage human mesenchymal stem cell (hMSC) derived adipocytes and induced pluripotent stem cells (iPSC) derived hepatocytes to physically implicate effector genes at these loci. We extracted proxy SNPs in LD with each sentinel variant and cross-referenced them with the ATAC-seq-determined open chromatin landscape in both cell types. We determined that 1,044 (2.7%) proxy SNPs in hMSC-derived adipocytes and 835 (2.2%) in iPSC-derived hepatocytes were in open chromatin out of a total of 38,117 proxies. Then leveraging our high-resolution, genome-wide, promoter-focused Capture C approach, we sought to characterize consistent contacts between open-proxy SNPs and candidate effector genes across the two cell types. We identified 33 contacts between open chromatin proxy SNPs and 22 shared gene promoters. Shared effector genes across the two cell types included *GPR146* (with two identical SNP promoter interactions plus other interactions to different proxy SNPs), *SORT1* and *PCBP1* (with interactions to different open proxy SNPs). These results suggest a degree of common genomic etiology influencing total cholesterol levels in both adipocytes and hepatocytes.

## PgmNr 3112: Disrupting early biogenesis peroxins restores growth to mutants of late-acting biogenesis peroxins.

### Authors:

R. Llinas; K. Burks; S. Burkhardt; K. Gonzalez; B. Bartel

[View Session](#) [Add to Schedule](#)

**Affiliation:** Rice University, Houston, TX.

---

Peroxisomes are organelles that house essential oxidative reactions in almost all multicellular eukaryotes. Peroxisome biogenesis disorders (PBDs) are a suite of autosomal recessive congenital disorders that are often lethal within the first year of life, including Zellweger syndrome, Neonatal Adrenoleukodystrophy, and Infantile Refsum Disease. These disorders are caused by mutations in peroxin (*PEX*) genes, which encode the proteins that generate and maintain peroxisomes. Over-accumulation of very-long-chain fatty acids and reduced synthesis of plasmalogen due to defective peroxisomes are hallmarks of PBDs and were often used to diagnosis the disease prior to the development of genomic sequencing tests. As in mammalian cells, plant cells rely on peroxisomes for breakdown of fatty acids stored in lipid droplets, and mutations in peroxin genes confer a range of developmental defects that are easily assayed to monitor peroxisome function in various *Arabidopsis* mutants. We screened for mutant suppressors that restored post-germinative growth to *pex12-1* and *pex6-1*, mutants that display severe growth defects and over-accumulation of lipid droplets stemming from disrupted import of peroxisome luminal proteins. Our independent *pex12-1* and *pex6-1* suppression screens recovered nearly identical early nonsense mutations in *PEX3B* that result in complete loss of protein. The loss of PEX3B protein likely results in reduced but not abolished PEX3 function because a *PEX3A* paralog remains intact. Paradoxically, both *pex3b* mutants markedly improved *pex12-1* or *pex6-1* growth without restoring luminal-protein import, suggesting that previously unrecognized *pex12-1* and *pex6-1* defects contribute to their stunted growth. Elucidating these suppression mechanisms may reveal novel roles for PEX3B in lipid droplet or peroxisome dynamics beyond well-established roles in pre-peroxisome budding from the ER and peroxisomal membrane protein insertion. Moreover, this study highlights PEX3 as a potential drug target for the development of PBD therapy.

## PgmNr 3113: Novel rare variants in *APOC3* and their association with serum triglycerides, diabetes and coronary heart disease.

### Authors:

C. Bejar <sup>1</sup>; S. Ralhan <sup>2</sup>; G.S. Wander <sup>2</sup>; J.R. Singh <sup>3</sup>; N.K. Mehra <sup>4</sup>; D.K. Sanghera <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatrics, University of Oklahoma Health Science Center, Oklahoma City, Oklahoma.; 2) Hero DMC Heart Institute, Ludhiana, Punjab, India; 3) Central University of Punjab, Bathinda, Punjab, India; 4) All India Institute of Medical Sciences and Research, New Delhi, India

Elevated serum triglyceride (TG) concentration is heritable and a strong independent risk factor for coronary heart disease (CHD). Apolipoprotein CIII (ApoC3) is an important regulator of lipid metabolism and plays a key role in regulating plasma triglycerides. Elevated plasma levels of ApoC3 are correlated with the increased risk of CHD. Recent exome sequencing studies on populations of European and African ancestry have identified rare loss of function (LOF) variants associated with reduced triglycerides and decreased risk of CHD. However, no such information is available in any population of South Asian (SA) ancestry despite the fact that SAs constitute ~1/4<sup>th</sup> of the global population and contribute the highest proportion of CHD burden compared to any other region in the world.

Here we performed targeted resequencing of 13 candidate genes of dyslipidemia (including *APOC3*) to identify functional variants underlying their biological functions. We sequenced 820 Sikhs from India with extreme hypertriglyceridemia (HTG) levels and used 2,239 samples for replication; all 3,059 participants were part of the Asian Indians Diabetic Heart Study. Many rare and loss of function variants identified earlier in Europeans and Africans including R19X were not present in our sample. However, our data revealed 23 novel rare variants (MAF <0.001) including 5 variants in 5'UTR, 3'UTR, insertion or deletions that were not present in any public database. Mean plasma TG levels were lower 57.8% among carriers of any of these variants compared to non-carriers ( $p=3.8 \times 10^{-5}$ ). Similarly, HDL levels were elevated 19.8% and LDL was reduced 12.8% among carriers compared to non-carriers. Collectively, these results highlight the importance of genetic evaluation of diverse ancestries for identifying clinically important therapeutic targets.

**Funding Acknowledgements:** The Sikh Diabetes Study/ Asian Indian Diabetic Heart Study was supported by NIH grants-R01DK082766 (NIDDK) and NOT-HG-11-009 (NHGRI) and grants from Presbyterian Health Foundation. Sequencing services were provided through the RS&G Service by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under U.S. Federal Government contract number HHSN268201100037C from the National Heart, Lung, and Blood Institute of the NIH.



## PgmNr 3114: Longitudinal multi-omic analyses of whole blood gene expression and metabolite levels in the MultiMuTHER study reveal differences in longitudinal variation patterns between 'omics.

### Authors:

J.S. El-Sayed Moustafa <sup>1</sup>; Y. Jiao <sup>2,3,4</sup>; G. Nicholson <sup>4</sup>; G. Leday <sup>5</sup>; M. Stevens <sup>6,7</sup>; M. Abdalla <sup>2,3,4</sup>; A. Ramisch <sup>6,7</sup>; C. Menni <sup>1</sup>; C. Holmes <sup>4</sup>; T.D. Spector <sup>1</sup>; M.I. McCarthy <sup>2,3,8</sup>; S. Richardson <sup>5</sup>; E.T. Dermitzakis <sup>6,7</sup>; K.S. Small <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 2) Wellcome Centre Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK; 3) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, OX3 7LE, UK; 4) Department of Statistics, University of Oxford, Oxford, OX1 3TG, UK; 5) Medical Research Council Biostatistics Unit, Cambridge, United Kingdom; 6) Institute of Genetics and Genomics of Geneva, Geneva, Switzerland; 7) Department of Genetic Medicine and Development, Faculty of Medicine, University of Geneva, Geneva, Switzerland; 8) Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, OX3 7LE, UK

The body is a dynamic system of interacting subunits whose combined influences define its phenotypic trajectories. Multi-omic datasets represent a snapshot of the system's physiological state, integrative analysis of which may enable a precision medicine approach to health monitoring and earlier identification of disease risk. Critically, to identify deviation on an individual-level indicative of adverse outcomes, we must first understand the extent of normal variation over time in individuals and on a population level. We therefore sought to investigate how gene expression and metabolite levels in blood change over time, and what genetic or environmental factors drive these changes.

To address this, we have generated gene expression (RNASeq) and metabolomics (1,197 metabolites) data in whole blood at three or more timepoints per individual from 335 TwinsUK subjects [age range 32-80; median=61yrs]. TwinsUK is a deeply-phenotyped cohort of twins with extensive 'omics data and repeat phenotypic measures. Time between first and last visit ranges from 3 to 8 years (median 6yrs).

We observed moderate correlation in metabolite and gene expression profiles within individuals across timepoints (median rho-metabolites=0.44; median rho-gene expression=0.28). Intriguingly, correlation of metabolomic profiles decreased with increased time between sampling (median rho Timepoint1-3=0.33;  $P < 2.2 \times 10^{-16}$ ), whereas correlations in gene expression profiles showed stability over time, with no significant reduction in correlation comparing two timepoints longest apart versus those closer together ( $P > 0.05$ ).

Expression levels of ~30% of genes showed significant longitudinal associations (FDR 5%), including genes previously associated cross-sectionally with age such as *ACAA2* and *CD248*, as well as *CASP2*, previously associated with premature ageing in Caspase-2 deficient mice. ~12% of metabolites

showed significant longitudinal associations ( $P < 5 \times 10^{-5}$ ), including classes indicative of both environmental and physiological changes over time. Significant metabolites included environmental toxins, as well as androgenic steroids including dehydroisoandrosterone sulfate (DHEA-S). Metabolite associations were replicated in a sample of 2,000 twins with three timepoints per subject.

Integrative multivariate analyses of gene expression and metabolomic data, together with whole genome sequencing, longitudinal phenotypic data and disease outcomes in TwinsUK are also currently underway.

## PgmNr 3115: Revisiting the start codons for mitochondrial *MT-ATP6*, *MT-ND1* and *MT-ND5* genes.

### Authors:

H. Cui; K. Haude; G. Richard; R.K. Bai

[View Session](#) [Add to Schedule](#)

**Affiliation:** Gene Dx, Gaithersburg, Maryland.

The human mitochondrial genome encodes 13 structural subunits of the mitochondrial respiratory chain, 22 tRNAs, and 2 rRNAs. All respiratory chain complexes with the exception of complex II contain components encoded by mitochondrial DNA. Approximately 8000 variants have been reported as sequence variants in the 13 coding genes; among these, 10 unique variants alter non-wobble positions within the start codons of three genes: *MT-ATP6*, *MT-ND1*, and *MT-ND5*. In total, these variants were reported in 746 individuals (PMID: 25489354; <http://www.mitomap.org>). In conjunction with our data from diagnostic mitochondrial genome sequencing, these findings raise doubt on the validity of the assigned start codons of *MT-ATP6*, *MT-ND1* and *MT-ND5* in the Revised Cambridge Reference Sequence (rCRS, GenBank #NC\_012920.1).

A homoplasmic m.8527 A>G variant in *MT-ATP6* has been reported in a patient with suspected mitochondrial disorder (PMID: 14697245). This variant changes the start codon from AUG to GUG (normally coding for valine). However, it was also observed in the healthy mother of the reported patient. This change is seen in two non-anthropoid species and has a global minor allele frequency of 0.45% (208/46,092) (PMID: 25489354, 28510580). We have observed m.8527 A>G in 162 homoplasmic individuals, most of whom are either asymptomatic or have phenotypes that are atypical of mitochondrial disorders. However, variant m.8528 T>C affecting the second nucleotide of the start codon has been published in multiple patients (PMID: 19188198, 26803244) and seen at our lab in a patient with suspected mitochondrial disorder. The next downstream methionine (Met) is distant at p.57, with the second nucleotide of this codon being polymorphic. This cumulative evidence suggests the resulting GUG codon at p.1 is a functional start codon.

In contrast, the start codons of *MT-ND1* and *MT-ND5* are close to the second Met located at p.3. Our lab observed 6 distinct nucleotide changes in these start codons in 629 homoplasmic individuals, including healthy adults. No changes at the non-wobble position of the p.Met3 codon have ever been reported or detected at high heteroplasmy/homoplasmy in our lab. Based on this observation, we suspect that the p.Met3 is more likely to represent the real start codon for the *MT-ND1* and *MT-ND5* genes. Further functional studies and clinical correlation are warranted to fully elucidate the start codon of these two genes.

## PgmNr 3116: The molecular mechanisms by which the *MTTL1* variant mt.3250T>C causes mitochondriopathy.

### Authors:

Lou <sup>1,2</sup>; T. Campbell <sup>1,3</sup>; J. Slone <sup>1</sup>; J. Brown <sup>1</sup>; M. Bromwell <sup>1</sup>; R. Bai <sup>4</sup>; T. Huang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, Cincinnati Children's Hospital, Cincinnati, OHIO, USA; 2) Key Laboratory of Laboratory Medicine, Ministry of Education, School of Laboratory Medicine and Life sciences, Wenzhou Medical University, Wenzhou, Zhejiang, China.; 3) College of Medicine, University of Cincinnati, Cincinnati, Ohio, USA; 4) GeneDx, Gaithersburg, Maryland, USA

---

Mitochondrial diseases can be caused by pathogenic mutations in the mitochondrial or the nuclear genome. The *MTTL1* gene is located in the mitochondrial genome (mtDNA) and encodes the mitochondrial tRNA-leucine. The variant mt.3250T>C in the *MTTL1* gene has been previously associated with mitochondrial dysfunction and mitochondriopathy in three families. Males and females were reported to be affected equally, with complete penetrance. Despite efforts to document the phenotypic presentation associated with this variant, the molecular mechanism resulting in disease is yet to be established. Here, we have identified a large, multigenerational family with the mt.3250T>C variant. Next generation sequencing (NGS) was performed on mtDNA from ten individuals (both affected and unaffected) to quantify mt.3250T>C heteroplasmy load. Fibroblast samples were collected from the affected individuals for mitochondrial function studies. In parallel, additional families with the mt.3250T>C variant have been collected and detailed phenotypes were analyzed. In the large, multigenerational family, we identified nine individuals with a medical history of exercise intolerance, chronic fatigue, muscle weakness and elevated lactate levels. Two members of this family passed from sudden cardiomyopathy in early adulthood. Surprisingly, all affected individuals in this family are males, with all female carriers being asymptomatic despite several of them demonstrating homoplasmic or near homoplasmic levels of the mt.3250T>C variant. Additional studies are currently in progress to investigate the molecular mechanisms underlying this sex-based difference in phenotypic penetrance as well as the mechanisms by which the mt.3250T>C variant causes mitochondriopathy.

## PgmNr 3117: A genome-wide mitochondria sequencing identified SNP that alters its inner membrane potential increases risk for fibromyalgia.

### Authors:

J. Smith-Voudouris <sup>1</sup>; G. Drury <sup>1</sup>; V. Verma <sup>1</sup>; A.J. Chabot-Doré <sup>1</sup>; M. Parisien <sup>1</sup>; A.G. Nackley <sup>3</sup>; S.B. Smith <sup>3</sup>; M.A.L. van Tillburg <sup>4</sup>; R.G. Boles <sup>5</sup>; W.E. Whitehead <sup>4</sup>; R.H. Gracely <sup>2</sup>; J.D. Mann <sup>6</sup>; W. Maixner <sup>3</sup>; L. Diatchenko <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Alan Edwards Centre for Research on Pain, McGill University, Montreal, QC, Canada; 2) Center for Pain Research and Innovation, University of North Carolina School of Dentistry, Chapel Hill, NC, United States; 3) Department of Anesthesiology, Center for Translational Pain Medicine, Duke University, Durham, NC, United States; 4) Center for Functional Gastrointestinal and Motility Disorders, Division of Gastroenterology and Hepatology, Department of Medicine, University of North Carolina, Chapel Hill, NC, United States; 5) Courtagen Life Sciences, Inc., Woburn, MA, United States; 6) Department of Neurology, Program on Integrative Medicine, University of North Carolina, Chapel Hill, NC, United States

---

Chronic pain (CP) is defined by the International Association for the Study of Pain as pain without apparent biological value that has persisted beyond normal tissue healing time. CP is estimated to affect 20% of U.S. adults; it affects quality of life, productivity, rates of workplace absenteeism, and has a greater economic impact than other health conditions. It is thus important to understand the progression of pain, and identify those at risk of developing CP. Previous research has unveiled that mitochondria play a role in the acute to chronic pain transition. We sought to further investigate the relationship between mitochondria genome and CP in a cohort of human subjects with at least one of five Complex Persistent Pain Conditions (CPPC): episodic migraines, irritable bowel syndrome, vulvar vestibulitis, temporomandibular disorders, and fibromyalgia. This cohort comprises representative US racial groups, ages 18-64, 85% females, n=752. High-coverage (>100x) mitochondria DNA sequencing from whole blood fractions was performed to determine allele content at each genomic position. We conducted a mitochondria-wide association study on CPPC using PLINK with sex, age, and self-declared race as co-variables. SNP rs28358579 was found significantly associated with fibromyalgia in women after Bonferroni correction ( $P=4.0 \times 10^{-3}$ , OR=2.9, C allele, MAF=6%). This result was replicated in the OPPERA cohort ( $P=2.8 \times 10^{-2}$ , OR=4.2, C allele, MAF=7%). The SNP is located in mitochondrial 16S rRNA and is not listed yet in the OMIM and ClinVar catalogs. In African-Americans the minor C allele for SNP rs28358579 appears at about 20%, when in Caucasians is almost non-existent. We explored the effect of this SNP in immortalized B lymphoblast cell lines made available from the 1,000 genomes project of African-American ancestry subjects. To test the ability of mitochondria to utilize oxidative phosphorylation we measured their inner mitochondrial membrane potential. A total of 20 samples from unique individuals, where half of them had the risk allele, were screened using JC-10 staining followed by 2D flow cytometry. We found that a distinct population of live cells with lower potential was significantly more abundant in samples that carried the C allele. This indicated that mitochondrial function was impaired in cells with the risk allele. Further investigation on the link between CPPC and mitochondrial membrane potential is under way.

## PgmNr 3118: High throughput identification of MCCRP2 genetic modifiers using *Caenorhabditis elegans*.

### Authors:

T. Maroilley<sup>1,2</sup>; S. Stasiuk<sup>1,2</sup>; K.H.T. Hassan Rahit<sup>1,2</sup>; M. Meode<sup>1,2</sup>; K.F. O'Connell<sup>3</sup>; M. Tarailo-Graovac<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Departments of Biochemistry, Molecular Biology and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada; 2) Alberta Children's Hospital Research Institute, University of Calgary, Calgary, AB, Canada; 3) Laboratory of Biochemistry and Genetics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 8 Center Drive, Bethesda, MD, United States of America

---

Microcephaly and chorioretinopathy 2 (MCCRP2) is a rare genetic disorder (RGD) due to biallelic loss of function variants in the *PLK4*, a gene that encodes a key regulator of centriole biogenesis. Centrioles are conserved, microtubule-derived organelles, involved in a number of essential processes, such as cell division (centrosomes) and sensation/signaling/motility (cilia). The first five centriole components were identified in *C. elegans*, and their orthologs were later shown to be conserved in other systems, including the *C. elegans* *PLK4* ortholog, ZYG-1 (ZYGote defective/embryonic lethal), a serine/threonine kinase. To establish our high throughput genomics approach to identification of genetic modifiers of MCCRP2, we used the *C. elegans* *zyg-1* mutant strain that results in complete embryonic lethality at restrictive temperature (24.5°C). We performed unbiased forward genetic mutagenesis screens using two different mutagens (e.g. ethyl methanesulfonate and N-ethyl-N-nitrosourea) leading to randomly located predominantly single nucleotide variants. Thus far, we have isolated more than 100 genetic modifiers that suppress the embryonic lethality of the *zyg-1/PLK4* mutant. To expedite the molecular identification of suppressors, we utilized Whole Genome Sequencing (WGS) followed by in-house bioinformatics approaches to call and prioritize the variants according to their effect on genes and biological functions. We show that our approach allowed us to identify candidate suppressor variants in the majority of WGS-tested suppressor genomes shortly after sequencing (~2-4 hours), greatly increasing efficiency over standard approaches. The candidate suppressors are further being validated by Sanger sequencing, segregation analyses and in some instances, CRISPR/Cas9. The majority of the suppressor genomes have a single genetic modifier; however, a subset of genomes were isolated with more complex genetic interactions, where more than one modifier gene was altered by mutagenesis. Notably, the vast majority of the identified modifier genes are conserved in humans, supporting the potential for translatability of our findings to patients with MCCRP2. Importantly, our high-throughput approach to understanding genetic modifiers shows a potential to accelerate the identification of genes and genetic networks not previously implicated in centriole function and thus RGDs like microcephaly, which in turn may enhance diagnosis, prognosis or provide new drug-able targets.

## PgmNr 3119: Germline-activating *RRAS2* mutations cause Noonan syndrome.

### Authors:

T. Niihori <sup>1,2</sup>; K. Nagai <sup>1</sup>; A. Fujita <sup>3</sup>; H. Ohashi <sup>4</sup>; N. Okamoto <sup>5</sup>; S. Okada <sup>6</sup>; A. Harada <sup>7</sup>; H. Kihara <sup>8</sup>; T. Arbogast <sup>2</sup>; R. Funayama <sup>9</sup>; M. Shirota <sup>10</sup>; K. Nakayama <sup>9</sup>; T. Abe <sup>1</sup>; S. Inoue <sup>1</sup>; I.C. Tsai <sup>2</sup>; N. Matsumoto <sup>3</sup>; E.E. Davis <sup>2</sup>; N. Katsanis <sup>2</sup>; Y. Aoki <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept of Med Gene, Tohoku Univ Sch of Med, Sendai, Japan; 2) Center for Human Disease Modeling, Duke Univ Med Center, Durham, NC; 3) Dept of Hum Genet, Yokohama City Univ Grad Sch of Med, Yokohama, Japan; 4) Div of Med Genet, Saitama Children's Med Center, Saitama, Japan; 5) Dept of Med Genet, Osaka Women's and Children's Hosp, Osaka, Japan; 6) Dept of Ped, Hiroshima Univ Grad Sch of Biomed & Health Sci, Hiroshima, Japan; 7) Dept of Ped Neurosurg, Takatsuki General Hosp, Osaka, Japan; 8) Dept of Ped, Onomichi General Hosp, Hiroshima, Japan; 9) Dept of Cell Proliferation, United Center for Adv Research and Translational Med, Tohoku Univ Grad Sch of Med, Sendai, Japan; 10) Div of Interdisciplinary Med Sci, United Center for Advanced Res and Translational Med, Tohoku Univ Grad Sch of Med, Sendai, Japan

---

Noonan syndrome (NS) is characterized by distinctive craniofacial appearance, short stature, and congenital heart disease. Approximately 80% of individuals with NS harbor mutations in genes whose products are involved in the RAS/mitogen-activating protein kinase (MAPK) pathway. However, the underlying genetic causes in nearly 20% of individuals with NS phenotype remain unexplained. We identified four de novo *RRAS2* variants in three individuals with NS. *RRAS2* is a member of the RAS subfamily, and is ubiquitously expressed. Three variants, c.70\_78dup; p. Gly24\_Gly26dup, c.216A>T; p. Gln72His, and c.215A>T; p. Gln72Leu have been found in cancers; our functional analyses showed that these three changes induced elevated association of RAF1 and that they activated ERK1/2 and ELK1. Notably, prominent activation of ERK1/2 and ELK1 by p. Gln72Leu associates with the severe phenotype of the individual harboring this change. To examine variant pathogenicity *in vivo*, we generated zebrafish models. Larvae overexpressing c.70\_78dup; p. Gly24\_Gly26dup or c.216A>T; p. Gln72His variants, but not wild-type *RRAS2* RNAs, showed craniofacial defects and macrocephaly. The same dose injection of mRNA encoding c.215A>T; p. Gln72Leu caused severe developmental impairments and low dose overexpression of this variant induced craniofacial defects. In contrast, the *RRAS2* c.224T>G; p. Phe75Cys change, located on the same allele with p. Gln72His in an individual with NS, resulted in no aberrant *in vitro* or *in vivo* phenotypes by itself. Together, our findings suggest that activating *RRAS2* mutations can cause NS and expand the involvement of *RRAS2* proto-oncogene to rare germline disorders. The discussion will be made including recent findings from other groups.



## PgmNr 3120: A possible crosstalk disorder between BMP15 and TGF- $\beta$ family ligands and receptors in Mayer-Rokitansky-Küster-Hauser syndrome.

### Authors:

B. Abdelmoula; R. Louati; W. Smaoui; K.H. Trigui; F. Abid; S. Kammoun; O. Kaabi; N. Abdellaoui; K. Chaker; H. Fourati; S. Aloulou; H. Yahia; M.H. Ben Ayed; A. Sbair; S. Fellah; S. Ben Ali; H. Chaari; A. Yaich; A. Kaabi; S. Mrabet; M. Ben Aziza; J. Jabr; S. Mallek; W. Kammoun; W. Ltaif; J. Dammak; H. Ben Ameer; N. Gouiaa; M.N. Mhiri; N. Bouayed Abdelmoula

[View Session](#) [Add to Schedule](#)

**Affiliation:** UR17ES36 Genomics of Signalopathies at the service of Medicine, Medical University of Sfax, Sfax, Tunisia

---

Mayer-Rokitansky-Küster-Hauser syndrome (MRKH) is a rare congenital disease (1/4500 female births) characterized by aplasia of the uterus and upper vagina, in young women with normal secondary sex characteristics and a normal female chromosome formula. Primary amenorrhoea is often indicative of the syndrome. MRKH syndrome can be isolated (type I). However it is most often associated with renal, vertebral, and, less frequently, auditory and cardiac abnormalities (MRKH type II or MURCS association).

Here, we report three clinical observations of girls with primary amenorrhea and for whom ultrasound revealed the absence or very rudimentary appearance of the uterus while the hormonal ovarian function was almost normal. The dysmorphological consultation revealed skeletal anomalies for all patients. The cytogenetic study revealed normal female chromosome formula (46, XX) whereas molecular study by direct sequencing of the two exons of the BMP15 gene, involved in early ovarian insufficiency, revealed the presence of a c.788insTCT / c.852C-T polymorphism of exon 2 of the BMP15 gene in the heterozygous state.

MRKH syndrome has often been reported as sporadic but familial forms have been reported in 20 percent of cases. The pathogenic mechanism of MRKHS may involve gene abnormalities, and there are various case reports associating MRKHS with the Wnt4 gene mutations. Analysis of genes mapped to regions in which deletion and duplication are frequently detected in patients with MRKHS has shown involvement of LHX1, HNF1B and TBX6 genes. We can suggest through our results a possible association between development of female reproductive tract, MRKHS and BMP15 polymorphisms. This association may involve a pathway crosstalk disorder between BMP15 and TGF- $\beta$  family ligands and receptors having many roles in primordial germ-cell development, sexual differentiation, and gonadal cell development.



## PgmNr 3121: VNTRs mediate the expression of genes involved in neurological disorders and familial cancers.

### Authors:

M. Bakhtiari; V. Bafna

[View Session](#) [Add to Schedule](#)

**Affiliation:** Computer Science and Engineering, University of California San Diego, La Jolla, CALIFORNIA.

---

**Background.** Variable Number Tandem Repeats have been implicated in multiple Mendelian and Complex disorders (Alzheimer's, Bipolar, familial cancer), usually due to repeating VNTR units regulating gene expression. While similar to short tandem repeats (STRs), VNTRs are harder to genotype due to larger and highly variant repeating units that confound traditional pipelines. Generative methods such as adVNTR (Bakhtiari, Genome Research 2018) now estimate genotypes accurately, but are computationally intensive, and cannot scale to computing genome-wide association with gene-expression and other quantitative traits.

**Method.** We developed a number of algorithmic and statistical strategies to identify VNTR genotypes at scale. Importantly, we designed neural networks that encoded each read based on its k-mer composition, and either assigned the read to a VNTR from a targeted list or discarded it. The filtered reads for each VNTR were then genotyped using adVNTR. Next, we assessed the effect of VNTR genotypes on expression of proximal genes in 27 tissues (GTEx data) after correction for technical co-variates, population structure, and sex of individuals. We used analysis of variance (ANOVA) to show that VNTR genotypes explain more variations in gene expression than nearby SNVs, as well as fine-mapping tool CAVIAR to confirm the causality of each variant, and determine if the VNTR directly mediates expression of the proximal gene in a specific tissue.

**Result.** We chose 13645 VNTRs from promoters, UTRs, and exons that could be genotyped using short-reads. Neural network filtering prior to genotyping using HMMs allowed us to genotype 100 individuals from the GTEx project in 10833 cpu-hours. We identified 435 VNTRs (VNTR eQTL) in 27 tissues with significant association between number of repeats and gene expression ( $p$ -value $<0.05$ ). These candidates were further validated by selecting for the lowest  $p$ -value among all SNPs within 100kb (320 VNTRs), explaining more variation than the top 20 eQTLs (301 variants), and being top-ranked in a statistical fine-mapping analysis of all variants using CAVIAR (298 VNTRs). To our knowledge, only four of the top 30 had previously been identified as causal variants using experimental methods, including *AS3MT* (Li, Nature Medicine, 2016), *HMGA2*, *KCNQ1OT1*, and *TERT*. The identified VNTR eQTLs mediate the expression of many genes with known associations to neurological disorders and familial cancers, illustrating the power of our approach.

## PgmNr 3122: A novel *SHH* variant associated with holoprosencephaly-3 modeled in *Xenopus*.

### Authors:

E. Mis <sup>1</sup>; L. Jeffries <sup>1</sup>; W. Ji <sup>1</sup>; J. McGrath <sup>2</sup>; J. McGlynn <sup>3</sup>; M. Konstantino <sup>1</sup>; M. Khokha <sup>1,2</sup>; S. Lakhani <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Pediatric Genomics Discovery Program, Department of Pediatrics, Yale University School of Medicine, New Haven, CT; 2) Department of Genetics, Yale University School of Medicine, New Haven, CT; 3) Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, Yale University School of Medicine

---

Holoprosencephaly-3 (HPE-3) is caused by loss-of-function variants in *SHH* and has wide clinical features. Even within families, abnormalities range from normal brain structure to presence of a single ventricle with no interhemispheric fissure and other midline birth defects. However, incomplete penetrance and the inaccessibility of variant-specific functional assays often result in inconclusive genetic diagnoses, especially in the case of inherited missense variants. We used whole exome sequencing to identify a novel variant, R310P, in a fetus with alobar HPE and the unaffected mother, whose mother was determined also to have the variant via Sanger sequencing. To assess the functional effects of R310P of SHH protein, we performed *in vivo* overexpression of WT SHH in *Xenopus tropicalis* and compared this to overexpression of SHH harboring the R310P variant. We analyzed *pax2* and *pax6* expression, as these are regulated by SHH, and found no change in *pax2* or *pax6* expression upon overexpression of the R310P variant, indicating a loss of normal SHH function. Additionally, we overexpressed other previously reported variants to test the accuracy and consistency of this assay. We aim to provide evidence that a *SHH* variant identified in a single family is causative for HPE-spectrum findings over three generations. We also test our hypothesis that *Xenopus* is an effective model for functional characterization of *SHH* variants.

## PgmNr 3123: Mining GWAS hits for therapeutic targets: *In vitro* validation of grip strength modulating genes using cell painting assays for phenotypic screening.

### Authors:

D. Waterworth <sup>1</sup>; R.A. Scott <sup>2</sup>; D. Rajpal <sup>3</sup>; D. Jhamb <sup>3</sup>; C. Watson <sup>4</sup>; R. Merrihew <sup>4</sup>; M. Truong <sup>5</sup>; M. Harpel <sup>4</sup>; D. Pertusi <sup>5</sup>; D. Cooper <sup>6</sup>; K. Brown <sup>5</sup>; L. Kallal <sup>5</sup>; J.-L. Klein <sup>4</sup>; L. Sarov-Blat <sup>1</sup>; A. Hinken <sup>7</sup>; J. Kerr <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics, GlaxoSmithKline, Collegeville, Pennsylvania.; 2) Human Genetics, GlaxoSmithKline, Stevenage, UK; 3) Computational Biology, GlaxoSmithKline, Collegeville, Pennsylvania.; 4) Novel Human Genetics, GlaxoSmithKline, Collegeville, Pennsylvania.; 5) Medicinal Science and Technology, GlaxoSmithKline, Collegeville, Pennsylvania.; 6) Research Statistics, GlaxoSmithKline, Collegeville, Pennsylvania.; 7) Innate Immunity RU, GlaxoSmithKline, Collegeville, Pennsylvania.

---

Frailty is a significant component of morbidity in old age and a predictor of mortality, although etiology remains unclear. To gain insight into the genetic influences on frailty, we previously performed a GWAS of grip strength (n=448,861) in UK Biobank identifying 208 significant loci and then used a variety of variant-to-gene (V2G) mapping methods (COLOC, IDEASpredict and functional coding variants) to identify putative effector genes at those loci. We identified 230 genes with varying degrees of mapping confidence and set out to use *in vitro* assays in relevant cell types to further resolve effector genes and validate those with potential to augment muscle size and function. Ideally, a good target for muscle frailty of aging would be a gene that generates more mitochondria, enhances growth, supports moderate autophagy and diminishes reactive oxidative stress (ROS) in muscle. To identify targetable genes from our GWAS list, we employed two strategies: first, we generated a 360-compound small molecule library through network analysis to identify interacting proteins, and secondly, we developed a lentiviral shRNA library to knockdown all 230 potential effector genes. We then utilized cell painting assays in both human muscle myotubes and U2OS cells to elucidate the impact of compound and genetic manipulations on mitochondrial morphology, cell growth and size, autophagy and oxidative stress.

Using multiparametric image analysis and Sammon projection mapping to cluster cellular manipulations by phenotypic similarity, we can build upon the GWAS identified target genes to assign biological context and develop integrated pathway maps. For example, an mTOR inhibitor from our compound library resulted in changes to cell morphology (spindly vs round shape), mitochondrial network distribution (punctate), and an increase in the golgi and endosomal network. Following analysis, we identified a second mTOR inhibitor as well as an inhibitor for another gene in the same pathway (S6Kα) that had similar phenotypic changes. This assay and analysis paradigm allow us to define a beneficial cell profile and mine the assay results for genetic manipulations with phenotypic similarity, delineate targets or compounds on the same pathways, and ultimately identify the best targets for therapeutic intervention.

## PgmNr 3124: Transcriptome analysis of single cell RNA-seq from Alzheimer disease admixture block brains: Differences in *APOE* e4 carriers risk conferred to local ancestry.

### Authors:

K. Celis<sup>1</sup>; P. Bussies<sup>1</sup>; A. Griswold<sup>1</sup>; F. Rajabli<sup>1</sup>; P. Whitehead<sup>1</sup>; K. Hamilton-Nelson<sup>1</sup>; J. Lorenzo<sup>1</sup>; G. Beecham<sup>1</sup>; E.H. Bigio<sup>2</sup>; M. Mesulam<sup>2</sup>; S. Weintraub<sup>2</sup>; C. Geula<sup>2</sup>; M. Gearing<sup>3</sup>; W. Scott<sup>1</sup>; M.A. Pericak-Vance<sup>1</sup>; J.M. Vance<sup>1</sup>; J. Young<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics (IHG), Univ Miami, Miami, Florida.; 2) Northwestern University Feinberg School of Medicine, Chicago, IL; 3) Emory University, Atlanta, GA

---

**Background:** Alzheimer Disease (AD) is the most common neurodegenerative cause of dementia. The apolipoprotein E epsilon 4 (*APOE4*) allele is the strongest genetic risk factor for late-onset Alzheimer Disease (LOAD). However, this risk for AD from *APOE4* is not homogenous between racial backgrounds: Asians > Non-Hispanic Whites (NHW) > African-American (AA). We have previously shown (Rajabli et al 2018) this difference is due to local ancestry surrounding the *APOE* gene. As gene expression can be difficult to interpret due to the cellular heterogeneity of the brain, single nucleus RNA-seq (snRNA-seq) allows us to evaluate cell-specific transcriptional changes that may accompany these risk variation from *APOE4* carriers.

**Methods:** We compared SnRNA-seq of frontal cortex (Brodmann area 9) from brains of homozygote African or NHW local ancestry *APOE4* carriers. Isolation of the nuclei was performed using a sucrose gradient protocol, followed by SnRNA-seq (10X Chromium platform) with analysis using Seurat for potential differences in expression between the samples. For cell-type identification, we used computational pipelines to align the raw sequencing data to a reference human transcriptome including intronic sequences and derive a gene expression matrix of ~22,000 genes across the cells.

**Results:** We obtained data from 5000-8000 nuclei sequenced at a median depth of ~120,000 reads per cell. We detected on average ~800 genes/cell. Transcriptome based resolution of nuclei resulted in similar clusters for African and NHW samples, classified based on previously established cell type specific gene sets as neurons (excitatory and inhibitory), oligodendrocytes, astrocytes, microglia, oligodendrocyte progenitors, and smaller subsets of endothelial and pericyte nuclei. We observed cluster-specific differential expression of *APOE* and differential proportion of a microglia sub-cluster, when comparing African vs NHW samples.

**Conclusions:** SnRNA-seq allows isolation of specific cell types expressing *APOE*, facilitating differential analysis between the *APOE4* local ancestry. This study is unique in its transcriptomic ancestral-specific focus in the identification of molecular pathways playing a role in AD pathogenesis.

## PgmNr 3125: Functional analysis of an Alzheimer disease-associated rare variant in SORL1.

**Authors:**

D.M. Dykxhoorn; C. Chambers; M.L. Cuccaro; J.M. Vance; H.N. Cukier; M.A. Pericak-Vance

[View Session](#) [Add to Schedule](#)

**Affiliation:** Hussman Institute for Human Genomics, University of Miami, Miami, Florida.

---

Alzheimer's disease (AD) is the most common neurodegenerative disease. In 2015, of the ~47 million individuals with dementia worldwide, AD accounted for 50-70% of these cases. The fundamental mechanisms underlying AD are not fully understood but result from both genetic and environmental components. Over 20 genetic loci have been consistently associated with AD with accumulating evidence points to alterations of the endolysosomal pathway as playing key roles in AD. The sortilin-related receptor 1 (SORL1) gene has been associated with AD through multiple genetic studies and has been suggested to function by guiding APP to the endocytic pathway. We have recently identified a family with multiple AD affected individuals that bear a single base pair deletion (g.4292delG) that is predicted to result in a frameshift alteration and early termination of the protein (p.Cys143Ser\*1). Induced pluripotent stem cells (iPSC) lines were developed to examine the cellular and molecular consequences in neuronal and glial cell (astrocytes and microglia) populations, which can otherwise only be collected postmortem. Each iPSC line generated was validated for pluripotency through immunocytochemical staining and shown to be negative for any large-scale chromosomal abnormalities via karyotyping. In addition, we have generated control iPSC lines that have a single copy of the SORL1 gene knocked out. The patient iPSC lines, the SORL1 knockout iPSC lines, and control lines were differentiated into cortical neurons and glia cells. No difference in the cellular differentiation and growth were observed for the SORL1 cells compared to the control lines. The SORL1 variant bearing and SORL1 genome edited glia cells showed deficits in fibrillary Amyloid beta uptake, while the SORL1 neurons showed elevated production of Ab42 levels. These results support the hypothesis that SORL1 plays a role in endolysosomal trafficking and APP processing.

## PgmNr 3126: Genetic variants control RNA editing rates in Alzheimer disease in diverse populations.

### Authors:

O.K. Gardner<sup>1</sup>; L. Wang<sup>1</sup>; D. Van Booven<sup>1</sup>; P.L. Whitehead<sup>1</sup>; K.L. Hamilton-Nelson<sup>1</sup>; L.D. Adams<sup>1</sup>; T.D. Starks<sup>2</sup>; N.K. Hofmann<sup>1</sup>; J. Lorenzo<sup>1</sup>; J.M. Vance<sup>1,3</sup>; M.L. Cuccaro<sup>1,3</sup>; E.R. Martin<sup>1,3</sup>; G.S. Byrd<sup>2</sup>; J.L. Haines<sup>4,5,6</sup>; W.S. Bush<sup>4,5</sup>; G.W. Beecham<sup>1,3</sup>; M.A. Pericak-Vance<sup>1,3</sup>; A.J. Griswold<sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, Florida.; 2) Maya Angelou Center for Health Equity (MACHE) / Wake Forest School of Medicine, Winston Salem, NC, USA; 3) The Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL, USA; 4) Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA; 5) Cleveland Institute for Computational Biology, Cleveland, OH, USA; 6) Case Western Reserve University School of Medicine, Cleveland, OH, USA

---

RNA editing, the post-transcriptional modification of RNA bases, is a regulatory feature modulating a variety of physiological processes. We have shown significant differences in the RNA editing rates between Alzheimer disease (AD) cases and cognitively intact controls across diverse ancestral backgrounds, at several specific genomic sites, including some in AD candidate genes (Gardner 2019). Editing QTLs (edQTLs) are genetic variants that influence the rate of RNA editing at a specific site. Whether underlying edQTLs are responsible for a portion of the observed differences in editing AD is unknown. Thus, we examined the correlation of DNA variants with RNA editing levels at sites that are differentially edited between AD and controls. Since genetic effects can differ across ancestries, we tested for edQTLs in both African-American (AA) and non-Hispanic White (NHW) populations.

We performed genome-wide edQTL analysis using whole transcriptome RNA sequencing from peripheral blood of 216 AD cases (105 AA, 111 NHW) and 212 cognitively intact controls (105 AA, 107 NHW), all over age 65. REDIttools software was used to identify editing events and a logistic regression model to identify differentially edited sites. DNA for the same subjects was genotyped on the Infinium Global Screening Array. edQTL mapping was performed using FastQTL adjusting for appropriate covariates. Associations with false discovery rate (FDR) less than 1% were considered significant.

We identified 34,770 and 24,918 edQTLs in AA and NHW, respectively. Interestingly, only 1,810 variants were edQTLs in both ethnicities. A subset of these edQTLs was associated with at least one RNA editing site differentially edited between AD cases and controls. These include edQTLs for sites located in known GWAS implicated AD candidate genes, including *SORL1* and members of the HLA and SLC gene families. In addition to known genetic candidates, we find evidence for genetic regulation of sites in functional AD candidates (e.g. *GAS6*). edQTLs for these candidates appeared at significantly different allele frequencies in cases and controls (Fisher's Exact Test,  $p < 0.05$ ).

Sites in genetic and functional candidate genes have edQTLs that differ in frequency in cases and controls. Furthermore, we found significant differences between edQTLs in AA and NHW, suggesting

ancestry specific background effects of edQTLs. These results imply that edQTLs may modulate functional outcomes of some genetic risk factors for AD.

## PgmNr 3127: Cerebrospinal fluid high-throughput proteomics and Mendelian randomization analyses infer a causal role of CD33 on Alzheimer's disease.

### Authors:

C. Yang<sup>1,2,3,4</sup>; F. Farias<sup>1,2,3,4</sup>; K. Mihindukulasuriya<sup>1,2,3,4</sup>; R. Davenport<sup>1,2,3,4</sup>; F. Wang<sup>1,2,3,4</sup>; J. Budde<sup>1,2,3,4</sup>; H. Rhinn<sup>5</sup>; C. Cruchaga<sup>1,2,3,4</sup>; O. Harari<sup>1,2,3,4</sup>; Alector

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) NeuroGenomics and Informatics, Washington University in St. Louis, St. Louis, MO; 2) Department of Psychiatry, Washington University in St. Louis, St. Louis, MO; 3) Hope Center for Neurological Disorders, Washington University in St. Louis, St. Louis, MO; 4) Charles F. and Joanne Knight Alzheimer's Disease Research Center, Washington University in St. Louis, St. Louis, MO; 5) Alector, San Francisco, CA

Alzheimer's disease (AD) is a neurodegenerative disease, affecting around 5.8 million people in the United States as of 2019. It has been widely studied and characterized, but there is still no effective treatment. Using large-scale Genome-wide association studies and whole genome/exome sequencing, researchers have identified common and rare variants associated disease risk that can be used to build prediction models. For instance, GWAS have uncovered 25 AD risk loci, including one mapped to the gene *CD33*. Our previous work has shown the importance of employing protein levels in cerebrospinal fluid (CSF) as endophenotypes for genetic studies. These can help the elucidation of the molecular mechanisms for bridging the genomic variation to the disease.

Here, we have profiled 1305 proteins in CSF samples from AD cases and controls. Several known proteins associated with AD risk are included, such as CD33 and 14-3-3 protein. After quality control (QC) for both proteomic and genotype data, we kept 713 proteins and 510 samples. We aim to identify the genetic architecture that governs protein levels and how these proteins modulate AD risk. To do so, we first identified protein quantitative trait loci (pQTLs) from CSF samples passing QC. We employed linear regression models after adjusting for age, gender, two genetic principle components. We identified 82 proteins with cis-pQTLs study-wide significant ( $P < 7.01 \times 10^{-11}$  after Bonferroni correction), replicating 28 previously reported cis-pQTLs from CSF (e.g. rs2673908 associated with the protein Siglec-9,  $P=6.1 \times 10^{-160}$ ) reported by Sasayama et al 2017. We also discovered 54 CSF proteins with novel cis-pQTLs. Furthermore, we performed genome-wide association analyses and discovered 20 proteins with novel trans-pQTLs ( $P < 7.01 \times 10^{-11}$ ). Next, we studied the causal effect of protein levels in AD, by employing Mendelian randomization approaches and employing study-wide significant pQTL as instrumental variables. For example, increased protein level of CD33 elevated AD risk (Wald-ratio test, Beta=0.73,  $P=4.997 \times 10^{-8}$ ), using cis-pQTL (rs3865444,  $P=5.11 \times 10^{-49}$ ) in the two-sample MR analysis. Our results will be useful for both research into the basic mechanisms of protein functions in pathways causing AD and applicable to discovery of novel biomarkers for AD risk.



## PgmNr 3128: Hypomyelinating leukodystrophy: Do heterozygous variants in *HSPD1* deserve a closer look?

### Authors:

S. Gannavarapu<sup>1</sup>; C. Prasad<sup>2,3,4</sup>; T.B. Balci<sup>2,3,4</sup>; A.N. Prasad<sup>3,4,5</sup>; T. Rupar<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology and Laboratory Medicine, Western University, London, Ontario, Canada; 2) Genetics Program of South Western Ontario, London, Ontario, Canada; 3) Department of Pediatrics, Western University, London, Ontario, Canada; 4) Children's Health Research Institute, London, Ontario, Canada; 5) Division of Pediatric Neurology, Faculty of Medicine, Western University, London, Ontario, Canada

---

Heterozygous variants in *HSPD1* (heat-shock 60kDA protein [HSP60]) are associated with adult-onset spastic paraplegia-13. Whereas, biallelic variants cause pediatric-onset hypomyelinating leukodystrophy, also known as mitochondrial HSP60 chaperonopathy. HSP60 participates in the folding and trafficking of proteins into mitochondria and is expressed abundantly in myelin-forming glial cells.

Exome sequencing identified a de novo heterozygous *HSPD1* variant (p.Leu47Val) in a 23-year-old male with consanguineous parents. At 2 years, he presented with progressive ataxia, developmental delay and dysmyelination showing a symmetric tigroid pattern in the centrum semiovale and periventricular white matter on his MRI. He also had ethylmalonic aciduria, with the presence of homozygous variants in *ACADS*, considered to be a benign trait. In silico analyses, immunoblotting and immunofluorescence (IF) were used to determine the predicted impact of the variant, HSP60 protein expression, and subcellular localization. MitoTracker Red CMXRos was used to study mitochondrial morphology and membrane potential. Multiple prediction score across different in silico tools (e.g. CADD, DynaMut, SIFT, Polyphen2) all categorized the variant as pathogenic.

We propose that certain heterozygous variants in *HSPD1* may lead to more severe features resembling the biallelic phenotype, due to a dominant-negative effect or multigenic interactions. IF imaging of HSP60 demonstrated co-localization with mitochondrion as expected. Preliminary findings suggest decreased HSP60 expression and mitochondrial membrane potential in the proband fibroblasts compared to unaffected, age-matched control cell lines. Alterations in membrane potential are highly suggestive of mitochondrial dysfunction. Even as we actively investigate changes in mitochondrial mechanisms and HSP60 protein complex organization in more disease-representative in vitro models such as patient-derived neural precursor cells, these findings highlight the potential mitochondrial contribution to the progressive loss of neural cells and consequent hypomyelination patterns observed in our patient.

## PgmNr 3129: A rescue based screen to functionally assess *de novo* missense variants lined to autism spectrum disorders using *Drosophila*.

### Authors:

J. Andrews<sup>1,2</sup>; P. Marcogliese<sup>1,2</sup>; S. Deal<sup>1,2,3</sup>; M. Harnish<sup>1,2,3</sup>; H. Graves<sup>1,2</sup>; Y. Chao<sup>1,2</sup>; S. Jangam<sup>1,2</sup>; V. Hemanjani Bhavana<sup>1</sup>; X. Luo<sup>1,2</sup>; N. Liu<sup>1,2</sup>; H. Chung<sup>1,2</sup>; C. Brady<sup>1,2</sup>; H. Pan<sup>2</sup>; H. Chao<sup>1,2,4,5</sup>; P. Lee<sup>1,2</sup>; K. Scott<sup>1,6</sup>; S. Yamamoto<sup>1,2,3,5</sup>; M. Wangler<sup>1,2,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, 77030; 3) Program in Developmental Biology, BCM, Houston, TX, 77030; 4) Department of Pediatrics, BCM, Houston, TX, 77030; 5) Department of Neuroscience, BCM, Houston, TX, 77030; 6) Program in Integrative Molecular and Biomedical Sciences, BCM, Houston, TX, 77030

---

Autism spectrum disorders (ASDs) are complex neurodevelopmental disorders primarily characterized by varying degrees of impaired social communication, repetitive and restricted behavior patterns and atypical responses to sensory stimuli. While the disease mechanism(s) underlying ASD remain unclear, there is strong support for a substantial genetic component. Large-scale sequencing efforts indicate an overlap in candidate genes and pathways between ASD and a number of Mendelian syndromes with ASD as a feature. However, both disease heterogeneity and an expanding list of candidate genes have made assessing their functional significance a challenge. In attempts to functionally annotate potential disease-causing variants we are conducting a functional screen in flies to analyze ASD candidate genes from the Simon's Simplex collection (SSC). The SSC contains whole-exome sequencing data from 2,500 families with a proband with non-syndromic ASD. The majority of probands possess *de novo* missense variants and approximately 80% of these genes are conserved in *Drosophila*. We have prioritized examination of these variants based upon independent genomic data from non-Autism cohorts, potential gene expression in the CNS, and the availability of *Drosophila* reagents. Accordingly, we have selected 126 *Drosophila* Mi[MIC] lines with insertions in between coding exons for conversion to loss-of-function T2A-GAL4 lines. We successfully converted 111 Mi[MIC] lines to T2A-GAL4 insertions. 62/111 of these T2A-GAL4 lines are homozygous lethal, and rescue experiments using reference and variant UAS human cDNA are currently underway. The remaining 49/111 T2A-GAL4 lines are currently being evaluated for alterations in stereotyped patterns of behavior. Together with the parallel over-expression based screen (see poster by Deal et al.), our rescue based screen provides an assay system to assess the function of human variants linked to ASD and other diseases.

## PgmNr 3130: Tissue-specific transcriptional signatures in reciprocal genomic disorders: Insights from mouse brain and human neuronal models.

### Authors:

R. Yadav<sup>1,2,3</sup>; D. JC Tai<sup>1,2,3</sup>; T. Aneichyk<sup>1,2,3</sup>; S. Erdin<sup>1,2,3</sup>; B. B Currall<sup>1,2,3</sup>; P. Razaz<sup>1,2,3</sup>; A. Stortchevoi<sup>1,2,3</sup>; U. Rudolph<sup>5</sup>; J. F Gusella<sup>1,2,3,4</sup>; M. E Talkowski<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genomic Medicine, MGH/Harvard Medical School, Boston, Massachusetts.; 2) Department of Neurology, Harvard Medical School, Boston, Massachusetts, USA.; 3) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.; 4) Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.; 5) Department of Comparative Biosciences, College of Veterinary Medicine, University of Illinois at Urbana-Champaign, Urbana, IL, USA

---

Reciprocal genomic disorders (RGDs) involve copy number variation (CNV), including both deletion (DEL) and duplication (DUP), of large genomic segments mediated by flanking duplicated sequence. Though individually rare, RGDs collectively account for 10-15% of broadly defined neurodevelopmental disorders and congenital anomalies. We previously demonstrated widespread, yet diverse and tissue-specific transcriptional changes associated with 16p11.2 RGD across 6 mouse tissues and CRISPR-engineered human neural stem cells (NSCs) and NGN2-induced neurons (iNs). That CNV alters a 740 kb gene-dense region (29 protein coding genes). Here, we compared these results to the transcriptional signatures associated with a larger 1.5 Mb CNV of 15q13.3 that spans only 7 genes (6 with measurable expression in brain) and is associated with variable phenotypes and lower penetrance than 16p11.2. For 15q13.3 analyses, we generated 119 RNASeq libraries across cerebellum, prefrontal cortex, and striatum from mice harboring CNV of the syntenic 7qc region, as well as from isogenic NSCs and iNs. Expression changed linearly with CNV for all 6 genes in the RGD. The only significant pathway changes in the mouse after multiple testing correction were observed in striatum, where differentially expressed genes (DEGs; FDR<0.1) were downregulated across several neuronal pathways, while the most significant pathways observed in the human neuronal models were associated with Wnt signaling in DEL NSCs. Co-expression analyses of brain tissues and cell lines isolated a single recurrent module that included all 15q13.3 genes and was highly enriched for DEGs across all tissues and neuronal models (maximum > 50% of DEGs in cerebellum), as well as modules enriched for constrained genes and genes associated with NDD and autism. Thus, the alterations due to 15q13.3 perturbation were far more modest than 16p11.2, yet more consistent across tissues. We next sought shared molecular pathways across these RGDs using a weighted z-score based meta-analysis in 279 RNAseq samples across the same three mouse tissues and CRISPR models. Significant commonalities between these RGDs converged on genes enriched for myelin sheath and dendritic function, as well as genes involved in mitochondrial respiratory chain function (FDR<0.1). These results suggest the pathogenic mechanisms associated with recurrent RGDs may converge on common molecular signatures, including changes in neuronal function and cellular respiration.

## PgmNr 3131: Quantifying genetic variability within intrinsically disordered regions of proteins.

**Authors:**

J.E. Miller; S.M. Dudek; A.T. Frase; M.D. Ritchie

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics, Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, Pennsylvania.

---

Characterizing the impact of genetic variants is not only important for defining the etiology of complex traits and diseases, but identifying new drug targets as well. Due to their clear potential for impacting a phenotype, coding regions are often targeted first when searching for disease causing variants or potential pharmacogenetic interactions. However, even after applying methods that characterize single nucleotide variants (SNVs) based on evolutionary constraint, population frequency, gene regulatory properties, and potential pathogenicity, many variants of unknown significance remain. One reason this problem may exist is due to SNV annotation methods over emphasizing the utility of structured protein regions that maintain a stable three-dimensional shape, when there has been growing awareness to the vital functions of intrinsically disordered regions (IDRs). Moreover, pathogenic SNV predictions often underperform in IDRs, suggesting that refining our understanding of these regions could improve our ability to identifying pathogenic variants. IDRs do not conform to a defined three-dimensional shape and are critical for cell signaling and gene regulatory processes. Furthermore, IDRs represent a keystone of an emergent property of the cell to generate membraneless organelles (e.g. liquid-liquid phase separation) for compartmentalizing biochemical processes. While IDRs tend to evolve more rapidly than more structured regions of proteins, there is evidence of purifying selection associated with functional regions within the IDRs. We hypothesize, that the identification of these functional regions with purifying selection can inform the interpretation of the impact genetic variants have on phenotypes. We designed a workflow to identify regions in the genome that code for IDRs and functional regions that reside within them. Next, we used BioBin, a program for rare variant binning, to count low-frequency and rare variants in disordered regions using 1000 genomes data. By comparing the ratio of non-synonymous to synonymous SNVs at a global level in the entire IDR compared to functional regions within the IDR we can capture the substitution rate that occurs across the genome. This approach was applied across individuals of diverse backgrounds as a proof of concept study for later testing in Alzheimer's disease cohorts. This study presents a new strategy for studying the genetic etiology of disease by investigating disordered regions of proteins.

## PgmNr 3132: Is juvenile absence epilepsy with photosensitivity related to de novo LoF variant in *CIC* gene and folate deficiency?

### Authors:

A. Todorova <sup>1,2</sup>; T. Todorov <sup>2</sup>; S. Ategin <sup>1,2</sup>; A. Maver <sup>3</sup>; B. Peterlin <sup>3</sup>; P. Dimova <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Chemistry and Biochemistry, Sofia Medical University, Sofia, Bulgaria; 2) Genetic Medico-Diagnostic Laboratory "Genica", Sofia, Bulgaria; 3) Clinical Institute of Medical Genetics, UMC Ljubljana, Ljubljana, Slovenia; 4) Neurosurgery Clinic, "St. Ivan Rilski" Hospital, Sofia, Bulgaria

---

Low concentration of 5-methyltetrahydrofolate (5-MTHF) in cerebrospinal fluid is associated with cerebral folate deficiency (CFD) syndrome, while folate levels in plasma are within the normal range. Mutations in several folate pathway genes (*FOLR1*, *DHFR* and *PCFT*) have been previously identified in CFD patients. Recently, some CFD cases with milder phenotype have been related to de novo mutation in human capicua (*CIC*) gene. Developmental delay and intellectual disabilities are the common characteristics with or without epilepsy.

Here we report a 14 year old girl, who manifested juvenile absence epilepsy and EEG photosensitivity two years ago. Since then she is on Valproic Acid treatment (1000 mg per day) with excellent response and no absences any more. She has no mental retardation, no behavioral problems, no developmental delay. In order to clarify the genetic cause of the juvenile absence epilepsy we performed array-CGH and whole exome sequencing (WES). The array-CGH showed normal results, while WES discovered a de novo frameshift mutation c.2414delC, p.Pro805GlnfsTer119 in the human Capicua gene (*CIC*). The mutation was confirmed by Sanger sequencing. The biochemical parameters in blood are: folic acid 12.0 to 22.0 nmol/l, ref. value 10.4 - 42.4. At the same time B12 was measured above the upper limit of the range: 684 pmol/l - ref. values 145-569. Moreover, MCH and MCV showed increased values in several blood tests (MCH 34 pg, ref. values 26.0-32.0; MCV 104 fl, ref. values 80.0-97.0), which corresponds to vitamin B12 and/or folic acid deficiency.

Could we consider the above mentioned results in blood as a marker for folic acid deficiency in cerebrospinal fluid and how it refers to the brain? Is it possible the higher B12 concentration to result from folic acid deficiency? Is the *CIC* gene loss-of-function (LoF) variant a disease causing in our case and why the clinical manifestation does not fit to the already published cases?

## PgmNr 3133: Exploiting *Drosophila* to examine RNA exosome-linked disease.

### Authors:

D.J. Morton <sup>1</sup>; B. Jalloh <sup>2</sup>; L. Kim <sup>1</sup>; T. Le <sup>1</sup>; C. Rounds <sup>2</sup>; M.C. Sterrett <sup>1</sup>; B. Brown <sup>1</sup>; S.W. Leung <sup>1</sup>; M.B. Fasken <sup>1</sup>; K.H. Moberg <sup>2</sup>; A.H. Corbett <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biology, Emory University, Atlanta, Georgia.; 2) Cell Biology, Emory University, Atlanta, Georgia.

---

The RNA exosome is an evolutionary-conserved 3'-5' ribonuclease complex critically important for both precise processing and complete degradation of a variety of cellular RNAs. Given, the crucial role of the RNA exosome in post-transcriptional regulation of RNA, it is not surprising that the complex is essential in systems examined thus far. The recent discovery that mutations in genes encoding structural exosome subunits cause tissue-specific diseases makes defining the role of the RNA exosome within specific tissues critically important to understand the basis of these diseases. The *Drosophila* system provides ideal tools to examine tissue-specific gene function. Mutations in the RNA exosome subunit 3 gene (*EXOSC3*) cause Pontocerebellar Hypoplasia Type 1b, an autosomal recessive neurodegenerative disease. The disease-causing mutations identified are missense mutations in evolutionarily-conserved amino acids. The tissue-specific defects these changes cause are challenging to understand based on current models of RNA exosome function with only limited analysis of the complex in any multicellular model *in vivo*. The goal of this study is to provide insight into how mutations in *EXOSC3* impact the function of the RNA exosome leading to disease. *EXOSC3* is an evolutionarily conserved subunit, termed Rrp40 in *Drosophila*. To begin to determine the functional consequences of the specific amino acid substitutions that cause disease in humans, we developed a model of RNA exosome-linked disease in *Drosophila* utilizing CRISPR/Cas9 gene editing to introduce disease-linked mutations in *Rrp40* (*EXOSC3*). Preliminary data examining disease-linked amino acid substitutions in Rrp40 of differing severity reveal behavioral and morphological phenotypes that align with disease severity found in patients. Next, we performed RNA-Seq analysis to define the spectrum of RNA targets affected. Principal component analysis revealed that PCH1b-linked alleles in flies segregated into distinct molecular groups based on their gene expression profiles. Gene expression patterns among the PCH1b mutants highlight large-scale differences in lncRNAs and mRNAs in each PCH1b-mutant providing a molecular basis for the phenotypes observed. These data provide evidence that the *Drosophila* model can be used to provide insight into tissue-specific function of the RNA exosome *in vivo* and explore the functional consequences of amino acid substitutions linked to disease phenotypes.



## PgmNr 3134: Deficiencies in vesicle-mediated transport mediated by *TRAPPC4* are associated with severe syndromic intellectual disability.

### Authors:

N.J. Van Bergen <sup>1,2</sup>; Y. Guo <sup>3</sup>; N. Al-Deri <sup>4</sup>; Z. Lipatova <sup>5</sup>; D. Stanga <sup>4</sup>; R. Murtazina <sup>5</sup>; V. Gyurkovska <sup>5</sup>; S. Zhao <sup>5</sup>; D. Pehlivan <sup>6,7</sup>; T. Mitani <sup>6</sup>; A. Gezdirici <sup>8</sup>; J. Antony <sup>9</sup>; F. Collins <sup>10,11</sup>; M.J.H. Willis <sup>12</sup>; Z.H. Coban Akdemir <sup>6</sup>; P. Liu <sup>6</sup>; J. Punetha <sup>6</sup>; J.V. Hunter <sup>13</sup>; J.N. Jhangiani <sup>14</sup>; J.M. Fatih <sup>6</sup>; J.A. Rosenfeld <sup>6</sup>; J.E. Posey <sup>6</sup>; R.A. Gibbs <sup>6,14</sup>; K. Wilson <sup>10</sup>; C. Troedson <sup>9</sup>; J.R. Lupski <sup>6,14,15,16</sup>; M. Sacher <sup>4,17</sup>; N. Segev <sup>5</sup>; H. Hakonarson <sup>3</sup>; J. Christodoulou <sup>1,2,18,19</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Brain and Mitochondrial Research Group, Murdoch Children's Research Institute, Royal Children's Hospital, Melbourne, Australia; 2) Department of Paediatrics, University of Melbourne, Melbourne, Australia; 3) Center for Applied Genomics (CAG) at the Children's Hospital of Philadelphia (CHOP), Philadelphia, USA; 4) Department of Biology, Concordia University, Montreal, Quebec, Canada; 5) Department of Biochemistry and Molecular Genetics, University of Illinois at Chicago, Chicago, IL, USA; 6) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, 77030, USA; 7) Section of Pediatric Neurology and Developmental Neuroscience, Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA; 8) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, 34303, Turkey; 9) TY Nelson Dept of Neurology and Neurosurgery, Children's Hospital at Westmead, Sydney, Australia; 10) Western Sydney Genetics Program, Children's Hospital at Westmead, Sydney, Australia; 11) Medical Genomics Department, Royal Prince Alfred Hospital, Sydney; 12) Department of Pediatrics, Naval Medical Center San Diego, San Diego, California, 92134, USA; 13) Department of Radiology, Baylor College of Medicine, Houston, Texas, 77030, USA; 14) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, 77030, USA; 15) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, 77030, USA; 16) Texas Children's Hospital, Houston, Texas, 77030, USA; 17) Department of Anatomy and Cell Biology, McGill University, Montreal, Quebec, Canada; 18) Victorian Clinical Genetics Services, Royal Children's Hospital, VIC, Australia; 19) Kids Research, The Children's Hospital at Westmead, Sydney, NSW, Australia

---

The transport protein particle (TRAPP) complexes regulate key trafficking events and are required for autophagy. TRAPPC4 is a core component of the TRAPP complexes and one of the essential subunits in GEF activity. Pathogenic variants in specific TRAPP subunits are associated with neurological disorders. In an effort to determine the genetic etiology, we undertook exome sequencing (ES) in three unrelated families of different ethnicities with seven affected children with features of early onset seizures, developmental delay, microcephaly, sensorineural deafness, spastic quadriplegia and progressive cortical and cerebellar atrophy. In all seven affected individuals we identified a rare homozygous variant in a non-canonical well-conserved splice site within *TRAPPC4* (hg19:chr11:g.118890966A>G; *TRAPPC4*: NM\_016146; c.454+3A>G). Haplotype analysis is in progress. *In silico* analysis predicted the variant to cause aberrant splicing. Consistent with this, experimentally we showed both a reduction in full-length transcript and a shorter transcript missing exon 3, suggestive of an incompletely penetrant splicing defect. TRAPPC4 protein levels were

significantly reduced whilst levels of other TRAPP complex subunits were unaffected. Native PAGE and size exclusion chromatography demonstrated a defect in TRAPP complex assembly and/or stability. Intracellular trafficking through the Golgi using the marker protein VSVG-GFP-ts045 demonstrated significantly delayed entry and exit from the Golgi in patient fibroblasts. Lentiviral expression of wildtype *TRAPPC4* in patient fibroblasts restored trafficking, suggesting that the trafficking defect was due to reduced *TRAPPC4* levels. Consistent with the recent association of the TRAPP complex in autophagy, we found that patient fibroblasts had a basal autophagy defect and a delay in autophagic flux, likely due to unsealed autophagosomes. These results were validated using a yeast *trs23* temperature sensitive mutation that exhibits constitutive and stress-induced autophagic defects at permissive temperature and a secretory defect at restrictive temperature. We provide evidence for a disease-association mutation in a member of the core TRAPP subunit, *TRAPPC4* that is associated with an autophagy defect. Our findings add to the growing number of TRAPP-associated neurological disorders.



## PgmNr 3135: A holoprosencephaly-associated transmembrane domain mutation of human SMOOTHENED (SMO) leads to hypomorphic activity for forebrain and craniofacial development in a zebrafish model.

### Authors:

M. Tanim-Nagai; S. Hong; E. Roessler; P. Hu; B. Carrington; R. Sood; M. Muenke

[View Session](#) [Add to Schedule](#)

**Affiliation:** NHGRI, National Institutes of Health, Bethesda, MD

---

**Background and purpose:** Holoprosencephaly (HPE) is the most common congenital anomaly of the forebrain and face in humans. Sonic hedgehog (SHH) is one of the best characterized HPE genes, and its signaling pathway plays crucial roles in numerous developmental processes including craniofacial development and neurogenesis. However, the disease-related roles of *SMO* genetic variation are largely unknown. We now report functional testing of seven non-synonymous variants of *SMO* derived from HPE molecular and clinical data. Here we characterize human *SMO* variants function by a novel zebrafish rescue assay and systems biology analysis.

**Materials and methods:** For our rescue assay, *SMO* synthetic mRNA (WT\_*SMO*) or each variant's site-directed metagenesis *SMO* mRNA (SDM\_*SMO*) was injected separately into fertilized one cell stage of CRISPR/Cas9 *smo* K/O zebrafish embryos. The phenotype and genotype were analyzed after injection. In our systems biological assay, co-injection of wild-type zebrafish with small doses of SHH was effective in modulating the ability of endogenous zebrafish *Ptch* to inhibit injected human WT\_*SMO*. Therefore, this approach was adopted for each variants SDM\_*SMO*.

**Results:** We determined that WT\_*SMO* and many of its variants could rescue the abnormal phenotype of *smo* K/O fish. Interestingly, the overall rescue pattern was different for each variant. There were three types of phenotypic rescue: 1) complete, 2) mainly eye, or 3) mainly body. Three variants partially rescued ( $P < 0.01$ ), and only one variant, with a mutation point in the transmembrane domain, was a strong hypomorph ( $P < 0.01$ ). Co-injection of SHH with each SDM\_*SMO* independently replicated these findings based on our analysis of the over-expression phenotypic severity and modulation of eye development-related gene expression by *in situ* hybridization; however, these effects were only modestly different between each variant compared to WT\_*SMO*.

**Conclusion:** Our results reveal that variants in human *SMO* may be modifiers of brain and craniofacial development and implicate in the transmembrane domain. We conclude that the disease genes should not always be studied in isolation; rather the molecular pathway interactions should be incorporated into biological assay development and interpretation. In summary, we have characterized the role of the *SMO* locus in forebrain and craniofacial development and infer that the trans-membrane domain is crucial for normal function.

## PgmNr 3136: Humanized *STXBP1* animal models as a precision medicine approach to detecting pathogenicity in clinical variants.

### Authors:

C. Hopkins <sup>1</sup>; T. Brock <sup>1</sup>; K. McCormick <sup>1</sup>; R. Fekete <sup>1</sup>; A. Johnson <sup>2</sup>; C. Rigby <sup>2</sup>; M. Bainbridge <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Nemametirx, Eugene, Oregon.; 2) STXBP1 Foundation, Philadelphia, Pennsylvania.; 3) Rady Children's Hospital, San Diego, California

---

Functional data on clinical variants addresses a growing problem in genomic data. Approx. 75% the variant entries in public databases are either unknown or VUS for their pathogenic consequence. Functional data is a significant contributor to pathogenicity assessment (PS3). Yet, even in cases where pathogenicity is known, clinicians and patients are often unsure of which therapeutics offer the best treatment for the specific variant. In response to these needs, our group has created a series of humanized animal models as Clinical Avatars for functional assessment and therapeutic best-fit discovery. CRISPR-based methods are used to insert human coding sequence as a gene replacement in *C. elegans*. When the human gene restores normal function in *C. elegans*, the system is ready for variant installation. In a demonstration with an epilepsy-associated gene, the human *STXBP1* coding sequence was inserted as gene replacement of *unc-18*. Next, a series of clinically-observed variants were installed in the hSTXBP1 humanized loci. Finally, change-of-function activity was quantified using a series of functional assays. CRISPR-based knock-out of *unc-18* results in loss of function characterized by uncoordinated movement. CRISPR-based knock-in of hSTXBP1 led to a rescue of function. Rescue of activity was observed to be complete (100%) by electrophysiology (up 4x from *unc-18* KO), 100% in liquid 'thrashing' assays for movement (up 3x), and 50% in a chemotaxis behavioral assay (up from 0%). Variants installed into hSTXBP1 locus showed a range of functional defects. For instance, selected variants of p.R292H (rs796053361 P), p.R406H (rs886041246 P) and p.R388X (rs121918321 P) are established pathogenic variants in *STXBP1* known to be autosomal dominant. Heterozygotes were examined in a modified chemotaxis assay for deficiencies in motivated behavior. The p.R388X variant creates a two-thirds truncation of STXBP1 protein and a dominant negative activity was observed in the heterozygote - approximately 3 fold more severe than a heterozygote with full gene KO. Variant activity profiling is revealing that gene function defects can be quantitatively measured in the nematode animal model. Larger sets of *STXBP1* variants (including VUS alleles) are in preparation. Drug testing is being implemented for pathogenic variants to find out which antiepileptic drugs (AED) are the therapeutic best fit for the variant-specific condition of *STXBP1* patients.

## PgmNr 3137: Clinical exome sequencing reveals KLHL40 mutation associated with cleft palate, severe nemaline myopathy, and fetal akinesia.

### Authors:

K. Avasthi <sup>1</sup>; S. Agarwal <sup>1</sup>; I. Panigrahi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, Uttar Pradesh, India; 2) Department of Pediatrics, Postgraduate Institute of Medical and Education Research, Chandigarh-160012

---

### Introduction

Congenital myopathies are heterogeneous group of neuromuscular disorders and characterised by hypotonia early onset muscular weakness, developmental delay. Congenital myopathies are classified into centranuclear myopathies, nemaline myopathies, core myopathies and congenital fiber-type disproportion based on major pathological features found in muscle biopsies.

### Material and Methods

The present study was conducted on 25 syndromic orofacial cleft patients present with different clinical features. After clinical evaluation, genomic DNA was extracted by standard protocol (Qiagen kit). Clinical exome sequencing (illumina platform) was performed in patient DNA sample and sequencing data was analysed for sequence variant and identified variant validation was done by Sanger sequencing method.

### Results

A homozygous nonsense variant in exon 1 of the KLHL40 gene (chr3:42727712G>A: Depth 53x) that results in a stop codon and premature truncation of the protein at codon 201 (p.Trp201Ter; ENST00000329777) was detected.

### Conclusion

This study suggest the involvement of KLHL40 mutation with fetal akinesia and severe nemaline myopathies, also associated with cleft palate. At present scenario there is no evidence which recognize that cleft palate associated with severe nemaline myopathies and fetal akinesia.

## PgmNr 3138: Genetic recapitulation and functional annotation of frequently mutated genes in myopathy.

### Authors:

J. Hwang <sup>1</sup>; H. Kim <sup>2</sup>; S. Park <sup>1</sup>; Y. Kim <sup>1</sup>; M. Choi <sup>3,4</sup>; S. Kim <sup>4</sup>; J. Chae <sup>4</sup>; Y. Ahn <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Rare Diseases, Center for Biomedical Science, KNIH, KCDC, Cheongju, Korea; 2) Department of Internal Medicine, College of Medicine, Institute of Kidney Disease Research, Yonsei University, Seoul, Korea; 3) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea; 4) Department of Pediatrics, Pediatric Clinical Neuroscience Center, Seoul National University Childrens Hospital, Seoul, Korea

---

In clinical exome-based variant identification for molecular diagnosis, researchers have encountered frequently mutated genes (FLAGS) with genetic variants of uncertain significance or as exome negatives. It has recurrently emerged as one of the contentious issues. However, regulatory functional implications of the FLAGS have still not been fully considered in rare genetic diseases. To evaluate the relevance of the FLAGS in myopathy, we conducted genetic recapitulation and comprehensive functional annotation. In seventeen trio-based whole exome analyses, we found that twenty putative genes intersected with the FLAGS. Using an integrative genetic-epigenetic approach, we demonstrated that epivariations were associated with regulatory changes underlying gene expression, DNA methylation, histone modification and open chromatin accessibility. Taken together, the twenty FLAGS genes may serve as a potential target in myopathy. Our findings might provide additional insight into understanding variable expressivity in rare diseases.

## PgmNr 3139: eQTL analysis in brain cortex samples of 3,833 individuals and 31,684 blood samples reveals distinct regulatory effects of disease-associated genetic variants.

### Authors:

N. de Klein <sup>1</sup>; E. Tsai <sup>2</sup>; O. El Garwany <sup>1</sup>; H. Runz <sup>2</sup>; L. Franke <sup>1</sup>; H. Westra <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, Groningen, The Netherlands; 2) Genetic Epidemiology Group, Translational Biology, Biogen, 125 Broadway, Cambridge, MA, USA

Many studies have shown how genetic variation affects gene expression (expression quantitative trait loci; eQTL), with the aim to identify the molecular downstream consequences of genetic risk factors for disease. However, since eQTL effect sizes are usually small, large sample-sizes are required to uncover both *cis*- and *trans*-eQTLs. We therefore performed a very large-scale eQTL meta-analysis in brain, encompassing 8,654 RNA-seq samples obtained from AMP-AD, Braineac, TargetALS, CMC, GTEx, PsychEncode and public ENA resources, using the same eQTL analysis procedure as eQTLgen. To study both *cis*- and *trans*-eQTL, we rigorously harmonized the brain RNA-seq and genotype data permitting us to study 3,833 unique individuals of European descent. Out of 19,844 tested protein-coding genes, 11,942 were *cis*-eQTL genes (*cis*-eGenes) and 713 were *trans*-eQTLs genes (FDR<0.05).

We contrasted these brain eQTLs to eQTLgen, a large-scale eQTL analysis in blood samples from 31,684 individuals. We observed that *cis*-eQTLs are often unique to either brain or blood: of the 16,989 *cis*-eGenes in blood and 11,942 *cis*-eGenes in brain, 9,017 showed a significant effect in both tissues, yet 17% showed an opposite allelic direction for the SNP with the largest eQTL effect in eQTLgen (lead eSNP). While the majority of these opposite effects are potentially explained by imperfect linkage, the eQTLgen lead eSNP was also the lead eSNP in brain for 26 of these genes, suggesting tissue dependent regulation. Finally, we observed that highly expressed genes in brain are proportionally less likely to be regulated in *cis* than medium expressed genes in blood.

We subsequently compared the overlap of *cis* and *trans*-eQTL that we found in blood and brain with GWAS associated SNPs. We observed that *trans*-eQTL analysis in brain is informative for SNPs, associated to neurological disorders: we observed a significant enrichment of *trans*-eQTL signals in brain ( $P=7.6 \times 10^{-3}$ ), whereas we did not find any *trans*-eQTL enrichment in blood ( $P = 1$ ).

Our eQTL results provide a context-specific resource that can assist in fine-mapping and SNP prioritization for neurological disorders. We show the importance of large scale eQTL meta-analysis in brain by identifying genetic regulators that either cannot be found in blood or have a different direction of effect.

## PgmNr 3140: RNA binding of FMRP that is related to Fragile X syndrome may play an important role in neurite formation.

### Authors:

H. Shimizu; E. Suzuki; H. Hohjoh

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Molecular Pharmacology, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan

---

Fragile X syndrome (FXS) is an inherited disorder caused by mutant Fragile X mental retardation gene 1 (*FMR1*) carrying aberrantly expanded CGG triplet repeats: aberrant CGG repeats inhibit the expression of *FMR1*. *FMR1* encodes fragile X mental retardation protein (FMRP) that is a RNA binding protein, and involved in the stabilization and translation of its bound mRNAs. FMRP is present in dendrites and dendritic spines of neurons and plays an important role in the local translation of mRNAs, which contributes to neuronal development and synaptic function. Some FXS patients show similar characteristics of autism spectrum disorder (ASD), a developmental disorder: they exhibit changes in neurite outgrowth and dendritic spine formation. Immature dendritic spines were detected in FXS model mice, and the decrease in dendritic spines and the reduction of basal dendrite length and branching were observed in visual cortex pyramidal neurons of *Fmr1* knockout mice. Thus, FXS appears to be somehow associated with neurite formation, but is not yet well understood. In this study, to see the effects of *Fmr1* (FMRP) on process formation, we generated two lines of mouse neuroblastoma Neuro2a (N2a) cells in which *Fmr1* were knocked out by gene editing using CRISPR-Cas9. The two N2a cell lines showed little or no expression of FMRP by western blotting; and interestingly, they exhibited severe impairment of process formation under differentiation conditions. When the expression plasmid encoding full-length FMRP was introduced into *Fmr1*-deficient N2a cells, the cells restored their ability of process formation. In contrast, when mutant FMRP lacking RNA binding was introduced to *Fmr1*-deficient N2a cells, the resulting cells showed little improvement in process formation. Therefore, the findings suggest that FMRP, which is encoded by *Fmr1* related to FXS, is involved in process (neurite) formation, and that RNA binding of FMRP may play an important role in process formation. We are trying to isolate candidate RNAs associated with FMRP and involved in process formation.

## PgmNr 3141: Characterizing the hnRNP gene family in neurodevelopmental disorder subtypes.

### Authors:

M.A. Gillentine<sup>1</sup>; T. Wang<sup>1</sup>; K. Hoekzema<sup>1</sup>; H. Guo<sup>1</sup>; B. Coe<sup>1</sup>; J. Rosenfeld<sup>2</sup>; H. Panjwani<sup>3</sup>; E. Torti<sup>4</sup>; A. Sulovari<sup>1</sup>; J. Gecz<sup>5,6</sup>; M. Nordenskjöld<sup>7</sup>; C. Romano<sup>8</sup>; F. Kooy<sup>9</sup>; K. Xia<sup>10</sup>; H. Peters<sup>11</sup>; C. Lajonchere<sup>12</sup>; G. Mirzaa<sup>13,14</sup>; P. Atwal<sup>15</sup>; A. Lehman<sup>16</sup>; C. Zweier<sup>17</sup>; L.M. Bird<sup>18,19</sup>; B. De Vries<sup>20</sup>; R. Bernier<sup>21,22</sup>; E.E. Eichler<sup>1,23</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genome Sciences, University of Washington, Seattle, Washington, USA; 2) Baylor Genetics Laboratories, Houston, Texas, USA; 3) Psychiatry and Behavior, University of Washington, Seattle, Washington, USA; 4) GeneDX, Gaithersburg, Maryland, USA; 5) Adelaide Medical School and the Robinson Research Institute, University of Adelaide, Adelaide, Australia; 6) South Australian Health and Medical Research Institute, Adelaide, SA 5000, Australia; 7) Department of Molecular Medicine and Surgery, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 8) Unit of Pediatrics & Medical Genetics, IRCCS Associazione Oasi Maria Santissima, Troina, Italy; 9) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 10) State Key Laboratory of Medical Genetics, School of Life Sciences, Central South University, Changsha, China; 11) Centre for Human Genetics, KU Leuven and Leuven Autism Research (LAuRes), Leuven, Belgium; 12) Autism Genetic Resource Exchange, Autism Speaks, Los Angeles, California, USA; 13) Seattle Children's Hospital, Seattle, Washington, USA; 14) Northwest Clinical Genomics Laboratory, University of Washington, Seattle, Washington, USA; 15) The Atwal Clinic: Genomic & Personalized Medicine, Jacksonville, Florida, USA; 16) Dept. of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 17) Institute of Human Genetics, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany; 18) Clinical Genetics/Dysmorphology, Clinical Pediatrics, University of California, San Diego, California, USA; 19) Rady Children's Hospital San Diego, San Diego, California, USA; 20) Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands; 21) Seattle Children's Autism Center, Seattle, Washington, USA; 22) Center on Human Development and Disability, University of Washington, Seattle, Washington, USA; 23) Howard Hughes Medical Institute, University of Washington, Seattle, Washington, USA

---

With the large number of genes implicated in neurodevelopmental disorders (NDDs), it is unlikely that each has a unique molecular pathogenesis. Variation in gene families likely results in shared clinical spectrums due to shared molecular pathology that may allow for targeted therapies. To identify gene families that fit this model, we assessed 10927 published exomes of patients with NDDs, highlighting the heterogeneous ribonucleoproteins (hnRNPs), a large family involved in RNA processing, as a family with significantly more *de novo* likely gene-disruptive (LGD) variants than expected. With further investigation of our targeted sequencing data from an additional 20990 probands, combined with clinical sequencing data generated from Baylor Genetics Laboratories and GeneDx, collaborations through GeneMatcher and Geno2MP, as well as the literature (in total, 64393 individuals), we have identified 140 patients carrying hnRNP variations. Nine hnRNPs (hnRNPU, K, H2, R, Q, UL1, D, F, A2B1) were identified in our initial dataset, three of which met genome-wide significance ( $p < 4.2 \times 10^{-7}$ ) for LGD variants using the union of two statistical models, four having

nominal significance for LGD variants ( $p < 0.05$ ), and one reaching significance for missense variants. Our NDD cohort has significantly higher burden of hnRNP variants compared to gnomAD, supporting their pathogenicity. Clinically, we have found a shared spectrum of intellectual disability/developmental delay, autism spectrum disorder, and MRI abnormalities, with hnRNPs of higher similarity having more shared phenotypes.

Functionally, human fetal single-cell transcriptomic data shows high levels of expression of these NDD-related hnRNPs among excitatory neurons, radial glia, ventral progenitor cells, and intermediate progenitor cells, with decreased expression among inhibitory and newborn neurons. This shared expression is notable, as hnRNPs are implicated in many processes, including neuronal development, and often work cooperatively. Using Cas9 gene-edited induced pluripotent stem cells and differentiated neuronal and glial cells, we are investigating functional consequences of patient informed variants, identifying changes in cell viability/proliferation, neuronal morphology and activity, as well as hnRNP-specific functionality, including effects on splicing and binding targets (via CLIP-seq) and expression (via RNA-seq and ribosome profiling), as well as RNA/protein trafficking.



## PgmNr 3142: Neonatal lethality and genetic modifiers in a new mouse model of Niemann-Pick disease type C.

### Authors:

J. Rodriguez-Gil <sup>1,2,3</sup>; D. Watkins-Chow <sup>1</sup>; F. Platt <sup>2</sup>; W. Pavan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Human Genome Research Institute, NIH, Bethesda, MD; 2) Department of Pharmacology, University of Oxford, United Kingdom; 3) University of Wisconsin-Madison School of Medicine and Public Health, Medical Scientist Training Program, Madison, WI

Niemann-Pick Disease, Type C (NPC) is a rare, fatal neurodegenerative disorder characterized by lysosomal accumulation of unesterified cholesterol and sphingolipids. These subcellular pathologies eventually cause phenotypes of hepatosplenomegaly and neurological degeneration leading to premature death. NPC is extremely heterogeneous in the timing of clinical presentation (prenatal to adulthood) and is associated with a wide spectrum of causative *NPC1* mutations. Miglustat (Zavesca) is the only approved therapy for NPC, and while it has been shown to improve clinical outcomes it is not a cure. To study the genetic architecture of NPC, including the clinical and genetic heterogeneity seen in this patient population, we have generated a new mouse model (*Npc1*<sup>em1Pav</sup>). Using this model, we identified strain-specific quantitative trait loci (QTL) affecting lifespan and characterized a novel lung phenotype causing neonatal lethality. Interestingly, *Npc1*<sup>em1Pav</sup> mutants on a C57BL/6J (B6J) genetic background had a more severe visceral pathology of foam cell accumulation than *Npc1*<sup>em1Pav</sup> mutants on a BALB/cJ background. These underlying pathological changes translated into B6J mutants having a significantly shorter lifespan ( $70 \pm 4.30$  days) than mutants on a BALB/cJ background ( $84 \pm 7.25$ ;  $p < 0.0001$ ), suggesting strain-specific modifiers contribute to disease severity. QTL analysis of backcross N2 mutants detected significant linkage to markers on chromosomes 1 (LOD=5.57) and 7 (LOD=8.91). Analysis of predicted deleterious sequence variants between B6J and BALB/cJ strains at these two loci resulted in a short list of candidate genes that can now be evaluated as potential modifiers of NPC severity and progression.

A significant reduction of viable *Npc1*<sup>em1Pav</sup> mice was also observed on the B6J background at Postnatal day 10 (P10) (10.5% vs. 25%,  $p < 0.0001$ ). Interestingly, the expected genotype frequency was observed at E18.5-E19.5 (26.5% vs. 25%,  $p = 0.034$ ) suggesting *Npc1*<sup>em1Pav</sup> mutants exhibit a previously uncharacterized lethal phenotype postnatally. C-sections accompanied by maternal fostering confirmed that the majority of *Npc1*<sup>em1Pav</sup> mutants died shortly after birth showing signs of respiratory insufficiency and abnormal surfactant pathology. A similar phenomenon reported in NPC patients with pre/perinatal onset indicates *Npc1*<sup>em1Pav</sup> may be the first valid animal model of this severely affected group.

## PgmNr 3143: Kinesinopathies: Emerging role in genetically undiagnosed Rett syndrome.

### Authors:

S. Kaur<sup>1,2</sup>; N.J. Van Bergen<sup>1,2</sup>; W.A. Gold<sup>3,4</sup>; K. Verhey<sup>5</sup>; L. Boyle<sup>6</sup>; C.J. Nowell<sup>7</sup>; S. Howden<sup>8</sup>; M. Dottori<sup>9</sup>; T. Roscioli<sup>10</sup>; S. Eggers<sup>11,12</sup>; S. Lunke<sup>11,12</sup>; Z. Stark<sup>2,12</sup>; C. Ellaway<sup>4,13</sup>; M.J. Friez<sup>14</sup>; B. Ben Zeev<sup>15</sup>; A. Torella<sup>16</sup>; V. Nigro<sup>16</sup>; J. Pinner<sup>17</sup>; R. Sachdev<sup>17</sup>; D. Milnes<sup>18</sup>; B. Schonewolf-Greulich<sup>19</sup>; Z. Tümer<sup>19</sup>; M.C. Kruer<sup>20</sup>; S. Bakhtiari<sup>20</sup>; B. Isidor<sup>21</sup>; F. Bilan<sup>22</sup>; N. Brown<sup>12</sup>; A. Hidalgo-Gonzalez<sup>2,23</sup>; W.K. Chung<sup>6</sup>; J. Christodoulou<sup>1,2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Brain and Mitochondrial Research Group, Murdoch Children's Research Institute, Parkville, VIC, Australia; 2) Department of Paediatrics, University of Melbourne, Parkville, VIC, Australia; 3) Molecular Neurobiology Lab, Kids Research, Westmead Children's Hospital, Westmead, NSW, Australia; 4) Disciplines of Genetic Medicine and Child and Adolescent Health, Sydney Medical School, University of Sydney, NSW, Australia; 5) Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI, USA; 6) Division of Molecular Genetics, Columbia University Medical Center, NY, USA; 7) Drug Discover Biology, Monash Institute of Pharmaceutical Sciences. Monash University, VIC, Australia; 8) Kidney Development, Disease and Regeneration, Murdoch Children's Research Institute, Parkville, VIC, Australia; 9) Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong, NSW, Australia; 10) Garvan Institute, Sydney, NSW, Australia; 11) Translational Genomics Unit, Murdoch Children's Research Institute, Parkville, VIC, Australia; 12) Victorian Clinical Genetics Services, Murdoch Children's Research Institute, Parkville, VIC, Australia; 13) Western Sydney Genetics Program, Children's Hospital at Westmead, Westmead, NSW, Australia; 14) Greenwood Genetic Center, Greenwood, SC, USA; 15) Sheba Medical Center, Tel HaShomer, Ramat Gan, Israel; 16) Department of Precision Medicine, University of Campania "Luigi Vanvitelli", Naples, Italy; 17) Centre for Clinical Genetics, Sydney Children's Hospital, Randwick, NSW, Australia; 18) Genetic Health Queensland, Royal Brisbane & Women's Hospital, Herston, QLD, Australia; 19) Kennedy Center, Department of Clinical Genetics, Copenhagen University Hospital, Rigshospitalet, Glostrup, Denmark; 20) College of Medicine, University of Arizona, Phoenix, AZ, USA; 21) L'institut du Thorax, INSERM, CNRS, UNIV Nantes, CHU de Nantes, Nantes, France; 22) Neurovascular unit and cognitive disorders, University of Poitiers, Poitiers, France; 23) Stem Cell Disease Modelling Facility, Murdoch Children's Research Institute, Parkville, VIC, Australia

---

Rett Syndrome (RTT) is a severe neurodevelopmental disorder that is mainly caused by mutations in the methyl-CpG-binding (*MECP2*) gene; however, between 5 – 40% of classical and atypical RTT patients do not have a genetic diagnosis. Classical RTT patients exhibit normal development until 6 -18 months of age after which developmental regression occurs. Moreover, patients develop characteristic abnormal hand movements, a hallmark of classical RTT. Atypical RTT patients have many features of classical RTT but do not meet all of the specific diagnostic criteria. There is an unmet need to provide a precise genetic diagnosis for these patients. To this end we have performed singleton genomic analysis on 40 of a cohort of 80 genetically undiagnosed RTT/RTT-like patients. To date, we have obtained a likely genetic diagnosis for 12 cases and identified 5 genes, including 2 novel candidate genes, potentially associated with RTT. We have identified a *de novo* heterozygous missense variant (c.744C>A; p.Asp248Glu) in the motor domain of kinesin-3 family member 1A

(*KIF1A*), which encodes a neuron-specific kinesin motor protein essential for ATP-dependent anterograde axonal transport of synaptic cargos. Through international collaborations, we have now identified 18 different heterozygous *KIF1A* variants in 25 patients with many clinical features seen in RTT including ataxia, spasticity, hypertonia, seizures and gait disturbances. All of these variants are within the highly conserved motor domain, and 11 variants are novel. *In-silico* tools predict all these variants to be pathogenic, and 3D molecular modelling predicts disrupted ATP hydrolysis and/or cycling rate. For 4 of the novel variants, we determined that the ability of mCitrine-tagged KIF1A-motor domain proteins to accumulate in neurite tips of differentiated SH-SY5Y cells is reduced by > 90-95% ( $p < 0.0001$ ) compared to the wild type protein. This suggests the variants prevent the movement of KIF1A along microtubules. Complementary studies using live-cell single-molecule KIF1A tracking, *in vitro* microtubule gliding assays and kinetic ATPase assays to determine variants' effect on KIF1A motor activity are under way. We are generating iPSC from patient's fibroblasts along with CRISPR/Cas-9 gene corrected controls to examine KIF1A-specific cargo trafficking defects in neurons. This work has identified a potentially novel disease-causing gene in RTT/ RTT-like patients which may be tractable to future targeted therapies.

## **PgmNr 3144: Constitutive activation of mTORC1 signaling induced by biallelic loss-of-function mutations in *SZT2* underlies a discernible neurodevelopmental disease.**

### **Authors:**

Y. Nakamura <sup>1</sup>; K. Kato <sup>1,2</sup>; N. Tsuchida <sup>3</sup>; Y. Takahashi <sup>2</sup>; S. Saitoh <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School, Nagoya, Japan; 2) Department of Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan; 3) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan

---

There have been increasing number of reports of *SZT2*-related neurological diseases, the main symptoms of which are epilepsy, developmental delay, macrocephaly and a dysmorphic corpus callosum. *SZT2* functions as a regulator of mechanistic target of rapamycin complex 1 (mTORC1) signaling in cultured human cell lines and mouse tissues. However, it remains to be determined whether mutations in *SZT2* in human patients alter mTORC1 signaling. In this study, we aimed to investigate the functional consequence of biallelic *SZT2* variants in Epstein-Barr virus-induced lymphoblastoid cell lines (LCLs) established from two patients with a typical *SZT2*-related neurodevelopmental disease. Increased phosphorylation of S6 kinase and S6 was identified in patient-derived cell lines under amino acid-starved condition, suggestive of constitutive hyperactivation of mTORC1 signaling. This result was validated by constitutive lysosomal localization of mTOR in patients' LCLs. Furthermore, patients' LCLs display an excessive response to slight amino acid stimulation. Our data suggest the loss-of-function nature of *SZT2* mutations in the patients, and consequent hyperactivation of mTORC1 signaling in response to both amino acid starvation and stimulation in their LCLs. By these functional analyses, the pathogenicity of newly identified *SZT2* variants can be determined, allowing for more detailed characterization of genotype-phenotype correlations.

## PgmNr 3145: Somatic L1 retrotransposition in cortical neurons and non-brain tissues of Rett patients and healthy individuals.

### Authors:

B. Zhao<sup>1,2,3</sup>; Q. Wu<sup>4,5</sup>; A.Y. Ye<sup>5,6,7</sup>; J. Guo<sup>2,8</sup>; X. Zheng<sup>2,3</sup>; X. Yang<sup>6</sup>; L. Yan<sup>6</sup>; Q. Liu<sup>9</sup>; T.M. Hyde<sup>10,11</sup>; L. Wei<sup>2,3,6</sup>; A.Y. Huang<sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA; 2) National Institute of Biological Sciences, Beijing, 102206, China; 3) Graduate School of Peking Union Medical College, Beijing, 100730, China; 4) School of Life Sciences, Peking University, Beijing, 100871, China; 5) Peking-Tsinghua Center for Life Sciences, Beijing, 100871, China; 6) Center for Bioinformatics, State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing, 100871, China; 7) Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China; 8) College of Life Sciences, Beijing Normal University, Beijing, 100875, China; 9) Laboratory of Clinical Investigation, National Institute on Aging, Baltimore, MD 21224, USA; 10) Lieber Institute for Brain Development, Baltimore, MD 21205, USA; 11) Departments of Psychiatry & Behavioral Sciences and Neurology, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

---

Somatic mosaicism results from genetic variations that occur in the somatic cells during the development and aging process of the human body. Retrotransposon is a major source of genetic variation, jumping in the human genome in a copy-and-paste manner via TPRT-mediated retrotransposition. Mounting evidence supports that human-specific LINE-1 (L1s) can retrotranspose postzygotically in human tissues and contribute to intra-individual genetic diversity. However, the genomic distribution, rate, accumulation, and selective pressure of somatic L1s insertions and their potential impact on carrier cells remain largely unclear.

Here, we developed HAT-seq (human active transposon sequencing), the first PCR-based bulk sequencing method capable of identifying rare somatic insertions in a small fraction of cells—even unique cells—from non-tumor human tissues. We profiled 9,181 somatic L1s insertions from 20 postmortem tissues obtained from five Rett patients and five matched healthy controls. Low-level clonal somatic insertions from both cortical neurons and non-brain tissues were thoroughly characterized using the 3' and 5' L1-genome junctions resolved by nested PCR and Sanger sequencing. To our knowledge, for the first time we reported somatic insertions in non-brain tissues of healthy individuals. In Rett patients, somatic insertions were significantly depleted in exons—mainly contributed by long genes—than healthy controls, implying that cells carrying *MECP2* mutations might be defenseless against a second exonic L1s insertion. We observed a significant increase of somatic L1s insertions in the brain compared with non-brain tissues from the same individual. Compared to germline insertions, somatic insertions were less sense-depleted to transcripts, indicating that they underwent weaker selective pressure on the orientation of insertion. When comparing Rett patients with matched healthy controls, we only observed a slight but not significant increase of somatic L1s insertion rate in the Rett group.

In conclusion, our results demonstrate that, in addition to neuronal cells, endogenous L1Hs could retrotranspose in a variety of non-brain tissues and contributes to the inter-cellular genomic diversity. We estimated the rate of somatic L1Hs insertions to be 0.38–1 per PFC neuron and 0.21–0.43 per non-brain cell in healthy individuals. Under pathological condition, MeCP2 dysfunction alters the genomic distribution of L1Hs retrotransposition in Rett patients.

## PgmNr 3146: A novel mutation in *MYBPC1* is associated with mild myopathy and tremor.

### Authors:

I. Inashkina <sup>1</sup>; J. Stavusis <sup>1</sup>; J. Geist <sup>2</sup>; B. Lace <sup>1,3</sup>; N. Wright <sup>4</sup>; S. Pajusalu <sup>5,6</sup>; D. Haubenberger <sup>7</sup>; C. Ward <sup>8</sup>; C.G. Bonneman <sup>9</sup>; A. Kontogianni-Konstantopoulos <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Latvian Biomedical Research and Study Centre, Riga, Latvia; 2) University of Maryland School of Medicine, Department of Biochemistry and Molecular Biology, Baltimore, MD; 3) Centre Hospitalier Universitaire de Québec, Quebec, Canada; 4) Department of Chemistry and Biochemistry, James Madison University, VA; 5) Department of Clinical Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia; 6) Department of Clinical Genetics, Institute of Clinical Medicine, Tartu University, Tartu, Estonia; 7) Clinical Trials Unit, Office of the Clinical Director, NINDS Intramural Research Program, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD; 8) University of Maryland Baltimore, Department of Orthopedics, Baltimore, MD; 9) Neuromuscular and Neurogenetic Disorders of Childhood Section, Neurogenetics Branch, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD

---

*MYBPC1* encodes the slow skeletal Myosin Binding Protein-C protein (sMyBP-C), which belongs to a family of sarcomeric accessory proteins with a major role in the regulation of cross-bridge cycling. sMyBP-C is expressed in both slow and fast twitch skeletal muscles, is highly modular, and plays both structural and regulatory roles.

Here we describe a three-generation family with a skeletal myopathy characterized by muscle weakness, hypotonia, skeletal deformities and spinal rigidity, accompanied by tremor. Exome sequencing analysis of four family member samples revealed the presence of a novel dominant missense mutation in the *MYBPC1* gene, c.742G>A p.E248K. The E248K mutation is located in a highly conserved linker region within the NH<sub>2</sub>-terminus of the molecule. Several *in silico* prognostic tools classify this mutation as deleterious. *In vitro* binding assays using wild type and mutant NH<sub>2</sub>-terminal sMyBP-C proteins revealed that the presence of the E248K mutation results in markedly increased binding to heavy meromyosin (HMM; ~3.5-fold) compared to wild type protein.

Our findings expand the phenotypic spectrum caused by alterations in the *MYBPC1* gene, which were previously associated exclusively with severe and lethal forms of distal arthrogryposis. Based on the familial segregation, absence from the general population, *in-silico* assessment, and altered binding properties of the E248K mutant protein, we can classify it as pathogenic by ACMG guidelines.

To evaluate the impact of the E248K mutation, we used CRISPR technology to generate the respective knock-in mouse model. Our studies show that the E248K mouse model faithfully recapitulates the myopathic phenotype and disease progression observed in patients. While homozygous mice exhibit neonatal lethality, heterozygous mice are significantly smaller, develop tremor, and exhibit behavioral, morphological and contractile deficits compared to their wild-type littermates.

## PgmNr 3147: *Clec16a* knockout compromises autophagy and mitophagy and facilitates ER stress-induced inflammation and neurodegeneration.

### Authors:

H.S. Hain<sup>1</sup>; R. Pandey<sup>1</sup>; M. Bakay<sup>1</sup>; B.P. Strenkowski<sup>1</sup>; H. Hakonarson<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.; 2) Department of Pediatrics, the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Recent studies revealed that CLEC16A is an important regulator of autophagy/mitophagy. Given its association with several autoimmune disorders, we generated an inducible global knockout (KO), *Clec16a*<sup>AUBC</sup> mice to investigate its role in autoimmunity. KO mice exhibited immune dysfunction, severe weight loss through lipolysis, and a neuronal phenotype with an increase in severity over time leading to morbidity. As observed previously, attenuation of CLEC16A disrupted autophagy, mitophagy and caused neurodegeneration. Accelerated mitophagy in cerebellum, cortex, trigeminal ganglia, dorsal root ganglia, and spinal cord was observed in KO mice as compared to control mice. ER stress and mitochondrial dysfunction results in increased oxidative stress and production of multiple proinflammatory mediators. Dysregulated OXPHOS signaling was observed in DRG's and splenic lysates of the *Clec16a*<sup>AUBC</sup> KO mice. KO mice also exhibited an inflammatory cytokine/chemokine profile. KO displayed elevated antibody levels, including IgM, IgA, Ig2b, IgG3 and autoantibodies in sera.

While treatment with a JAK/STAT inhibitor (tofacitinib) partially rescued the lipodystrophic phenotype and improved survival, it only rescued the autophagy in neuronal tissues and did not alter the neuronal phenotype. STAT proteins in neuronal tissues showed no significant change with tofacitinib-treatment as compared to untreated *Clec16a*<sup>AUBC</sup> KO mice. This may be due to the negative feedback loop associated with JAK/STAT/SOC1 signaling.

Our data indicate loss of function variants in *CLEC16A* that are associated with decreased CLEC16A levels may contribute to autoimmunity through elevated ER Stress resulting in dysregulated mitophagy and autophagy, which contributes to adipose lipolysis and production of inflammatory mediators. Drugs modulating ER stress, mitophagy/ autophagy or the JAK/STAT pathway partially reverse the process and may be effective in treating and preventing symptoms of autoimmune disorders in individuals with risk associated *CLEC16A* variants. A combination of such drugs may be optimal for this.



## PgmNr 3148: Unravelling the mechanism underlying *IRF2BPL* in a neurodevelopmental disorder with regression, abnormal movements, loss of speech, and seizures.

### Authors:

P.C. Marcogliese<sup>1,2</sup>; D. Dutta<sup>1,2</sup>; F. Fazal<sup>1,2</sup>; Y. Wang<sup>1,2</sup>; R.C. Spillmann<sup>3</sup>; H. Chung<sup>1,2</sup>; V. Shashi<sup>3</sup>; M.F. Wangler<sup>1,2,4</sup>; S. Yamamoto<sup>1,2,4,5</sup>; L.D.M. Pena<sup>6,7</sup>; H.J. Bellen<sup>1,2,4,5,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX; 3) Division of Medical Genetics, Department of Pediatrics, Duke University Medical Center, Durham, NC; 4) Program in Developmental Biology, BCM, Houston, TX; 5) Department of Neuroscience, BCM, Houston, TX; 6) Cincinnati Children's Hospital, Cincinnati, OH; 7) University of Cincinnati College of Medicine, Cincinnati, OH; 8) Howard Hughes Medical Institute, Baylor College of Medicine, Houston, TX

---

We recently identified 7 individuals who carry *de novo* variants in *IRF2BPL* and are affected with neurological symptoms. Five individuals who carry *IRF2BPL* nonsense variants exhibit a severe neurodevelopmental disorder with regression, abnormal movements, loss of speech, and seizures (NEDAMSS, OMIM-618088). Two individuals with missense variants display milder phenotypes of seizures with global developmental delay or autism spectrum disorder. We showed that the fruit fly *IRF2BPL* ortholog, called *pits*, is broadly expressed, including in the nervous system. Complete loss of *pits* is embryonic lethal, whereas knockdown with RNAi in neurons leads to neurodegeneration, revealing a requirement for this gene in proper neuronal function and maintenance. The *IRF2BPL* nonsense variants behave as severe loss-of-function alleles in flies, while the missense variants lead to milder phenotypes. Our results show that *IRF2BPL/pits* are required in the nervous system in humans and flies, and that their loss leads to a variety of neurological phenotypes.

To better understand the biological function of *IRF2BPL*, the impact of novel variants, and to identify potential targets for therapeutic intervention we embarked on detailed fly studies. Based on cancer studies *IRF2BPL* has been shown to be an E3 ligase that ubiquitinates  $\beta$ -catenin, the transcriptional co-activator responsible for transcribing Wnt target genes. We find that Wnt protein levels are reduced when *IRF2BPL/pits* is elevated, and that Wnt levels are elevated when *IRF2BPL/pits* is reduced. Hence we hypothesize that excess Wnt signaling in patients with NEDAMSS may contribute to the disease. We have generated new models of neurodegeneration caused by loss of *pits*. Firstly, we generated *pits* loss of function clones in the fly retina which causes rapid phototransduction and synaptic transmission defects as measured by electroretinograms. Secondly, we spatially and temporally reduced *Pits::GFP* in neurons with the deGradFP system in flies which results in a decreased lifespan. Using these phenotypes we have generated a platform for drug testing using approved, blood-brain barrier penetrable drugs that inhibit Wnt signaling. Together, our data indicate that *IRF2BPL* regulates WNT protein levels directly which may be important for human disease.

## PgmNr 3149: Single-cell RNA-seq in a mouse model of Parkinson disease reveals potential disease mechanisms affecting subpopulations of dopaminergic neurons.

### Authors:

S. McClymont; A. McCallion

[View Session](#) [Add to Schedule](#)

**Affiliation:** McKusick Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, Maryland.

---

Degeneration of dopaminergic (DA) neurons is the hallmark feature of Parkinson disease (PD). However, not all DA neurons perform the same functions nor are they equally sensitive to degeneration by PD-causing mutations. To help illuminate the subpopulations of DA neurons preferentially degenerated in PD and begin to understand the mechanisms of degeneration, we are examining the transcriptomes of DA neurons from a mouse model of PD.

We are using single-cell RNA-seq (scRNA-seq) to interrogate the consequences of the *SNCA* A53T familial PD mutation on developing (P7) DA neurons of the ventral midbrain. We have collected transcriptomes from >2,400 DA neurons from wildtype, hemizygous, and homozygous mutants bearing the A53T mutation (B6;C3-Tg(*Prnp*-*SNCA*\*A53T)83Vle/J).

Our initial goals were three-fold: 1) Identify subpopulations of DA neurons and their marker genes; 2) Establish which subpopulations are preferentially susceptible to the PD mutation and; 3) Identify differentially expressed genes between genotypes within affected populations to explore disturbed genes and pathways to assess potential mechanisms of action.

We are able to resolve distinct subpopulations of midbrain DA neurons within our scRNA-seq data: the ventral tegmental area (VTA), substantia nigra (SN), periaqueductal grey area (PAG), and a neuroblast population of developing neurons. We are able to further identify five novel subpopulations and associated markers within the SN (three) and PAG (two).

In comparing the distribution of the three genotypes across these subpopulations of DA neurons, we observe an under-representation of mutation-bearing neurons in the neuroblast, VTA, and one of the SN subpopulations, with commensurate increases of mutant cells in the remaining two SN clusters.

We have examined these affected subpopulations for differential gene expression and are investigating the mechanisms by which the mutation might affect neuron differentiation or degeneration. Of note, the A53T mutation appears to decrease the number of DA neurons present in the VTA and within this cluster, the *Aldh1a1* gene is differentially regulated by genotype (adjusted p-value = 0.014), with 100% of wildtype cells and only 55% of homozygous cells expressing *Aldh1a1*. Functional validation via RNAscope of the anatomical location of these subpopulations, their novel marker genes, and the effect of the A53T mutation on select genes and subpopulations is ongoing.

## PgmNr 3150: Spastic paraparesis and dopa-responsive dystonia associated to pathogenic variant in GCH1 gene.

### Authors:

M. Guerra; J.C. Prieto; J.C. Villalta-Fallas; R. Quero; C. Manotas; A. Paredes

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human genetics, Instituto de Genética Humana, Pontificia Universidad Javeriana, Bogotá, Bogotá, Colombia

---

Dopa-responsive dystonia includes a group of inherited disorders characterized for an alteration in the dopamine metabolism, that usually has onset in childhood or adolescence and which main symptom is dystonia that has an overwhelming and sustained response to low doses of L-dopa. It also can be associated with mild parkinsonism, marked diurnal fluctuations and other symptoms although, those are not specific and can be associated to other neurological diseases. It is usually caused by monogenic mutations and has autosomal recessive or dominant inheritance mechanisms. Pathogenic variants have been identified in these genes: GCH-1 (Guanosine triphosphate cyclohydrolase 1), SPR (sepiapterin reductase) and TH (tyrosine hydroxylase). The most common pathophysiological mechanism is a deficiency in bipterin which is a cofactor for tyrosine hydroxylase implicated in dopamine synthesis and is caused by a pathogenic variant in the GCH-1 gene.

In this case report we present a 30-year-old female patient, with a rare clinical presentation that started with gait difficulty, spasticity in lower limbs and hyperreflexia at 6 years old. Initially spastic paraparesis diagnosis was made. Then symptoms remained stable and at 22 years old she presented neck dystonia, so L-dopa treatment was started having well response with complete disappearing of dystonic movements, for that reason dopa-responsive dystonia diagnosis was made. To confirm this diagnosis a sequencing panel testing for dystonia including GCH1, TH and SPR genes was made, showing a pathogenic heterozygote variant in GCH1: c.602G>A (p.Gly201Glu) already correlated with Dopa-responsive dystonia in data bases, nevertheless, this variant in the cases described in the literature, has never been related with spastic paraparesis.

## PgmNr 3151: Curation of mosaic variants in developmental brain disorders: Recommendations from the ClinGen Brain Malformation Curation Expert Panel.

### Authors:

A. Lai <sup>1</sup>; A. Soucy <sup>1</sup>; M. Evenson <sup>3</sup>; R. Siegert <sup>2</sup>; M. DiStefano <sup>2</sup>; C. Yuskaitis <sup>1</sup>; L. Rodan <sup>1</sup>; E. Yang <sup>1</sup>; C. Moufawad El Achkar <sup>1</sup>; G. Mochida <sup>1</sup>; L. Smith <sup>1</sup>; C. Shain <sup>1</sup>; J. Barkovich <sup>7</sup>; Y.S. Lee <sup>3</sup>; D. Miller <sup>1</sup>; E. Marsh <sup>9</sup>; J. Chelly <sup>5</sup>; R. Guerrini <sup>6</sup>; C.A. Walsh <sup>1</sup>; H. Mefford <sup>4,8</sup>; G. Mirzaa <sup>4,8</sup>; A. Poduri <sup>1</sup>; T. Yu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Boston Children's Hospital, Boston, Massachusetts; 2) Broad Institute, Cambridge, Massachusetts; 3) Washington University School of Medicine in St. Louis, St. Louis, Missouri; 4) Seattle Children's Hospital, Seattle, Washington; 5) Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch-Graffenstaden, France; 6) University of Florence, Firenze FI, Italy; 7) University of California San Francisco, San Francisco, California; 8) University of Washington, Seattle, Washington; 9) Children's Hospital of Philadelphia, Philadelphia, Pennsylvania.; 10) Broad Institute, Cambridge, Massachusetts.

---

Somatic mutations arising during the course of development are increasingly recognized as important causes of human diseases, now extending beyond just cancer into the realm of developmental disorders. Proper interpretation of somatic mutations requires consideration of a number of factors not necessarily accounted for in the current 2015 ACMG/AMP Sequence Variant Interpretation Guidelines, which focuses predominately on germline variation. These include: (1) logical rules for assessing mosaic variants that may be detected only at low allele fraction or that (2) may be variably present in different tissues within the same individual, (3) the use of reference databases of somatic variants, and (4) guidelines for predicting gain-of-function activity (since many disease-associated somatic mutations are biochemically activating as opposed to loss-of-function). New and generalizable frameworks are necessary to address these issues.

The ClinGen Brain Malformation Expert Panel was recently formed to curate genes and variants associated with developmental brain malformations, a significant fraction of which have been linked to somatic mutations. In particular, somatic mutations in the MTOR pathway (e.g. *MTOR*, *AKT3*, *PIK3CA* and *PIK3R2*) are known to cause a spectrum of brain overgrowth syndromes ranging from focal cortical dysplasia to dysplastic megalencephaly. Targeted genetic testing for low-level mosaicism has recently become available for many of these malformations.

To curate these genes and variants, we have drafted a framework for the curation of somatic mosaic (post-zygotic) variants that addresses many of the issues unique to somatic variants. We will describe the finalized *MTOR*, *AKT3*, *PIK3CA*, and *PIK3R2* specific variant classification criteria, and present the results of pilot testing of 31 published variants, including 14 likely pathogenic/pathogenic variants, 7 variants of uncertain significance, 9 benign/likely benign, and one with conflicting evidence. Our work should provide a useful framework, not only for standardizing the curation of MTOR pathway mutations, but also for the interpretation of somatic mutations in other settings as well.

## **PgmNr 3152: Family-based genome-wide association study identifies 13q12 associated with major depressive disorder and Alzheimer's disease.**

### **Authors:**

C. Xu <sup>1</sup>; P. Acevedo <sup>1</sup>; B. Su <sup>2</sup>; K. Wang <sup>3</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Health & Biomedical Science, UTRGV, Brownsville, Texas.; 2) Canadian Academy of Natural Health, Toronto Canada; 3) Department of Family and Community Health, School of Nursing, West Virginia University, Morgantown, WV, 26506

**Introduction:** Major depressive disorder (MDD) and Alzheimer's disease (AD) are comorbidity; while MDD may be associated with the development of AD. However, limited study was observed on shared genetic basis between two disorders.

**Material and methods:** we conducted genome-wide association study (GWAS) of MDD and AD to search for novel- and shared - genetic variants for these disorders using a family-based design and to examine whether MDD associated genes are associated with AD (pleiotropic effects) based on 1266 AD cases and 1279 healthy relatives as well as 422 MDD patients and 1688 non-MDD individuals using PLINK and haplotype analysis using generalized estimating equations (FBAT- GEE) statistics in PBAT software.

**Results:** Family-based GWAS identified 12 single nucleotide polymorphisms (SNPs) associated with MDD with  $p < 10^{-5}$ . The most significantly associated SNP was rs930436 ( $p = 1.28 \times 10^{-6}$ ) within PAXIP1 gene, followed by rs1376088 ( $p = 1.97 \times 10^{-6}$ ) within LOC152594. Interestingly, among the 12 SNPs, rs12875434 within LOC646164/DDX39AP1 gene at 13q12 was associated with both MDD ( $p = 3.54 \times 10^{-6}$ ) and AD ( $p = 0.03216$ ). At 13q12, 12 SNPs were associated with MDD; four SNPs were associated with MDD and AD located in one haplotype block. Haplotype analyses further supported the associations of LOC646164/DDX39AP1 gene at 13q12 with MDD and AD.

**In conclusion,** this is the first report of that the variants located 13q12 region were associated with MDD and AD based on GWAS. Future replication and functional studies are required to validate our findings.

## PgmNr 3153: Integrating multi-omics data for gene-level association test identifies schizophrenia risk genes.

### Authors:

Y. Jiang<sup>1,2,3</sup>; D. Zhou<sup>2</sup>; R. Chen<sup>1,2</sup>; C. Chen<sup>3,4</sup>; B. Li<sup>1,2</sup>; C. Liu<sup>3,5,6</sup>; The PsychENCODE Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt University, Nashville, TN, USA; 2) Vanderbilt University Medical Center, Nashville, TN, USA; 3) Central South University, Nashville, TENNESSEE.; 4) Xiangya Hospital, Central South University, Changsha, Hunan, China; 5) SUNY Upstate Medical University, Syracuse, NY, USA; 6) Shaanxi Normal University, Xi'an, Shaanxi, China

Transcriptome-wide association study has been successfully used to explain GWAS signals for psychiatric disorders. Those studies used RNA-Seq data to impute gene expression and performed gene-level association analyses. However, the malfunction of risk genes that affect disorders may occur not only at transcription level, but also at translation level. Different levels of gene expression should have their unique signals that may help identify risk genes. Integrating multi-omics data in gene expression prediction can increase the statistical power.

The PsychENCODE BrainGVEX project generated multi-omics data for over 400 samples in human brain prefrontal cortex region. The omics types include genotypes, RNA-Seq, Ribo-Seq, etc. Ribo-Seq measures gene expression during translation by capturing ribosome-wrapped RNA, which is closer to traits and should be more straightforward to decipher disease risks. However, Ribo-Seq sequences reads of only 27 bps, which may be not so accurate as RNA-Seq. Therefore, gene expression predicted by RNA-Seq and Ribo-Seq data should be complementary at some extent and integrating both models should provide more information.

We performed gene-level association tests to identify schizophrenia risk genes based on GWAS summary data from the Psychiatric Genomics Consortium. To build expression models for RNA-Seq and Ribo-Seq data, we trained the correlation between in-cis SNPs and gene expression. For each gene, we used elastic net with 10-fold cross-validation to select SNPs within gene body and 1Mb flanking region. We used a generalized Berk-Jones test to identify risk genes from the joint two models.

As proof of concept, we first evaluated unique signals in RNA-Seq and Ribo-Seq data. When conditioning on Ribo-Seq data, the correlation between actual and predicted expression for RNA-Seq data dropped from 0.61 to 0.54. When conditioning on RNA-Seq data, the correlation for Ribo-Seq dropped from 0.67 to 0.56. Using the RNA-Seq model only, we identified 342 schizophrenia risk genes. Using the Ribo-Seq model only, we identified 119 risk genes. Integrating the two models, we identified 370 risk genes, which outperforms any single model. Comparing with single models, 307/342 and 113/119 genes were preserved in the joint model. 119 of the 370 genes were located within 500Kb flanking region of 50 known GWAS loci. In a word, integrating multi-omics data can increase the power of identifying risk genes.

## PgmNr 3154: Chromatin accessibility and chromatin QTLs (cQTLs) in the anterior cingulate cortex: Unique and shared characteristics with the dorsolateral prefrontal cortex.

### Authors:

A. Ashley-Koch <sup>1</sup>; M.E. Garrett <sup>1</sup>; L. Song <sup>2</sup>; A. Safi <sup>2</sup>; J. Bryois <sup>3</sup>; G.E. Crawford <sup>2,4</sup>; P.F. Sullivan <sup>3,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC 27701, USA; 2) Center for Genomic and Computational Biology, Duke University, Durham, NC 27708, USA; 3) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-17177 Stockholm, Sweden; 4) Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC 27708, USA; 5) Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599-7264, USA; 6) Department of Psychiatry, University of North Carolina, Chapel Hill, NC, 27599-7264, USA

---

**Background:** To identify non-coding genetic variation relevant to schizophrenia (SCZ) risk, we previously characterized chromatin accessibility in the dorsolateral prefrontal cortex (DLPFC) and identified eight colocalizations of SCZ-associated variants and DLPFC cQTLs, providing a functional context for those genome-wide association (GWA) loci (Bryois et al. 2018). We now extend that work to the anterior cingulate cortex (ACC), which is also implicated in SCZ pathology. **Methods:** We performed ATAC-seq on postmortem ACC samples from 48 SCZ cases and 48 controls. We performed differential chromatin analysis to identify regions associated with SCZ risk and association analysis to identify cQTLs in ACC, DLPFC, or both tissues. **Results:** We identified 89,057 ATAC-seq peaks in the ACC samples, which largely overlapped ones previously identified in the DLPFC. There were no significant differences in ACC chromatin accessibility between SCZ cases and controls. After merging ATAC-seq peaks from ACC and DLPFC for each subject, we identified 32,055 cQTLs significant in both ACC and DLPFC: 21 cQTLs were in a SCZ GWA locus. Five cQTLs were unique to the ACC, 405 cQTLs were unique to the DLPFC, and no tissue-specific cQTLs overlapped SCZ GWA loci. A common haplotype (77% frequent) on chromosome 2 associated with open chromatin drives the most significant DLPFC cQTL. For all ACC-specific cQTLs, the associated SNP was >1kb from the ATAC-seq peak. In contrast, 25% of DLPFC-specific cQTL SNPs were <1kb from the peak and 5% were <100bp. The distance from the SNP to the peak in cQTLs common to both tissues were more similar to DLPFC-specific cQTLs. No gene ontology enrichment was observed among ACC-specific cQTLs. However, DLPFC-specific peaks were enriched for targets of MIR-355 and the interferon gene family (FDR  $q < 0.05$ ). Among cQTLs shared by both tissues, several terms related to the major histocompatibility complex were enriched (FDR  $q < 0.0001$ ). **Conclusions:** Regions of chromatin accessibility and cQTLs are generally shared between ACC and DLPFC in SCZ cases and controls, consistent with Sieberts et al. (2019) demonstrating most eQTLs are shared across brain regions. Nonetheless, some unique characteristics exist. DLPFC cQTL SNPs reside closer to the chromatin peak compared with ACC cQTL SNPs. DLPFC-specific and shared cQTLs are enriched in pathways believed to be important for SCZ risk, underscoring the utility of our approach to dissect SCZ risk mechanisms.



## PgmNr 3155: Allele-specific open chromatin in human iPSC neurons elucidates functional non-coding disease variants.

### Authors:

S. Zhang <sup>1</sup>; H. Zhang <sup>1</sup>; M. Qiao <sup>2</sup>; S. Zhao <sup>2</sup>; Y. Zhao <sup>2</sup>; A. Kozlova <sup>1</sup>; J. Shi <sup>3</sup>; A.R. Sanders <sup>1,4</sup>; G. Wang <sup>2</sup>; S. Sengupta <sup>1</sup>; S. West <sup>1</sup>; M. Streit <sup>1</sup>; M. Chen <sup>5</sup>; C.A. Cowan <sup>6</sup>; W.J. Greenleaf <sup>7</sup>; Z.P. Pang <sup>8</sup>; P.V. Gejman <sup>1,4</sup>; X. He <sup>2</sup>; J. Duan <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Psychiatric Genetics, North Shore Univ Health System, Evanston, Illinois.; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 4) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL; 5) Department of Medicine, University of Chicago, Chicago, IL; 6) Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA; 7) Department of Genetics, Stanford University, Stanford, CA; 8) Department of Neuroscience and Cell Biology and Child Health Institute of New Jersey, Rutgers University, New Brunswick, NJ

---

Chromatin state plays an important role in regulating gene expression and cellular differentiation. Open chromatin landscape during human cortical neurogenesis has been well demonstrated for its ability in influencing gene expression during neurodevelopment; however, the effect of genetic variation on chromatin openness pertaining to neurodevelopmental disorders remain poorly understood. Using human induced pluripotent stem cell (iPSC)-derived neurons as a neurodevelopmental model, we identified abundant open-chromatin regions absent in brains and thousands of genetic variants exhibiting allele-specific open-chromatin (ASoC). We found that ASoC variants are prevalent within the human genome, and neuronal cell type-specific ASoC variants were over-represented in brain enhancers and more likely to affect brain gene expression (eQTL), histone acetylation (hacQTL), and DNA methylation (meQTL). More importantly, neuronal ASoC variants were highly enriched for genetic risk of schizophrenia (SZ), intelligence, educational attainment, bipolar, major depression and Alzheimer's disease. For SZ, we identified 20 ASoC variants showing genome-wide significant associations at 17 risk loci, followed by a systematic functional confirmation using multiplex CRISPRi epigenomic perturbation combined with single-cell RNA-seq (scRNA-seq) in NPCs. In addition, precise nucleotide editing of the top ASoC variant associated with SZ at the *VPS45* locus verified the regulatory effect of the ASoC variant and its regulated genes. Moreover, in combination with computational fine-mapping, we further confirmed the regulatory activities and *cis*-targeted genes. Our results provide the first snapshot of the ASoC landscape in major neuronal cell types relevant to neurodevelopmental disorders, depicted a framework for prioritising functional disease variants, and support ASoC as an effective functional readout of regulatory noncoding disease risk regions.



## PgmNr 3156: Metabolic dysfunction as a contributor to 3q29 deletion syndrome phenotypes.

### Authors:

R.M. Pollak<sup>1</sup>; R.H. Purcell<sup>2</sup>; T.P. Rutkowski<sup>1</sup>; T. Malone<sup>1</sup>; K. Pachura<sup>3</sup>; G. Wynn<sup>3</sup>; G.J. Bassell<sup>2</sup>; M.P. Epstein<sup>1</sup>; T. Caspary<sup>1</sup>; P.A. Dawson<sup>3</sup>; D.P. Jones<sup>4</sup>; S.T. Warren<sup>1</sup>; D. Weinshenker<sup>1</sup>; M.E. Zwick<sup>1</sup>; J.G. Mulle<sup>1</sup>; The Emory 3q29 Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA.; 2) Department of Cell Biology, School of Medicine, Emory University, Atlanta, GA.; 3) Department of Gastroenterology, Department of Pediatrics, School of Medicine, Emory University, Atlanta, GA.; 4) Division of Pulmonary, Allergy, and Critical Care Medicine, School of Medicine, Emory University, Atlanta, GA.

---

3q29 deletion syndrome (3q29Del) is caused by a rare (~1:30,000) 1.6 Mb heterozygous deletion on chromosome 3 and is associated with a wide range of phenotypes, including mild to moderate intellectual disability; increased risk for autism, ADHD, and schizophrenia; and reduced birth weight and growth deficits. Individuals with 3q29Del have significantly reduced birth weight compared to matched controls (15.04oz difference,  $p=1.5E-6$ ); this phenotype is recapitulated in the mouse model of 3q29Del created at Emory University (1.81g difference, 8% of body weight,  $p=8.8E-12$ ). Based on the consistency of this phenotype and the presence of four mitochondria- or metabolism-associated genes within the 3q29 interval (*Pcyt1a*, *Tfrc*, *Bdh1*, and *Senp5*), we hypothesize that there may be a previously unidentified metabolic dysfunction in individuals with 3q29Del. Using untargeted metabolomics in liver tissue from our mouse model, we found that palmitoylcarnitine was significantly increased in 3q29 animals after adjusting for sex ( $p<0.05$ ,  $FC=0.314$ ). In stratified analyses, phosphatidate was significantly increased in female mutants versus wild-type (WT) littermates ( $p=0.02$ ,  $FC=1.997$ ), with males showing a similar trend ( $p>0.05$ ). Metabolome-wide association studies identified the carnitine shuttle and fatty acid metabolism pathways as significantly associated with these target metabolites ( $p<0.05$ ), suggesting that mitochondrial function and fat metabolism may be impacted by the 3q29 deletion. We used metabolic chambers to measure gross features of metabolism in our mouse model, including food and water consumption, activity, and energy expenditure. On standard chow, the 3q29Del mice were not significantly different from their WT littermates for any of the parameters measured. These data confirm that the 3q29Del-associated weight deficit and metabolic features are not due to decreased feeding or increased energy expenditure, and instead likely result from altered metabolism. Studies are underway to assess specific deficits in mitochondrial function and fatty acid metabolism in the mouse model of 3q29Del; preliminary results suggest that the 3q29Del-associated weight deficit is ameliorated when mice are fed a high-fat diet. Together with our metabolomics results, these data may highlight metabolic pathways relevant to intellectual disability, autism, and other neurodevelopmental and neuropsychiatric phenotypes.

## PgmNr 3157: Weighted gene co-expression network analysis of the human frontal cortex reveals novel insights into the schizophrenia-associated 3q29 deletion syndrome.

### Authors:

E. Sefik <sup>1</sup>; R.H. Purcell <sup>2</sup>; G.J. Bassell <sup>2</sup>; E.F. Walker <sup>3</sup>; J.G. Mulle <sup>4</sup>; The Emory 3q29 Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Neuroscience PhD Program, Emory University, Atlanta, GA; 2) Department of Cell Biology, Emory University School of Medicine, Atlanta, GA; 3) Department of Psychology, Emory University, Atlanta, GA; 4) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA

---

**Background:** The 3q29 deletion is a recurrent copy number variant resulting in hemizygous deletion of 21 protein coding genes; the syndrome is characterized by developmental delay (DD) and increased risk for autism spectrum disorders (ASD) and schizophrenia (SZ). The functional role of 3q29 genes and the mechanism underlying the neurodevelopmental sequelae of their hemizygosity remain largely unknown. **Objectives:** To interrogate functional consequences of the 3q29 deletion, we built an undirected gene co-expression network of the human frontal cortex and leveraged the principle of guilt-by-association to generate predictions about pathways and candidate genes contributing to syndrome phenotypes. **Methods:** We applied weighted gene co-expression network analysis (WGCNA) to publicly available RNA-Seq data (Genotype-Tissue Expression Project) and grouped genes into clusters that are likely functionally related. We established preservation in another dataset (BrainSpan Developmental Transcriptome Project), indicating high reliability of the network. We then identified clusters that harbor 3q29 genes and defined their top neighbors. We tested whether known risk genes for SZ, ASD and DD are overrepresented among these neighbors. Consequently, we generated a list of candidate drivers that are highly co-expressed with 3q29 genes and are associated with risk for SZ, ASD and/or DD. We tested the enrichment of this list for biological pathways and tissue expression. **Results:** The resulting data suggest that 3q29 interval genes are organized into seven highly correlated groups with overlapping effects on the same biological pathways. We identified *UBXN7* as a hub-gene, suggesting central influence on a large network of neighbors. We also revealed that ten 3q29 genes have top neighbors that are enriched for SZ, ASD and/or DD-risk, resulting in a prioritized list of 163 candidate drivers. This list maps onto key biological processes such as chromatin organization, protein-protein interactions at synapses and NMDA receptor activity. Intriguingly, this list is enriched for expression in the granular layer of the cerebellum, a brain region with abnormal volume in 3q29 deletion carriers. **Conclusion:** The molecular pathways and prioritized gene list identified in this study will inform new directions of neurobiological inquiry. Emerging transcriptional profiling data from the Emory 3q29 Project (both human cellular models and mouse models) will be used to validate our predictions.

## PgmNr 3158: Differentially expressed genes and eQTLs implicate oxidative stress and mitochondrial function in risk for PTSD.

### Authors:

M. Garrett <sup>1</sup>; X. Qin <sup>1</sup>; D. Mehta <sup>2</sup>; MIRECC. Working Group <sup>3</sup>; PTSD. Initiative <sup>4,5</sup>; M. Hauser <sup>1</sup>; A. Ashley-Koch <sup>1</sup>; Psychiatric Genomics Consortium PTSD Group

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC, USA.; 2) Institute of Health and Biomedical Innovation, School of Psychological and Counselling, Queensland University of Technology, Kelvin Grove, Queensland, Australia.; 3) Durham Veterans Affairs Medical Center, Durham, NC, USA; The VA Mid-Atlantic Mental Illness Research, Education, and Clinical Center, Durham, NC, USA; Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA.; 4) Gallipoli Medical Research Foundation, Greenslopes Private Hospital, Greenslopes, Queensland, Australia.; 5) Queensland University of Technology, Queensland, Australia.

---

PTSD is a complex psychiatric disorder that can develop following traumatic events. The Psychiatric Genomics Consortium PTSD group (PGC-PTSD) has collected over 20,000 multi-ethnic PTSD cases and controls and has identified both genetic and epigenetic factors associated with PTSD risk (Duncan et al., 2018; Wolf et al., 2018). To further investigate biological mechanisms underlying PTSD risk, we examined three PGC-PTSD cohorts comprising 977 subjects to identify differentially expressed genes among PTSD cases and controls. Whole blood gene expression was quantified with the HumanHT-12 v4 Expression BeadChip for 726 OEF/OIF veterans from the VA MIRECC (Ashley-Koch et al., 2015), 155 samples from the INTRuST Clinical Consortium (Bomyea et al., 2019), and 96 Australian Vietnam War veterans (Mehta et al., 2017). Probe and sample quality control, background adjustment, and normalization were performed using limma and batch effects were corrected using ComBat. The proportion of monocytes and lymphocytes for each sample was estimated using an approach described in Logue et al. (2015). Differential gene expression analysis was performed in each cohort separately controlling for age, ethnicity, sex (if applicable), and cell proportions using limma, followed by meta-analysis using METAL. Inflation and bias were controlled using the empirical null distribution. Finally, we performed expression QTL (eQTL) analysis in the MIRECC cohort followed by pathway analysis. While no gene survived false discovery, the most differentially expressed genes were *EXOC6* ( $p=9 \times 10^{-6}$ ) and *IL1B* ( $p=3 \times 10^{-5}$ ). *EXOC6* has been nominally associated with major depressive disorder (MDD) risk, a condition commonly comorbid with PTSD (Wray et al., 2012). *IL1B*, a pro-inflammatory cytokine, is associated with PTSD in humans (Hovhannisyan et al., 2017) and involved in stress-enhanced fear learning in rodents (Jones et al., 2015). We also identified 2,710 FDR significant eQTLs in the MIRECC cohort. eQTL target genes were enriched for mitochondrion and oxidoreductase GO terms. Mitochondrial function is associated with PTSD (Mellon et al., 2019), as well as with innate immune response in MDD (Culmsee et al., 2019). In conclusion, we observed differential gene expression between PTSD cases and controls in genes and pathways involved in oxidative stress, innate immune response and associated with psychiatric disorders often comorbid with PTSD. These data provide further insight into PTSD pathomechanisms.

## PgmNr 3159: Transcriptomic network signature of synaptic pruning in prefrontal cortex with increased copy number of complement component 4.

### Authors:

M. Kim <sup>1,2</sup>; J.R. Haney <sup>3,4</sup>; M.J. Gandal <sup>1-4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles; 2) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles; 3) Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles; 4) Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles

---

**Introduction:** Genome-wide association studies (GWAS) have successfully identified hundreds of genomic regions associated with schizophrenia. Subsequent fine-mapping of the strongest association signal, which lies in the major histocompatibility complex (MHC) locus, found that risk for schizophrenia was partially mediated by increased expression of complement component 4A (*C4A*) arising from complex structural variation of *C4* genes. Although *C4A* is believed to partake in synaptic pruning, its precise function in human brain—and its relation to other schizophrenia risk variants—remains difficult to examine due to lack of an appropriate experimental system and lack of evolutionary conservation in model organisms.

**Methods:** Here, we integrated genotype array and RNA-seq data of frontal cortex samples from 344 individuals with schizophrenia and 422 controls as part of the PsychENCODE project. To provide human brain-specific functional annotations, we imputed *C4* structural alleles and generated *C4A*-seeded co-expression networks stratified by *C4A* copy number. We then assessed whether resulting networks show enrichment for known pathways, cell-types, and GWAS SNP-heritability.

**Results:** *C4A*-seeded networks expanded in size with increasing *C4A* copy number and exhibited greater connectivity in frontal cortex compared to other brain regions. Positive co-expression partners were enriched for astrocytic and microglial markers as well as NFkB signaling pathway genes, whereas negative co-expression partners were enriched for synaptic/neuronal markers. We find that negative—but not positive—co-expression partners of *C4A* exhibited substantial enrichment for SNP-heritability in schizophrenia, providing strong evidence for additive polygenic effects in schizophrenia. Consistent with this result, negative co-expression partners were expressed in a limited set of neuronal cell-types that have previously been shown to be enriched for schizophrenia GWAS signals.

**Conclusion:** Functional genomic analyses of a large-scale human brain dataset could provide novel insights into *C4A*-mediated biological process and implicate that frontal cortex might be particularly susceptible to this process. We identify a gene expression network signature likely representing synaptic pruning that integrates multiple genetic risk factors for schizophrenia and thereby highlights some level of convergence in schizophrenia pathophysiology.

## PgmNr 3160: An iPSC-derived neuronal model of the schizophrenia-associated 3q29 deletion.

### Authors:

R.H. Purcell <sup>1</sup>; E. Sefik <sup>2</sup>; E. Evans <sup>1</sup>; S. Karne <sup>1</sup>; M.M. Murphy <sup>2</sup>; R.M. Pollak <sup>2</sup>; T. Mosley <sup>2</sup>; M.M. Merritt-Garza <sup>1</sup>; N. Raj <sup>1</sup>; Z.T. McEachin <sup>1</sup>; J.G. Mülle <sup>2</sup>; G.J. Bassell <sup>1</sup>; The Emory 3q29 Project

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Cell Biology, Emory University, Atlanta, GA.; 2) Department of Human Genetics, Emory University, Atlanta, GA.

---

A major goal in psychiatric genetics is to determine the biological mechanisms of schizophrenia risk genes. A fundamental challenge is that common variants confer very low risk individually and therefore must act in concert with many other factors. By contrast, rare, high-risk copy number variants (CNVs) offer a more direct path from genotype to understanding risk gene mechanisms. The 3q29 deletion is a highly-penetrant, recurrent CNV that is associated with neurodevelopmental disorders and confers a striking 40-fold increased risk for schizophrenia, and therefore represents a high-value target for mechanistic study.

The Emory 3q29 Project (<http://genome.emory.edu/3q29/>) was formed to advance study of the 3q29 deletion, and unites a team of investigators toward understanding the phenotypic spectrum, natural history, and molecular mechanism of the deletion. One specific goal of The Emory 3q29 Project is to investigate the effects of this CNV on neuronal development. In this ongoing study we have now obtained peripheral blood mononuclear cells from six individuals with the 3q29 deletion enrolled in The Emory 3q29 Project and created induced pluripotent stem (iPS) cell lines. Neurotypical controls were recruited for comparison and matched by age, sex, and race. Following in-depth characterization of iPSC lines, we used dual-SMAD inhibition to differentiate all lines to PAX6, Nestin, and SOX2 positive neural progenitor cells (NPCs). Transcriptional profiling of NPC lines is underway and will be presented at the meeting.

All NPC lines have been further differentiated to forebrain neuronal cultures that stain positive for MAP2, beta-III tubulin, CTIP2, TBR1, CUX1, and SATB2. Preliminary morphological analyses indicate that 3q29del neurons display increased neurite branching and shorter primary neurites at 2 weeks post-differentiation (wpd). Additionally, CTIP+ neurons at 4wpd display somewhat simplified neuronal complexity. Transcriptomic and proteomic analyses of 3q29del neurons are ongoing. These preliminary results suggest that haploinsufficiency of multiple 3q29 genes impinges on key neuronal pathways, impairing development and maturation.

**Acknowledgements:** This study is supported by NIH MH110701 (J.G.M and G.J.B) and the Emory Integrated Genomics Core (EIGC), Emory Stem Cell Core, and Emory Integrated Computational Core, which are subsidized by the Emory University School of Medicine.

## PgmNr 3161: Convergent pathways of 22q11.2 deletion and schizophrenia in human neurons.

### Authors:

O. Pietilainen<sup>1,2</sup>; R. Nehme<sup>1,2</sup>; M. Artomov<sup>1</sup>; A. Ganna<sup>1</sup>; T. Singh<sup>1</sup>; S. McCarroll<sup>1</sup>; S. Hyman<sup>1,2</sup>; B. Neale<sup>1,3,4,5,6</sup>; L. Barrett<sup>1,2</sup>; A. Palotie<sup>1,3,4,5,6</sup>; M. Daly<sup>1,3,4,5,6</sup>; K. Eggen<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA; 2) Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA; 3) Psychiatric & Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 5) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA; 6) Department of Neurology, Massachusetts General Hospital, Boston, MA

---

The most common human genomic rearrangement is a 3 Mb deletion on chromosome 22q11.2, occurring once in every 4,000 live births. The 22q11.2 deletion syndrome is associated with variety of maladies with particularly high risk to schizophrenia and childhood-onset neurodevelopmental disorders. Yet, it is unclear how the deletion acts to manifest the high risk for these presumably cerebrocortical phenotypes, and how much could be learned from these disorders by studying this high impact regulatory variant. To gain insight into these questions, we performed RNA sequencing in human induced pluripotent stem cells (hiPSC), neuronal progenitors (NPs) and post-mitotic excitatory neurons derived from 49 donors, of which 20 carried the 22q11.2 deletion. We were able to detect a statistically significant reduction in gene expression of genes located within the deletion region across all three cell stages (FDR<5%) where the deletion explained ~50% of variance in the RNA abundances. The majority (89%) of the differentially expressed genes were located outside the deletion and was mostly cell-stage specific, suggesting that the regulatory effect of the 22q11.2 deletion changes with progression of neuronal development. We then turned our attention to genes that were differentially regulated by the stem cells and early NPs. We found that the differentially expressed genes in these early developmental stages had more than expected protein-protein interactions with genes that have been previously linked with early-onset neurodevelopmental disorders ( $p=0.046$ ). However, this enrichment was not present in neurons. We next asked whether gene variants in the differentially expressed genes contribute disproportionately to the heritability of common neuropsychiatric traits and disorders. We found that variants in up-regulated genes in neurons from deletion carriers were significantly enriched for heritability particularly for schizophrenia ( $p=4.2 \times 10^{-6}$ ). Finally, the 22q11.2 deletion using CRISPR was sufficient to induce the directionality of effects observed in the patient collection. Together, these results suggest that the 22q11.2 deletion has a regulatory effect on pathways that converge with the polygenically defined schizophrenia in neurons, and with genetic variants implicated in early neurodevelopmental disorders in progenitors.



## PgmNr 3162: Exonic resequencing of the genes encoding the ionotropic glutamate receptors of kainate subtypes in patients with schizophrenia.

### Authors:

M. Cheng; T. Hu; S. Hsu; H. Tsai

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Psychiatry, Yuli Branch, Taipei Veterans General Hospital, Hualien county, Taiwan

---

Schizophrenia is a chronic debilitating mental disorder with a high genetic component in its etiology. The ionotropic glutamate receptor kainate type plays a critical role in regulating the synaptic transmission and the functions of various synaptic receptors. Genes encoding the ionotropic glutamate receptors of kainate subtypes can be considered as the candidate genes of schizophrenia. There has been growing debate that rare mutations in synapse-associated protein genes are rare and may contribute to parts of the pathogenesis of schizophrenia in certain patients. This study aims to identify the rare and pathogenic mutations of the glutamate ionotropic receptor kainate type subunit (GRIK) gene family members in patients with schizophrenia. We resequenced the exon regions of 5 genes (*GRIK1*, *GRIK2*, *GRIK3*, *GRIK4*, and *GRIK5*) in 420 unrelated patients with schizophrenia using ion semiconductor sequencing method. We identified several protein-disrupting mutations in patients with schizophrenia, and *in silico* analysis showed that some of these mutations were damaging or pathological to the protein function. Notably, we identified one frameshift deletion (*GRIK1*:c.70\_71delTT) in one schizophrenic patient and two heterozygous nonsense mutations (*GRIK2*:c.898C>T; *GRIK4*:c.1024C>T) in two patients with schizophrenia. The *GRIK1*:c.70\_71delTT mutant was predicted to change the reading frame (*GRIK1*<sup>Phe24fs</sup>), resulting in a completely different translation from the original GRIK1 protein. The *GRIK2*:c.898C>T mutant was predicted to cause the stop codon in codon 300 of the GRIK2 protein (*GRIK2*<sup>Arg300Ter</sup>). The *GRIK4*:c.1024C>T mutant was predicted to cause the stop codon in codon 342 of the GRIK4 protein (*GRIK4*<sup>Gln342Ter</sup>). Because, these three protein-disrupting mutations were absent in the ExAC dataset and 1517 healthy controls from Taiwan BioBank, we presumed these three rare mutations are schizophrenic-associated variants. The study suggests that the GRIK genes harbor rare pathogenic mutations in certain patients with schizophrenia, supporting rare coding variants may contribute to the genetic architecture of schizophrenia. Future studies using cells or transgenic mouse models carrying these mutations are necessary to understand the functional consequence of the GRIK gene family mutants *in vitro* and *in vivo* and, most importantly, to elucidate whether and how they contribute to the etiology of schizophrenia.

## PgmNr 3163: Alternative splicing: Non-canonical isoform REEP6.1 exhibits tissue-specific function.

### Authors:

N. Wu; Q. Liang; S. Agrawal; H. Liu; S. Fu; R. Bertrand; Y. Li; R. Chen

[View Session](#) [Add to Schedule](#)

**Affiliation:** Baylor College of Medicine, Houston, TEXAS.

---

Alternative splicing plays a role in increasing the complexity of gene products in eukaryotic cells. However, knowledge about the extent that the sequence complexity is translated to function complexity among tissues remains limited. For example, it is largely unknown whether isoforms that are expressed in a tissue-specific manner correlate with their tissue-specific functions. Previous studies from our lab have linked a deficiency in Receptor Expression Enhancer Protein 6 (*REEP6*) with retinitis pigmentosa (RP), an inherited retinal disease. Two isoforms of *REEP6* have been identified: both are conserved in mice and humans. While the canonical isoform, *REEP6.2*, is expressed broadly in many organs, the non-canonical isoform *REEP6.1* is expressed mainly in rod photoreceptor cells. Here, we investigated whether *REEP6.1* has retina-specific functions that cannot be substituted by *REEP6.2*. *REEP6.1* contains a 27-amino-acid region encoded by Exon 5 of the *REEP6* gene. *Reep6* exon 5 deletion mice (E5 mice) have been generated, which eliminates the *Reep6.1* isoform while leaving the *Reep6.2* isoform unchanged. Phenotypes of different *Reep6* genotypes, including *Reep6*<sup>E5/+</sup>, *Reep6*<sup>E5/E5</sup>, *Reep6*<sup>E5/-</sup>, with *Reep6*<sup>6+/-</sup> and *Reep6*<sup>-/-</sup> serving as positive and negative controls respectively are characterized. *Reep6*<sup>E5/E5</sup> mice showed a moderate retinal degeneration phenotype while *Reep6*<sup>E5/-</sup> mice showed a phenotype as severe as the complete knock-out (*Reep6*<sup>-/-</sup>). In parallel, we performed rescue experiments by overexpressing *Reep6.1* and *Reep6.2* cDNA in *Reep6*<sup>-/-</sup> mice. We found the *Reep6.2* is less efficient than *Reep6.1* in rescuing the retinal defect. Taken together, our results indicate that the retina specific isoform of *Reep6* plays a distinct function from the canonical isoform, providing critical information for further investigation of its function in the retina, as well as future development of therapy.



## PgmNr 3164: Novel compound heterozygous, nonsense variants p.L150\* and p.Y3565\* of the *USH2A* gene in a Chinese pedigree links to Usher syndrome type IIA.

### Authors:

J. Fu <sup>1</sup>; J.L. Fu <sup>1</sup>; J. Cheng <sup>1</sup>; Q. Zhou <sup>2</sup>; M.A. Khan <sup>1</sup>; C. Duan <sup>2</sup>; H. Lv <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Key Laboratory of Epigenetics and Oncology, the Research Center for Preclinical Medicine, Southwest Medical University, Luzhou, Sichuan, China; 2) Department of Ophthalmology, the Affiliated Hospital of Southwest Medical University, Luzhou, Sichuan, China

---

**Purpose:** Usher syndrome includes a group of genetically and clinically heterogeneous autosomal recessive disease with retinitis pigmentosa and hearing deficiencies. The relationships between the Usher syndrome-caused genes and the resultant Usher syndrome phenotypes in the patients are highly variable. **Methods:** In the present study, an Usher syndrome Chinese family was recruited, targeted next-generation sequencing, Sanger sequencing, and segregation analysis were fulfilled. The expression profile was analyzed. Functional effects of the pathogenic variants for *USH2A* were analyzed. **Results:** We revealed novel, nonsense compound heterozygous variants c.T449G (p.L150\*) and c.T10695A (p.Y3565\*) in the *USH2A* gene, which showed co-segregated with the disease phenotype causing Usher syndrome type IIA in our Chinese pedigree. The p.L150\* variant produced truncated small protein with losing almost all functional domains of *USH2A*; whereas the p.Y3565\* variant is located in one of the fibronectin type 3 domains with losing several fibronectin type 3 domains at the C-terminus of *USH2A* by producing truncated protein. We found that the *USH2A* mRNA was high expressed in the retina of the eye and this protein was highly conserved by comparing *H. sapiens* *USH2A* to eight other species. **Conclusion:** This is the first to reveal two novel, pathogenic compound heterozygous variants c.T449G (p.L150\*) and c.T10695A (p.Y3565\*) in the *USH2A* gene in a Chinese family with Usher syndrome type IIA, thereby expanding *USH2A* mutation spectrums. Our discoveries can help understand molecular pathogenesis including RP and Usher syndrome type IIA, developing strategies of diagnosis, therapy and genetic counseling in this Usher syndrome type IIA disease. **Funding:** The project was funded by the National Natural Science Foundation of China (31701087, 30371493, and 81672887).

## PgmNr 3165: Submicroscopic 13q32.1 deletions causing congenital microcoria modify the regulatory landscape of neighboring genes.

### Authors:

L. Fares Taie <sup>1</sup>; B. Nedelec <sup>1</sup>; C. Angée <sup>1</sup>; P. David <sup>2</sup>; S. Gerber <sup>1</sup>; S. Crippa <sup>3</sup>; B. Passet <sup>4</sup>; J.L. Vilotte <sup>4</sup>; N. Chassaing <sup>5</sup>; J. Kaplan <sup>1</sup>; D. Lupiañez <sup>6</sup>; C. Kostic <sup>3</sup>; P. Calvas <sup>5</sup>; J.M. Rozet <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory of Genetics in Ophthalmology (LGO). INSERM UMR1163. Imagine – Institute of Genetic Diseases, Paris Descartes University, 75015 Paris, France; 2) LEAT Imagine – Institute of Genetic Diseases, Paris Descartes University, 75015 Paris, France; 3) University of Lausanne· Ophthalmology. Jules-Gonin Eye Hospital. Fondation Asile des Aveugles. CP 5143 1002, Laussane, Switzerland; 4) Génétique Animale et Biologie Intégrative, Institut Nationale de la Recherche Agronomique, UMR1313. 78352 Jouy-en-Josas, France; 5) Service de Génétique Clinique, Hôpital Purpan, 31300 Toulouse, France; 6) Berlin Institute for Medical Systems Biology (BIMSB).Max Delbrück Center for Molecular Medicine. 13125, Berlin, Germany

**Purpose.** Congenital microcoria (MCOR) is a rare autosomal dominant disease characterized by unreactive pinhole pupils lacking iris dilator muscle (DM), iris transillumination and primary open angle glaucoma (POAG; > 30%). Previously, we reported 13q32.1 deletions as the cause of the disease. *GPR180* and *TGDS* were consistently altered, but the loss of either gene individually does not cause MCOR. We propose that MCOR is due to the modification of the regulatory landscape of the region which encompasses *dct* involved in iris pigmentation, *Sox21* that regulates the expression of *Pax6* involved iris development and *Abcc6* playing a role in the maintenance of the intraocular pressure (IOP), the increase of which is a major determinant of POAG. The purpose of the study was to assess this hypothesis by analyzing mice carrying deletions of variable size within or around the minimal MCOR deletion (35KB).

**Results.** Heterozygous animals carrying the 35 Kb deletion, presented with moderate, yet statistically significant, reduction in pupil diameter. RNAseq analysis revealed deregulation of a single neighboring gene, *Sox21*. The gene displayed ectopic expression in the iris, starting before the formation of the muscle during embryonic life, through to adulthood. Ablation of insulators within the MCOR deletion had no effect on neighboring genes. ChIPseq-H3K27Ac data identified two iris *Dct* enhancers upstream the deletion that might interact with *Sox21*. 4C analysis from the viewpoint of 2 kb surrounding *Sox21* showed the presence of ~1Mb TAD at the MCOR locus and that *Sox21* interaction extend from *Dct* through to *Uggt2*. TAD organization did not change dramatically in MCOR mice. Combined analysis of RNAseq and ChIPseq datasets revealed the deregulation in MCOR irises of two *Sox21* targets, including a protein known to play a major role in the maintenance of IOP.

**Discussion.** Here, we show that MCOR deletions alter the regulatory landscape of 13q32.1 in the iris. The ectopic *Sox21* expression in the developing iris likely changing the fate of the progenitors that should give rise to the DM. We suggest that in the absence of their endogenous targets (*Tgds* and *Gpr180*), *Dct* enhancers might activate *Sox21* more efficiently by altering promoter competence or by

reducing genomic distances within the TAD. Finally, we propose that POAG in MCOR patients might be secondary to the deregulation *Sox21* instead of *Abcc6*.

## PgmNr 3166: Mutations in the functional amyloid PMEL lead to pigmentary glaucoma.

### Authors:

M.A. Walter <sup>1</sup>; T. Footz <sup>1</sup>; O. Lehmann <sup>1,2</sup>; W.T. Allison <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Medical Genetics, Univ Alberta, Edmonton, Alberta, Canada; 2) Dept Ophthalmology, Univ Alberta, Edmonton, Alberta, Canada; 3) Dept Biological Sciences, Univ Alberta, Edmonton, Alberta, Canada

---

Pigmentary glaucoma (PG) is the most common secondary glaucoma making it a major cause of blindness worldwide. Recently, we have reported that mutations in the melanosome structural protein premelanosome protein (PMEL) underlie PG, the first known genetic cause of this form of blindness in humans (Lahola-Chomiak et al., 2018). Functional analysis of non-synonymous PMEL missense mutations found in patients revealed 7/9 mutations displayed defects in proteolytic cleavage and/or melanosome fibril formation. CRISPR-Cas9 disruption of the homolog *pmel* in zebrafish caused profound pigmentation defects and glaucoma-like phenotypes further supporting PMEL's role in PG. However, the mechanism by which PMEL mutations cause glaucoma is yet to be elucidated. PMEL is a rare protein that forms functional amyloid fibers at normal physiological conditions; indeed PMEL fibers constitute the scaffold in melanosomes upon which melanin is synthesized and deposited. We hypothesize that PMEL mutations are gain-of function mutations that switch PMEL from a functional to a pathological amyloid. In the eye, the trabecular meshwork located in the anterior segment is the primary drainage route of aqueous humor. We now show that PMEL-containing cellular debris is phagocytosed by ocular trabecular meshwork cells which may compromise cell survival. As loss of trabecular meshwork cells results in IOP homeostasis deregulation, these findings could reveal one of the initiating steps of pigmentary glaucoma pathogenesis.

## PgmNr 3167: *CPH1*, a new gene for dominant optic atrophy.

### Authors:

C. Delettre<sup>1,2</sup>; C. Mégy<sup>1,2</sup>; A. Tarres-Solé<sup>3</sup>; F. Hensen<sup>3</sup>; M. Pequignot<sup>1,2</sup>; M. Quiles<sup>1,2</sup>; C.P. Hamel<sup>1,2,4</sup>; A. Müller<sup>1,2</sup>; G. Manes<sup>1,2</sup>; A. Roubertie<sup>1,4</sup>; B. Bocquet<sup>1,2,4</sup>; H. Wilhelm<sup>5</sup>; U. Kellner<sup>6</sup>; N. Weisschuh<sup>7</sup>; G. Lenaers<sup>8,9</sup>; B. Wissinger<sup>7</sup>; X. Zanlonghi<sup>10</sup>; I.A. Meunier<sup>1,4</sup>; M. Solà Vilarrubias<sup>3</sup>; J.N. Spelbrink<sup>11</sup>; E. Sarzi<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institut des Neurosciences de Montpellier, INSERM U1051, Montpellier, Hérault, France; 2) University of Montpellier; 3) Molecular Biology Institute Barcelona, Spain; 4) Centre of Reference for Genetic Sensory Diseases, CHU, Montpellier, France; 5) University Eye Hospital, Centre for Ophthalmology, Tübingen, Germany; 6) Rare Retinal Disease Center, Siegburg, Germany; 7) Institute for Ophthalmic Research, Tübingen, Germany; 8) Institut MitoVasc, France; 9) Université d'Angers, France; 10) Clinique Jules Verne, Nantes, France; 11) Radboud Center for Mitochondrial Medicine, , Netherlands

---

Hereditary Optic Neuropathies (HON) are genetically heterogeneous diseases. HON can be isolated, affecting exclusively the retinal ganglion cell, or syndromic. HON are linked to genes coding proteins with a major role in the mitochondrial function. Our purpose is to characterize a new gene causing HON, called CPH1.

Whole-exome sequencing (WES) method was used to characterize the genetic cause of a new form of Dominant Optic Atrophy associated with a maculopathy. Sanger sequencing was used to verify the mutations. Then, we used fibroblasts to characterize mitochondrial dysfunction in patient cells.

WES analysis was performed on several members of a very large family with a dominant optic atrophy (DOA) and a photoreceptor anomaly in the central part of the retina. A dominant mutation was identified in a candidate gene called CPH1. This gene encodes a protein involved in the mitochondrial function. We then screened *CPH1* in a large cohort of DOA families and found additional mutations. Our results on patients' fibroblasts show that CPH1 protein expression is decreased compared to controls, whereas the mRNA level of CPH1 remained unchanged. To uncover the structural features underlying CPH1 mutations, we determined CPH1 crystal structure. Structural analysis suggests that mutations cause a disarrangement crucial motif of the protein. Using patient fibroblasts, we validated that CPH1 variant destabilizes the protein and affects mitochondrial function. We used morpholinos and CPH1 RNA mutant to model the disease in zebrafish. First results show defect in eye development in mutant fish.

These results suggest that CPH1 is a novel gene associated with dominant optic atrophy. Mechanistic study indicated that the mutation in this gene involves severe mitochondrial dysfunction.

## PgmNr 3168: Loss of function variants rs1971050, rs10038177 and rs10038058 of *WDR36* are not associated with primary open angle glaucoma in a Cameroon population.

### Authors:

F. Mbacham<sup>1,4</sup>; Y. Bilong<sup>1,2</sup>; A. Nomo<sup>1,2</sup>; H. Dongmo<sup>3,6</sup>; C. Tah<sup>3</sup>; J.C. Katte<sup>1,5</sup>; W. Mbacham<sup>1,3,6</sup>; E. Sobngwi<sup>1,3,5</sup>; B. Assumpta<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculty of Medicine and Biomedical Sciences, University of Yaounde 1, Cameroon; 2) Ophthalmology Unit, Gynecologic-Obstetrics and Pediatric Hospital, Yaounde, Cameroon; 3) Biotechnology Center, University of Yaounde 1, Cameroon; 4) Institut Supérieure des Sciences et de la Santé, Université des Montagnes, Banganté, Cameroun; 5) National Obesity Center, Yaounde Central Hospital, Cameroon; 6) Faculty of Science, University of Yaounde 1, Cameroon

In primary open angle glaucoma (POAG), the most common form of glaucoma in people of African origin, it is established that retinal ganglion cells are lost due to apoptosis. Loss of *WDR36* (OMIM 609669) function has been shown to result in activation of the p53 stress- response pathway, a key regulator of apoptosis. However, there is controversy surrounding its contribution in the pathogenesis of POAG. We aimed to establish association between rs1971050, rs10038177 and rs10038058, *WDR36* gene polymorphisms and POAG. We assessed 798 glaucoma medical records and a selected 209 POAG cases were further reduced to 87 participants eligible for inclusion. Out of these, we got through to 20 participants residing in Yaounde, 16 of whom were included and 4 not included (denied to participate). Ten more from daily consultations were enrolled. A total of 26 cases completed the study and 19 controls were matched for age and gender. Dried blood spots on whatman filter paper grade 3 were used for DNA extraction by the Chelex method. Polymerase chain reaction (PCR) was conducted with two sets of primers on rs1971050, rs10038177 and rs10038058 followed by restriction fragment length polymorphism (RFLP) using restriction enzymes *AluI* and *ApoI* to determine the genotypes. Digestion of rs10038177 revealed in the POAG group C/T (7.7%) and T/T (92.3%) and in the non POAG group C/T (10.5%) and T/T (89.5%) genotypes. Digestion of rs10038058 revealed G/A (11.5%) and G/G (88.5%) genotypes in the POAG group and in the non-POAG group G/A (10.5%) and G/G (89.5%) genotypes. Heterozygote and homozygote mutant genotypes were thus expressed in both groups for rs10038177 and rs10038058. Only the homozygous mutant form of rs1971050 was evidenced in both groups of our study population T/T (100%). There was a non-significant clustering of the heterozygous genotype of rs10038177 C/T ( $p = 0.5$ ) with the advanced glaucoma disease process. Correlation tests showed no significant association of single nucleotide polymorphisms (SNPs) rs1971050, rs10038177 or rs10038058 with POAG ( $OR = <1.0$ ,  $p = 1.0$ ). In conclusion, advanced glaucoma clustered with the heterozygote mutant form of rs10038177 and may request greater sample size and other dependent associated genes to be investigated. Homozygous mutant genotypes of rs1971050, rs10038177 and rs10038058 may not be associated with the disease process of primary open angle glaucoma. **Key words:** POAG, *WDR36*, SNP, *AluI*, *ApoI*.

## PgmNr 3169: Role of archaic APOBEC3 genes in population-specific virus-host interactions.

### Authors:

D. Ebrahimi <sup>1</sup>; C. Richards <sup>1,2</sup>; M. Carpenter <sup>1,3</sup>; R. Harris <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN; 2) Department of Microbiology and Immunology, Stanford University, Stanford, CA; 3) Howard Hughes Medical Institute, University of Minnesota, Minneapolis, MN

---

Human chromosome 22 codes for seven APOBEC3 (A3) enzymes, four of which, A3D, A3F, A3G, and A3H have antiviral activities. These enzymes can inhibit retroviruses such as HIV-1 by hypermutating cDNA cytosine nucleobases or by other deaminase-independent mechanisms. Our multi-omics studies suggest that A3G and A3H are the main sources of viral hypermutation in non-African and sub-Saharan African populations, respectively. We found that the dominant A3H allele in sub-Saharan Africa has a significant sequence similarity to the Neanderthal/Denisovan A3H gene. In contrast, the dominant A3G allele outside Africa shares significant sequence similarity with the Neanderthal/Denisovan A3G gene. Taken together, these data suggest that both modern sub-Saharan Africans and non-Africans have inherited their ability to hypermutate retroviruses from archaic humans. It is possible that exposure to ancient retroviruses were positively (and independently) selected for these two genes in different human populations.

## PgmNr 3170: Mouse chitinases have no synergistic effects on degradation of chitin.

### Authors:

M. Kimura <sup>1,2</sup>; T. Umeyama <sup>1</sup>; S. Wakita <sup>1</sup>; K. Okawa <sup>1</sup>; M. Masayoshi <sup>1</sup>; F. Oyama <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Chemistry and Life Science, Kogakuin Univ, Hachioji, Tokyo, Tokyo, Japan; 2) Research Fellow of Japan Society for the Promotion of Science (DC2), Koujimachi, Chiyoda-ku, Tokyo, 102-0083 Japan

---

Humans and mouse express two active chitinases, chitotriosidase (Chit1) and acidic mammalian chitinase (AMCase). Chit1 and AMCase have been implicated in various pathophysiological conditions such as chronic inflammatory diseases. Chit1 is increased in individuals with Gaucher disease, atherosclerosis disease and Alzheimer's disease. AMCase is increased in individuals with mouse models of asthma. However, the contribution of mammalian chitinases to the pathogenesis of these diseases and the physiological condition remains to be determined. *Serratia marcescens* express high levels of chitinases, Chitinase A (*Serratia* ChiA), Chitinase B (*Serratia* ChiB) and Chitinase C (*Serratia* ChiC). *Serratia* ChiB is a well-studied bacterial chitinase and structurally similar to the mammalian chitinases. It has been reported that *Serratia* ChiA and ChiB have synergistic effect on the chitin degradation. Mouse Chit1 and AMCase are expressed in monocytes and macrophages. Although Chit1 and AMCase co-exist in various tissues and cells, where the potential synergic functioning of these enzymes have not been determined. We determined combinatory effects of Chit1 and AMCase in natural chitin substrates using by fluorophore-assisted carbohydrate electrophoresis (FACE) method. Our results show that Chit1 and AMCase have no synergistic effect under physiological conditions. Thus, Chit1 and AMCase function independently in any tissue regardless of the pH or the substrates, in contrast to observations regarding some bacterial chitinases.



## PgmNr 3171: Comparision of functional properties between human and mouse chitotriosidase enzymes.

### Authors:

T. Watanabe <sup>1</sup>; M. Kimura <sup>1,2</sup>; A. Ikejiri <sup>1</sup>; K. Sekine <sup>1</sup>; F. Oyama <sup>1</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Department of Chemistry and Life Science, Kogakuin University, Tokyo, Japan.; 2) Research Fellow of Japan Society for the Promotion of Science (DC2), Koujimachi, Chiyoda-ku, Tokyo, 102-0083 Japan

---

Chitin is a polymer of  $\beta$ -1, 4-linked *N*-acetyl-D-glucosamine (GlcNAc), and is an integral component of the exoskeletons of crustaceans and insects, the microfilarial sheaths of parasites and the fungal cell walls. Although mammals do not produce chitin, two active chitinases, chitotriosidase (Chit1) and acidic mammalian chitinase, have been identified in mouse and human. Chit1 has attracted considerable attention due to its increased expression in individuals with different pathological conditions such as Gaucher disease, chronic obstructive pulmonary disease (COPD), Alzheimer's disease, atherothrombosis, diabetes mellitus and cystic fibrosis. Thus, Chit1 may play important roles in many pathophysiological conditions and can be used for biomarkers monitoring the disease state. In this study, we directly compared the enzymatic properties between human and mouse Chit1 proteins using artificial and natural chitin substrates by combination of the colorimetric analysis and fluorophore-assisted carbohydrate electrophoresis (FACE) method. The chitinolytic activity of human Chit1 is higher than that of mouse Chit1. Furthermore, human Chit1 exhibited much higher transglycosylation activity over mouse counterpart. These results suggest the functional differences of human and mouse Chit1 proteins *in vivo*.

## PgmNr 3172: Production of chitooligosaccharides from chitin and chitosan by mouse acidic mammalian chitinase.

### Authors:

C. Takasaki; S. Wakita; F. Oyama

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Chemistry and Life Science, Kogakuin University, 2665-1 Nakano, Hachioji, Tokyo, Japan

---

Chitin is a polysaccharide composed of  $\beta$ -1,4-glycosidic linked *N*-acetyl-D-glucosamine (GlcNAc). It functions as the major structural component of exoskeleton of crustaceans and insects, microfilarial shells of nematodes, and cell walls of fungi. Chitin is partially deacetylated by chemically or enzymatically and becomes chitosan. Although humans and mice do not synthesize endogenous chitin and chitosan, they produce two active chitinases, chitotriosidase and acidic mammalian chitinase (AMCase). AMCase has attracted considerable attention because AMCase levels are altered in various diseases such as asthma, allergic inflammation, gastric cancer, ocular allergy and dry eye syndrome. In addition, polymorphisms of AMCase have been shown to be associated with bronchial asthma in humans. Mouse AMCase mRNA is predominantly expressed in the stomach. Recombinant AMCase and its catalytic domain had the highest activity at pH of around 2.0, when it produces primarily dimer of GlcNAc [(GlcNAc)<sub>2</sub>] with lower activities at more neutral pH (pH 3.0~7.0). Several recent reports have indicated that chitooligosaccharides derived from both chitin and chitosan possess biomedical activities including anti-tumorigenic and anti-inflammatory effects. It still remains elusive whether AMCase can produce chitooligosaccharides. Here, we report that AMCase produce chitooligosaccharides from chitin and chitosan. AMCase produced more (GlcNAc)<sub>2</sub> from  $\beta$ -chitin than  $\alpha$ -chitin. It is been known that  $\alpha$ -chitin and  $\beta$ -chitin differ in their degree of deacetylation (DD). Thus, we next examined the degradation products of chitosan with 69%, 73%, 84% and 95% DD by AMCase. AMCase degrade chitosan substrates up to 84% DD. In response to increase of DD, the amounts of products were decreased, and 94% DD chitosan remained undetected. These results indicate that AMCase can degrade chitin and chitosan and produce chitooligosaccharides.

## PgmNr 3173: Expression and characterization of chitinases in crab-eating monkey, a non-human primate.

**Authors:**

M. Uehara; E. Tabata; F. Oyama

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Chemistry and Life Science, Kogakuin Univ, Hachioji, Tokyo, Japan

---

Crab-eating monkey is one of the important nonhuman primate animal models in basic and applied biomedical researches. Chitin is a polymer of *N*-acetyl-D-glucosamine. Although mammals do not produce endogenous chitin, they express two active mammalian chitinases, acidic chitinase (CHIA) and chitotriosidase (CHIT1). Because the levels of expression of these chitinases are increased under many pathological conditions such as Gaucher disease and mouse models of asthma, both chitinases may play important roles in the pathogenesis. Moreover, CHIA can function as a protease-resistant major glycosidase under the gastrointestinal conditions. Here, we investigated the expression of chitinases in crab-eating monkey. Expression level of CHIA mRNA was high in monkey stomach. CHIT1 mRNA was highest in the lung among other monkey tissues. When the chitinolytic activity was measured using monkey stomach and lung extracts, they showed strong chitinolytic activities. Then, we measured the chitinolytic activity of *E. coli*-expressed monkey CHIA and CHIT1 enzymes to clarify their respective enzymatic function. The highest activity of both chitinases was detected at pH 5.0. In addition, we also detected robust chitinolytic activity of CHIA under strong acidic condition to neutral condition. These results indicate that crab-eating monkey CHIA possess distinctive properties for biomedical purposes in various organ.

## PgmNr 3174: Studying functional *de novo* mutations linked to autism spectrum disorders in *Drosophila* using an overexpression screen.

### Authors:

S. Deal<sup>1,2</sup>; P.C. Marcogliese<sup>2,3</sup>; J.M. Harnish<sup>2,3</sup>; J. Andrews<sup>2,3</sup>; H.K. Graves<sup>1,2</sup>; Y.-H. Chao<sup>2,3</sup>; S. Jangam<sup>2,3</sup>; V.H. Bhavana<sup>3</sup>; X. Luo<sup>2,3</sup>; N. Liu<sup>2,3</sup>; H. Chung<sup>2,3</sup>; C. Brady<sup>1,2</sup>; H. Pan<sup>2</sup>; H.-T. Chao<sup>2,3,4,5</sup>; P.-T. Lee<sup>2,3</sup>; K.L. Scott<sup>3</sup>; M.F. Wangler<sup>1,2,3,4,5</sup>; S. Yamamoto<sup>1,2,3,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Developmental Biology, Baylor College of Medicine (BCM), Houston, TX, 77030; 2) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX, 77030; 3) Department of Molecular and Human Genetics, BCM, Houston, TX, 77030; 4) Department of Pediatrics, BCM, Houston, TX, 77030; 5) Department of Neuroscience, BCM, Houston, TX, 77030

---

Autism Spectrum Disorders (ASD) encompasses a complex spectrum of neurodevelopmental disorders. Sequencing efforts reveal a highly heterogeneous genetic component to ASD. In our lab, we work to uncover molecular pathways involved in neurological diseases by testing functionality of human disease associated variants in *Drosophila*. Currently, we are testing *de novo* ASD-associated missense mutations identified in the Simons Simplex Collection, a cohort of 2500 simplex families where one child has non-syndromic ASD. From the collection, we selected 300 variants in 237 genes for functional studies in *Drosophila*. We generated loss-of-function (LOF) gene trap alleles that express GAL4 in the pattern of the fly gene of interest (T2A-GAL4 lines) in order to determine if the variant has functional consequences *in vivo*. With the 126 MiMIC T2A-GAL4 lines and 214 UAS-human cDNAs we have made thus far, we have been performing high-throughput phenotypic analysis and rescue experiments (see poster by Andrews et al). In parallel, we are over-expressing reference or variant human cDNAs to assess phenotypic consequences. This is a quick method for analyzing functionality of the variant to identify if the ASD variant behaves as a LOF (reference causes phenotype but variant does not) or a gain-of-function (GOF, variant causes phenotype but reference does not) allele. We are expressing the reference and variant cDNA using three different drivers (ubiquitous, wing-specific, and eye-specific) and scoring for lethality or phenotypes such as reduced wing size, abnormal wing veins, or rough-eyes. Thus far, we have tested 56 reference and 60 variant cDNAs using this approach, and we have observed a LOF effect in 26.7% (16 variants) and a GOF effect in 1.7% (1 variant) of the variants tested. By using this overexpression approach to understand variant function, we have identified several interesting candidates for further study including *GRK4*, *MINK1*, *PC*, and *SOGA3*. We have also identified a variant in *TSC2*, a known ASD gene, to be a LOF allele, which supports our method to screen for ASD candidate genes. We are currently dissecting our hits for their role in neural development using additional human genetic data sets as well as *Drosophila* genetic toolkits. This screen allowed us to quickly probe the function of hundreds of uncharacterized variants in ASD candidate genes and provides a framework to study the function of these genes in neural development.

## PgmNr 3175: Single-cell RNA sequencing analysis identifies knockout-specific and conserved immune cell subsets in *Clec16a* knockout mice.

### Authors:

M. Bakay <sup>1</sup>; M. Gonzalez <sup>1</sup>; B. Strenkowski <sup>1</sup>; F. Mafra <sup>1</sup>; R. Pandey <sup>1</sup>; H. Hakonarson <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

*CLEC16A* is genetically linked with many inflammatory and autoimmune diseases. Its implication in these diverse diseases suggests that *CLEC16A* could be a critical regulator of aberrant autoimmune responses, however the exact mechanism of its action is still largely unknown. To investigate the role of *CLEC16A* in autoimmunity we generated an inducible ubiquitous *Clec16a* knockout (KO) mouse. The global *Clec16a* KO mice exhibit a complex phenotype: immune dysfunction, weight loss, and neurodegeneration. Recently we demonstrated that loss of *CLEC16A* in immune cells leads to an abnormal mitophagy, cell death and immune dysfunction.

Here, we performed single-cell RNA sequencing of immune cells isolated from peripheral blood of nine mice combined into three sample sets (*Clec16a* KO control-vehicle and control-tamoxifen), using the standard 10x Genomics method. The Seurat package in R was used to analyze data and visualize results. After QC, 23,455 cells were characterized from three separate conditions (11,243 from *Clec16a* KO, 5,849 from control-vehicle, 6,363 from control-tamoxifen). Ingenuity Pathway Analysis was used for further assessment of gene function. Automated clustering generated 19 distinct cell clusters. Cells were identified into one of the known cell populations based on canonical gene markers in each cluster. The transcriptional profiles of these individual cells enabled us to identify 19 clusters. Based on canonical gene markers in each cluster we identified 13 of them, with 6 still pending for characterization. For two the most prevalent subtypes (T, B) and also NK-cells, we observed clear separation between cells derived from KO and controls. The specific subset of CD8<sup>+</sup> T cells was preferentially enriched in blood of *Clec16a* KO versus vehicle and tamoxifen treated control (9.4%, 6.4 and 6.5%, respectively), the only subset with very limited number of differentially expressed genes (DEGs), total of 38. Next, we identified signature genes for each cluster and DEGs between KO and control for each known cell subtype. Only a small portion of DEG (~10%) was down-regulated and included genes involved in iron homeostasis signaling. The most significant up-regulated genes were involved in oxidative phosphorylation and cellular signaling. Mitochondrial dysfunction was in a top of tox list, followed by NRF2-mediated oxidative stress response. These result provide additional support for the direct involvement of *CLEC16A* in regulation of mitochondrial function.

## PgmNr 3176: Tissue specific compensatory regulation of gene expression associated with copy number variants in *Danio rerio*.

### Authors:

S. Brown <sup>1</sup>; J.Y. Chen <sup>2</sup>; K.H. Brown <sup>3</sup>; X. Shi <sup>4</sup>; C. Lee <sup>5,6,7</sup>; K.P. Dobrinski <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Tampa, Department of Biology, 401 W. Kennedy Blvd, Tampa, FL 33606; 2) Massachusetts General Hospital, Department of Neurology, 55 Fruit Street, Boston, MA 02114; 3) Portland State University, Department of Biology, 1825 SW Broadway, Portland, OR 97201; 4) University of North Carolina at Charlotte, Department of Bioinformatics and Genomics, 9201 University City Blvd, Charlotte, NC 28223; 5) The Jackson Laboratory for Genomic Medicine, 10 Discovery Drive, Farmington, CT USA 06032; 6) Ewha Womans University, 52 Ewhayeodae-gil, Bukahyeon-dong, Seodaemun-gu, Seoul, South Korea; 7) Xi'an Jiaotong University, 28 Xianning W Rd, Jiao Da Shang Ye Jie Qu, Beilin Qu, Xian Shi, Shaanxi Sheng, China, 718900

---

Copy number variants (CNVs) refer to the loss or gain of copies of a DNA region within the genome. While some CNVs enrich the diversity of an organism and play a role in species evolution, CNVs may also be linked to certain diseases such as neurological disorders, early onset obesity, and cancer. CNVs can affect gene expression sometimes leading to disease by direct overlap of the gene or by indirect effects (outside the gene location). These indirect associations may be the result of CNVs containing regulatory elements such as miRNAs. *Danio rerio* (zebrafish) is an excellent model organism for human development as well as disease. Zebrafish share 70% of genetic similarities with humans with 84% of genes associated with human disease also found in zebrafish. Additionally, zebrafish genomes contain a large amount of variation with 14.6% of the zebrafish reference genome being variable. This level of variation is more than four times the percentage of reference genome sequence covered by similarly common CNVs in humans and makes this organism an interesting model to investigate the effects of CNV on gene expression as well as how cells may compensate for these effects. This study seeks to compare CNVs, miRNA expression levels and gene expression levels within the kidney and liver of zebrafish. Array comparative genomic hybridization was used to determine CNV regions in 30 adult Tubingen zebrafish while expression arrays were used to measure gene expression as well as miRNA expression in kidney and liver in these same 30 fish. Expressive quantitative trait loci (eQTL) analysis (linear regression model in R) was used to explore how CNVs may affect gene expression and the regulatory role played by miRNAs. eQTL analysis was performed for cis associations with a 1-Mb (megabase) window upstream and downstream from the transcription probe midpoint to CGH midpoint. This analysis found less than 100 associations in the kidney and over 800 associations in the liver. Around 40% of associations in both tissues were negative (increased copy number – decreased expression, decreased copy number – increased expression) and most associations were indirect where the gene was not overlapped by CNV. Ongoing analyses with these associations will investigate gene functionality including immune function and potential disease susceptibility impacted by CNVs and the compensatory regulation carried out by the cell through mechanisms such as miRNAs.

## PgmNr 3177: Combing ACMG criteria with deep learning algorithms to classify VOUS mutations in genetic sequencing for rare diseases.

### Authors:

Y. Liu <sup>1</sup>; H. Qu <sup>1</sup>; A. Wenocur <sup>1</sup>; L. Tian <sup>1</sup>; H. Hakonarson <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Applied Genomics CAG, The children's hospital of Philadelphia CHOP, Philadelphia, PA; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Division of Human Genetics, Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

---

Contradictory evidence in functional classification of genetic variation based on the American College of Medical Genetics and Genomics (ACMG) guidelines has been a frequently seen issue. Based on analysis of 140,883 genetic variation in the Human Gene Mutation Database (HGMD), benign classification BP1 (2.99%) and BS2 (0.74%), were identified among variants met criteria to be classified as likely pathogenic/pathogenic. PM1 is commonly observed among variants classified as benign (28.45%), while PM2 and PM1 are commonly identified among variants met criteria to be classified as likely benign (48.91% and 42.95%, respectively). In most cases, these variants were re-assigned as VOUS (variants of uncertain significance) and medical genetic geneticist must review the results with patient medical history cautiously to reach a conclusion. Deep learning architectures such as deep neural networks (DNN), recurrent neural networks (RNN) and convolutional neural networks (CNN) have been applied to multiple biomedical fields, including drug discovery/toxicology tests, medical imaging analysis and bioinformatics. We applied deep learning model for the most common type of monogenic diabetes, maturity onset diabetes of the young type 3 (MODY 3), with HGMD reported variants in HNF1A gene as cases and likely benign/benign variants as controls in training data, the ACMG criteria service as the feature vectors to predict the confident score of VOUS variants. In this study, we demonstrated that deep learning algorithm could be applied to weight ACMG criteria then further differential VOUS variants for well-known genetic diseases.

## PgmNr 3178: Whole genome sequence analysis of *Mycobacterium tuberculosis* from Morocco.

### Authors:

L. Meriem <sup>1</sup>; E. Nadia <sup>2</sup>; B. Fatna <sup>2</sup>; L. Abdelhay <sup>2</sup>; B. Yassine <sup>2</sup>; Z. Adil <sup>2</sup>; E. Rachid <sup>2</sup>; F. Mohammed <sup>2</sup>; S. Samir <sup>3</sup>; E. Mostafa <sup>2</sup>; I. Azeddine <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Faculty of medicine and pharmacy rabat, Mohammed V university, Rabat, Morocco; 2) Laboratoire de Microbiologie de l'Hôpital Militaire d'Instruction Mohamed V Rabat/Maroc ?Rabat, Morocco.; 3) Service des Brulées de l'Hôpital Militaire d'Instruction Mohamed V Rabat/Maroc ?Rabat, Morocco

---

Drug-resistant tuberculosis (TB) remains a major challenge to global health and to healthcare in Morocco. In 2017, a total of 30897 cases of TB were recorded in Morocco, of which 1.8 % were multidrug-resistant TB (MDR-TB). Extensively drug-resistant TB (XDR-TB) occurs at a much lower rate, but the impact on the patient and hospital is severe. Current diagnostic methods such as drug susceptibility testing and targeted molecular tests are slow to return or examine only a limited number of target regions, respectively. Whole genome sequencing (WGS) has been shown to provide a rapid and comprehensive view of the genotype of the organism, and thus enable reliable prediction of the drug susceptibility phenotype within a clinically relevant timeframe. In addition, it provides the highest resolution when investigating transmission events.

To characterize the genetic determinants of resistance to antituberculosis drugs, we performed a whole genome analysis study on clinical isolates of *Mycobacterium tuberculosis* from different part of Morocco. A mutation rate analysis was followed by a spoligotyping determination for a phylogenetics SNP basec analysis. In addition to mutations in established and recently described resistance-associated genes, novel mutations were discovered for resistance to fluoroquinolones, ethambutol with lower concentration. This study shows that the majority of MDR and pre XDR strains belong to the Euro-American LAM9 sublineage while all the susptible strains belong to the haarlem H4. In addition all strains shared more than 360 SNP's, however uniq SNP's among MDR-TB and susptible strains suggested a higher genomic diversity among the LAM9.

Our findings provide an improved understanding of MDR-TB in Morocco, potentially significant in achieving the goal of precision medicine with respect to MDR-TB prevention and treatment.



## PgmNr 3179: Identification and functional characterization of a novel *ENAM* mutation in the signal peptide region in an autosomal dominant hypoplastic amelogenesis imperfecta family.

### Authors:

J.-W. Kim <sup>1,2</sup>; J. Kang <sup>2</sup>; Y.J. Kim <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular Genetics, Seoul National University School of Dentistry, Seoul, Korea, Republic of; 2) Pediatric Dentistry, Seoul National University School of Dentistry, Seoul, Korea, Republic of

---

Amelogenesis Imperfecta (AI) is a rare hereditary condition that results in developmental enamel defects, affecting the quality and/or quantity of enamel. The affected suffer from early tooth loss, eating difficulties and dental pain. Formation of enamel (amelogenesis) starts from ameloblasts secreting proteins that gradually thicken extracellular layer. ENAM is one of key matrix proteins in amelogenesis. We recruited a three generation family with a unique form of rough hypoplastic AI. The enamel shows irregular local hypoplastic area in all teeth. The affected enamel was not only thin but also hypomineralized based on reduced contrast with dentin. Pedigree analysis suggested autosomal dominant inheritance. Whole exome sequencing of the proband revealed a novel *ENAM* missense mutation (c.101T>C). This mutation would change leucine to proline at 34 codon position in the signal peptide region (39 amino acids). Sanger sequencing confirmed perfect segregation of this mutation with affected status in this family. Short N-terminal segment (42 amino acids) was amplified and cloned into pEGFP-N1 vector. PCR mutagenesis was performed to introduce the identified mutation. HEK293T cells were transfected with wild type and mutant vectors. Cell lysate and culture media were collected and western blot was performed to compare protein level of wild type and mutant vectors. It confirmed that secretion of mutant protein is greatly impaired. Confocal microscopy showed co-localization of ER marker with mutant protein, indicating ER retention. These results will not only expand mutational spectrum of the *ENAM* but also improve our understanding of molecular pathogenesis in AI.

**Keywords:** amelogenesis imperfecta; hypoplastic; whole exome sequencing; ENAM; confocal microscopy

## PgmNr 3180: Contribution of host genetics to pandemic influenza H1N1/09 vaccine response.

### Authors:

B.Y.H. Bai<sup>1,2</sup>; E. Theodoridis<sup>3</sup>; S. O'Farrell<sup>3,4,5</sup>; A. Lorenc<sup>3</sup>; V.N. Petrova<sup>1,2</sup>; L. Moutsianas<sup>6,7</sup>; A.C. Hayday<sup>3,4,5</sup>; C.A. Anderson<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Sanger Institute, Hinxton, United Kingdom; 2) University of Cambridge, Cambridge, United Kingdom; 3) Peter Gorer Department of Immunobiology, King's College London, London, United Kingdom; 4) ImmunoSurveillance Laboratory, Francis Crick Institute, London, United Kingdom; 5) Biomedical Research Centre, Guys & St Thomas NHS Foundation Trust & King's College London, London, United Kingdom; 6) Bioinformatics Research Services, Genomics England Limited, London, United Kingdom; 7) William Harvey Research Institute, Queen Mary University of London, London, United Kingdom

---

**Background:** Vaccines have enormous impact on global health, but their protective efficacy varies between individuals, and we have limited understanding of the immune mechanisms responsible. Systems vaccinology can be applied to model the variation in immune response to vaccines, but few studies thus far have examined the influence of host genetics. **Methods:** We consider in vivo response to pandemic influenza H1N1/09 vaccine (Pandemrix), in a previously-recruited cohort of 178 UK adults (Sobolev et al. 2016, DOI: 10.1038/nr.3328). We integrate matched genetic, transcriptomic (pre-vaccination, day 1 post-vaccination, day 7 post-vaccination), and vaccine-induced antibody response data (pre-vaccination, day 63 post-vaccination). **Results:** 1481 genes were differentially expressed post- vs. pre-vaccination (FDR<0.05, abs. log<sub>2</sub>FC>1). Coexpression gene modules related to innate immunity were upregulated at day 1 (e.g. monocyte and dendritic cell modules). Modules related to adaptive immunity were upregulated at day 7 (e.g. cell cycle and plasma cell modules). At 4630 genes, we find cis-eQTLs that influence post-vaccination gene expression (LFSR<0.05), of which 1894 are vaccine-response eQTLs, where the eQTL effect size differs pre- vs. post-vaccination. A subset of these vaccine-response eQTLs affect differentially expressed genes, and genes with expression that correlates with antibody response. For example, *ELL2* (MIM 601874), a gene involved in B cell antibody production, has upregulated expression at day 7 (FDR=8.2e-10, log<sub>2</sub>FC=2.2) that correlates with antibody response (p=0.04, r=0.31), and is modulated by a day 7-specific eQTL (LFSR=0.03). **Discussion:** By integrating longitudinal transcriptomic and antibody response data with matched genetic data, we identify loci where host genetics impacts immune response to Pandemrix. We are now performing colocalization of our vaccine-response eQTLs with independent QTL studies (e.g. in vitro stimulated immune cell subsets) and associations from GWAS of immune-mediated diseases. This will allow us to pinpoint cellular contexts and shared mechanisms through which these genetic effects on immune response manifest—we expect to present these additional results at the meeting.

# **PgmNr 3181: Gene-level, allele-specific and total read count-based mixed effects model improves power and speed for cis-QTL mapping, fine-mapping, and prediction.**

**Authors:**

Y. Liang <sup>1</sup>; S.E. Stephane <sup>2,3</sup>; F. Aguet <sup>4</sup>; K. Ardlie <sup>4</sup>; T. Lappalainen <sup>2,3</sup>; H.K. Im <sup>1</sup>; GTEx Consortium

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Department of Medicine, University of Chicago, Chicago, IL.; 2) Department of Systems Biology, Columbia University, New York, NY; 3) New York Genome Center, New York, NY; 4) Broad Institute of Harvard and MIT, Cambridge, MA

---

An increasing number of studies measure RNA-seq and other molecular traits with allele-specific (AS) reads. However, most methods do not take full advantage of the information provided by the combination of total and AS reads. Although there have been efforts to unify total and AS reads, they tend to be computationally intensive and are typically not used in large-scale studies such as GTEx with several hundred samples. In addition to QTL mapping, multi-SNP models for building prediction models and fine-mapping have become increasingly important for post-GWAS analyses approaches such as PrediXcan and COLOC. However, to our knowledge, there are no multi-SNP methods that take advantage of AS observations.

The gain in speed and power is achieved by using a linear mixed effects model combining total and gene-level AS read counts. Key features of our approach are 1) the use of a random effect to directly model the common environmental conditions to which both copies of the person's haplotypes are subjected; 2) a second random effect to account for the noise around the expected fraction of AS reads given the individual's heterozygosity; 3) the borrowing of information across the gene body using gene-level AS counts that are calculated using phASER.

We implement the single-SNP version of the method for QTL-mapping. To fit the multi-SNP model for prediction, we develop a penalized likelihood-based algorithm. Furthermore, to make our framework compatible with existing fine-mapping tools, which assume independent observations, we perform a whitening procedure on the three observations of each individual. Finally, we apply the methods to GTEx RNA-seq data to study the cis-regulation of gene expression.

We applied our methods to RNA-seq of 670 whole blood samples in GTEx version 8 data. Our eQTL mapping method (mixQTL) was more than 10 times faster and more powerful than existing methods such as WASP and RASQUAL. Compared to GTEx standard eQTL calling pipeline results, mixQTL was more powerful while maintaining the type I error calibration. Furthermore, our multi-SNP method showed better performance in identifying causal variants both in simulations and real data.

We present here a suite of fast and powerful methods that improve cis-eQTL mapping, fine-mapping, and prediction by taking full advantage of total and AS counts available in many large-scale studies. We demonstrate the type I error calibration and superior power using both real and simulated data.

## PgmNr 3182: Noncoding deletions expose a gene that is critical for intestinal function.

### Authors:

D. Oz Levi <sup>1,2</sup>; T. Olender <sup>2</sup>; I. Bar-Joseph <sup>3,4</sup>; Y. Zhu <sup>5</sup>; D. Marek-Yagel <sup>3,4</sup>; I. Barozzi <sup>5</sup>; M. Osterwalder <sup>5</sup>; A. Alkelai <sup>6</sup>; E.K. Ruzzo <sup>7</sup>; Y. Han <sup>8</sup>; H. Reznik-Wolf <sup>3,4</sup>; C. Hartman <sup>4,9</sup>; R. Shamir <sup>4,9</sup>; B. Weiss <sup>10</sup>; R. Shapiro <sup>4,9</sup>; B. Pode-Shakked <sup>10</sup>; D.M. Imai <sup>11</sup>; D. Coleman-Derr <sup>12</sup>; B. Black <sup>13</sup>; J.M. Wells <sup>14,16</sup>; R. Kleta <sup>15</sup>; D.B. Goldstein <sup>6</sup>; E. Pras <sup>3,4</sup>; A. Visel <sup>5</sup>; D. Lancet <sup>2</sup>; Y. Anikster <sup>4,10</sup>; L.A. Pennacchio <sup>5,17</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) DNA Science Team, MyHeritage, Or Yehuda, Israel; 2) Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel; 3) The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Ramat Gan, Israel; 4) The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel; 5) Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA; 6) Institute for Genomic Medicine, Columbia University Medical Center, New York, NY, USA; 7) Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles; 8) Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708, USA; 9) Schneider Children's Medical Center, Petach Tikva, Israel; 10) Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Ramat Gan, Israel; 11) Comparative Pathology Laboratory University of California 1000 Old Davis Rd, Building R1 Davis, CA 95616, USA; 12) Plant Gene Expression Center, USDA-ARS, Albany, CA, 94710, USA; 13) Cardiovascular Research Institute, University of California, San Francisco, CA 94143-3120, USA; 14) Division of Developmental Biology at Cincinnati Children's Hospital Medical Center, Cincinnati OH 45229; 15) Centre for Nephrology, University College London, London, UK; 16) Division of Endocrinology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, 45229, USA; 17) Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA

---

Large-scale genome sequencing is poised to provide an exponential increase in the discovery of disease-associated mutations, but the functional interpretation of such mutations remains a major challenge. Here we identify deletions of a regulatory sequence (intestine-critical region, ICR) on human chromosome 16 that cause inherited severe and intractable congenital diarrhea in affected infants from seven families of Jewish Iraqi origin. Transgenic mouse reporter assays show that the ICR robustly activates tissue-specific transcription *in vivo* during development of the gastrointestinal system. Targeted deletion of the ICR in mice caused symptoms recapitulating all major aspects of the human condition. Guided by the gastrointestinal *in vivo* activity pattern of the ICR, we performed transcriptome analysis in the developing gut of mouse embryos, which identified a neighboring unannotated open reading frame (*Percc1*) whose expression was lost in ICR knockout animals. Targeted deletion of *Percc1* in mice recapitulated the phenotypes observed in ICR deletion mice and patients, whereas an ICR-driven *Percc1* transgene rescued all phenotypes found in ICR knockout mice. Taken together, our results uncover a novel human gene critical for intestinal function and underscore the need for targeted *in vivo* studies for interpreting the growing number of clinical genetic findings that do not affect known protein-coding genes.

## PgmNr 3183: Evaluating variants of unknown significance with CRISPR-engineered human-yeast ortholog complementation.

### Authors:

H. Ji <sup>1</sup>; A. Chu <sup>2</sup>; A. Hooker <sup>1</sup>; B. Leland <sup>3</sup>; K. Yerramsetty <sup>3</sup>; A. Edwards <sup>2</sup>; B. St. Ongoe <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Medicine/Oncology, Stanford Univ, Stanford, California.; 2) Stanford Genome Technology Center, Palo Alto, California; 3) Inscripta, Boulder, CO

---

Most catalogued human genetic variants lack any functional characterization. These variants of unknown significance pose a dilemma for the diagnosis of human diseases whether they are hereditary disorders arising from germline mutations or cancer with somatic alterations. Moreover, without any substantive functional characterization, it remains difficult to determine the phenotypic consequences of most human genetic variation, be it polymorphism, rare variant or mutation. There are a limited number of high throughput methods for functional characterization of human variants. To address this challenge, we are using a novel CRISPR yeast genome engineering technology to study variants of unknown significance – our study focuses on the functional complementation of specific human yeast gene orthologs. Notably, some human genes provide complementary function to their yeast ortholog which has been deleted or reduced in expression. We are leveraging this functional ortholog complementation for parallel evaluation of human variants in human genes. Specifically, we are focused on complementing yeast genes that are essential or non-essential. For this study, we identified four essential yeast genes with their human orthologs provide complementation of lethal growth defects. These human genes have germline mutations that lead to a variety of genetic disorders. First, human gene orthologs are engineered for approximately 300 coding variants. To knockdown endogenous yeast genes, we rely on a yeast strain that has stable integration of gRNAs specific to over 1,117 essential genes and 514 genes required for robust growth. Among this set of human genes include NSHDL, a gene where germline mutations cause congenital hemidysplasia with ichthyosiform erythroderma and limb defects (CHILD). To identify germline variants, we used resources such as gnomAD and ClinVar. Using structural prediction information or CADD scores, we ranked up to 300 variants that are introduced into the human gene, eliminate the yeast ortholog expression and then these strains are tested for perturbations in growth. Subsequently, we determine which human variants affected growth. Our approach holds the possibility of enabling rapid, efficient and highly parallel functional characterization of human genetic variants.

## PgmNr 3184: A mix pool of iPSCs from hundreds of genetically different donors for multiplexing genetic studies.

### Authors:

Y. Chan <sup>1,2</sup>; E.T. Lim <sup>1,2</sup>; G.M. Church <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, Harvard Medical School, Boston, Massachusetts.; 2) Synthetic Biology, Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts.

---

Genome wide association studies (GWAS) have been used to discover variants associated with a plethora of traits and diseases. Fine mapping and validation of the variant is not straightforward as manipulation of the genome of a human individual in-vivo is not feasible. However, with the invention of induced pluripotent stem cells (iPSCs), we can reprogram human adult cells into iPSCs and differentiated them to a plethora of adult tissue types for testing in-vitro. Therefore, we can create iPSCs from carriers as well as non-carriers, differentiate and test whether the genetic effect discovered in GWAS can be replicated in-vitro. To enable such studies, one would require a very large collection of individual iPSCs that have whole-genome sequencing data as well as a method to reprogram all of them into iPSCs in a feasible and cost effective manner. Here, we describe our approach for achieving just that. We use lymphoblastoid cell-lines (LCLs) from both the 1000 Genomes Project and Personal Genome Project where whole-genome sequencing data is available for every donor. By pooling together a large number of LCLs from many different donors (>100 donors), we reprogrammed the mixed pool of LCLs into iPSCs by nucleofection of the Yamanaka factors. We found that the LCLs do become adherent and expressed stem-cell markers. To determine to what extent the individual donor LCLs were successfully reprogrammed, i.e. donor diversity, we employed our method which uses cell sorting, DNA sequencing and an EM-algorithm to predict relative donor proportion. The method does not require single-cell or exogenous DNA barcodes but exploits SNPs as natural barcodes for individual donors. Using the method, we found that the majority of LCLs were successfully reprogrammed into iPSCs. We then proceed to determine if the iPSCs are indeed pluripotent by performing directed differentiation into the 3 different germ layers, i.e. ectoderm, mesoderm and endoderm. Using the same method, we determined the donor diversity for each of the 3 germ layers and discovered that the majority of donors iPSCs can be differentiated into all 3 germ layers. Finally, we applied our method to test for genotype specific effects of inducing pluripotency as well as exiting pluripotency from the mixed pool. In summary, we describe a method that enables the multiplex reprogramming of LCLs into iPSCs and they become a valuable resource for testing and validating genotype specific effects discovered in GWAS.

## PgmNr 3185: Vulnerability to stress: The moderating role of *HOMER1* gene in gene x environment interaction.

### Authors:

Y. Chen <sup>1</sup>; L. Wang <sup>2</sup>; Q. Zhang <sup>2</sup>; D. Jiang <sup>1</sup>; Y. Gan <sup>1</sup>; G. Chen <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Peking University, China; 2) WeGene, Shenzhen, China

---

Early life experiences and stress events play a crucial role in future development and affect future orientation as well. Future orientation is the way individuals tend to think about the future which includes planning, goal-setting and delay-discounting. The present research aims to explore whether *HOMER1*(rs7713917) polymorphism can moderate relationship between early life stress and future orientation. 14,675 Chinese Han people involved in the current study, in which were 8267 female (56.3%). Future-Oriented Coping Inventory and Adolescent Self-Rating Life Events Checklist were fulfilled by participants, and DNA was collected to test polymorphism of *HOMER1*. The results indicated that genotype of *HOMER1* rs7713917 was not associated with future orientation( $\beta = -0.01$ , 95% CI = [-0.18, 0.18],  $t = -0.04$ ,  $p = 0.969$ ), early life stress was negatively associated with future orientation ( $\beta = -1.42$ , 95% CI = [-1.60, -1.25],  $t = -15.75$ ,  $p = 0.000$ ), and that *HOMER1* polymorphisms moderated this association ( $\Delta R^2 = 3.3\%$ ,  $F[5, 14619] = 99.64$ ,  $\beta = -0.18$ , 95 %CI = [-0.36, -0.01],  $t = -2.06$ ,  $p = 0.039$ ). Specifically, individuals with AA genotype and in high early life stress expressed higher future orientation compared with those with in low chronic stress. In contrast, individuals with GA/GG genotype displayed lower vulnerability to environmental changes. Then we separated samples into 3 groups according to their type of hometown to check the interaction effect ( $\Delta R^2 = 5.6\%$ ,  $F[5, 2229] = 26.41$ ,  $\beta = -0.61$ , 95 %CI = [-1.08, -0.16],  $t = -1.27$ ,  $p = 0.204$ ), and found the same mode in town group, but found no associations in city ( $\Delta R^2 = 3.2\%$ ,  $F[5, 9140] = 60.22$ ,  $\beta = -0.18$ , 95 %CI = [-0.36, 0.09],  $t = -2.06$ ,  $p = 0.039$ ) and countryside groups ( $\Delta R^2 = 2.7\%$ ,  $F[5, 3284] = 17.99$ ,  $\beta = -0.02$ , 95 %CI = [-0.38, 0.34],  $t = -0.10$ ,  $p = 0.920$ ).



## PgmNr 3186: Generation and *in vivo* characterization of human *NOTCH* transgenes to study the functional impact of disease-associated variants in *Drosophila*.

### Authors:

S.-A. Yang <sup>1</sup>; P.C. Marcogliese <sup>1</sup>; A. Phillips <sup>2</sup>; J.L. Salazar <sup>1</sup>; H.K. Graves <sup>1</sup>; S. Yamamoto <sup>1,3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.; 2) Rice University, Houston, TX; 3) Program in Developmental Biology, Baylor College of Medicine, Houston, TX; 4) Department of Neuroscience, Baylor College of Medicine, Houston, TX; 5) Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX

---

Notch signaling is a well-studied signaling pathway that is conserved throughout metazoan evolution. While the *Drosophila melanogaster* genome encodes a single Notch receptor, the human genome encodes four paralogs (*NOTCH1-4*). Variants in human *NOTCH* genes have been found to cause diverse Mendelian disorders and have been identified in various types of cancer. Some Notch related diseases in human are thought to be caused by loss-of-function alleles (amorphs or hypomorphs), whereas others are considered to be due to gain-of-function alleles (hypermorphs, antimorphs or neomorphs). Classification of *NOTCH* mutations into different functional classes has primarily been performed using mammalian cell-based signaling assays *in vitro* or knock-in mouse models *in vivo*. However, considering the number of newly identified missense variants found in human *NOTCH* genes is increasing at a rapid pace due to utilization of whole-exome sequencing in clinical settings, a higher-throughput assay system to experimentally determine the functional consequence of disease associated *NOTCH* variants *in vivo* is crucial.

To this end, we first generated a series of transgenic constructs and *Drosophila* lines that allow the expression of full length human *NOTCH1-4*. Intriguingly, these transgenes were unable to activate Notch signaling in flies, and *NOTCH1* even seemed to act in a dominant negative manner. Because human NOTCH proteins might not be compatible with fly Notch ligands, Delta (DLL1-4 in humans) or Serrate (JAG1-2 in humans), we next generated a series of transgenic lines that can function in a ligand-independent manner. Interestingly, ligand-independent NOTCH1 and NOTCH2 were able to strongly activate Notch signaling in flies, with NOTCH3 and NOTCH4 having reduced strengths. These results suggest that the intracellular domains of human NOTCH proteins have the ability to functionally interact with Mastermind (MAML1-3 in humans) and Suppressor of Hairless (RBPJ in humans) to activate fly Notch signaling *in vivo*. We are currently continuing to introduce several disease-linked NOTCH1 and NOTCH2 mutations into these constructs to assess their functions. This system will provide a platform for analyzing human *NOTCH* variants that affect the intracellular domain systemically and efficiently *in vivo*.



## PgmNr 3187: Genetic manipulation of zebrafish is a powerful tool to determine the pathogenicity of holoprosencephaly mutations.

### Authors:

S. Hong <sup>1</sup>; P. Hu <sup>1</sup>; E. Roessler <sup>1</sup>; J.H. Jang <sup>1,3</sup>; B. Carrington <sup>2</sup>; R. Sood <sup>2</sup>; M. Muenke <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland.; 2) zebrafish Core, NHGRI, NIH, Bethesda, Maryland; 3) University of Maryland, College Park, College of Computer, Mathematical, and Natural Sciences, College Park, Maryland

---

Holoprosencephaly (HPE) is the most common malformation of the forebrain in humans. It is classified anatomically by the degree of separation of the eye field and telencephalon into discrete left and right structures. Its phenotypic spectrum is broad and can range from mild craniofacial features to the extreme of severe microcephaly, cyclopia and a proboscis. Both strong autosomal dominant driver mutations, in a handful of key developmental genes, as well as modifying environmental and genetic factors are considered likely factors to explain this phenotypic variability.

Therefore, from a set of 330 HPE probands evaluated using a targeted capture panel of 150 developmental genes we selected variants in a number of known or suspected HPE genes, both published and un-published, for zebrafish assay development. Our ultimate goal is to annotate specific disease-associated individual variants and to look for any associated co-morbid genetic modifiers present in the same family.

For example, we recently employed a sensitive assay of human *FGF8* variants in zebrafish to demonstrate that the spectrum of isoforms of FGF8 produced by alternative splicing can provide key insights into the genetic susceptibility to human malformations. In addition, we described novel mutations in the FGF core structure that have both subtle and profound effects on ligand post-translational processing and biological activity. Finally, we solved a case of apparent digenic inheritance of novel variants in *SHH* and *FGF8*, two genes known to functionally co-regulate each other in the developing forebrain, as a simpler case of FGF8 diminished function. We now extend these investigations to include nearly a hundred non-synonymous variants in *SHH*. Here we show that the application of functional testing brings considerable clarity to the genotype/phenotype analysis and affirms many of the original pathogenetic assertions. Direct variant testing in zebrafish ultimately extends, and often replaces, pathogenicity determinations made bioinformatically.

## **PgmNr 3188: The Canadian Rare Diseases Models and Mechanisms (RDMM) Network: Connecting novel gene discoveries to functional characterization research in model organisms.**

### **Authors:**

P. Pavlidis <sup>1,2</sup>; K.M. Boycott <sup>3</sup>; P.M. Campeau <sup>4</sup>; S. Rogic <sup>1,2</sup>; C. Oriel <sup>5</sup>; P. Hieter <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Michael Smith Laboratories, University of British Columbia, Vancouver, BC, Canada; 2) Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada; 3) CHEO Research Institute, University of Ottawa, Ottawa, ON, Canada; 4) Department of Pediatrics, Centre de Recherche du CHU Ste-Justine, Montreal, Canada; 5) Maternal Infant Child Youth Research Network, Vancouver, BC, Canada

---

Advances in genomics has transformed our ability to identify the molecular cause of rare diseases. Yet for most candidate genes, we lack insight into their biological function, how mutations identified in patients affect them, or what therapies could be useful. Model organisms (MOs) represent powerful experimental systems to confirm the pathogenicity of gene variants, characterize the gene's biological function, and identify potential therapies. For these reasons, the Canadian RDMM Network ([www.rare-diseases-catalyst-network.ca](http://www.rare-diseases-catalyst-network.ca)) was established in 2014 to catalyze and fund connections between clinicians discovering new disease genes and researchers able to study equivalent genes and pathways in MOs.

The RDMM Network has created a rapid and direct pathway from gene discovery to functional characterization studies in MOs. The central resource of the RDMM Network is a web-based Canadian directory of MO researchers ('the Registry') built to facilitate identification of suitable collaborators for applying clinicians. As of June 2019, more than 500 researchers have registered over 6,500 genes of interest. With the aid of the computational inference built into the Registry, this translates to the coverage of close to 7,500 human genes. RDMM uses a committee process to identify and review potential clinician-MO researcher matches and approve \$25,000 CAD in catalyst funding. Since 2014, we have made 84 connections and funded 98 functional characterization studies. Besides the scientific insights into the molecular mechanisms of rare disease and possible novel therapies, these collaborations also lead to high impact papers, long-term collaborations, and external grants.

Leveraging on the success of the first four years, last year we expanded the impact and reach of the RDMM Network: we established international linkages with emerging similar networks in Europe, Australia, Japan and the United States. To facilitate further community uptake and adaptation of the concepts we established, we have made the RDMM Registry portable, customizable and linkable with other instances, and our committee processes freely available.

The Canadian RDMM Network continues to expand its utility by adopting policies and processes to support global collaborations irrespective of the software used. In so doing, we will continue to create meaningful collaborations between clinicians and MO researchers, generate new knowledge, and

advance rare disease research locally and globally.

## PgmNr 3189: Defining the missing heredity and phenotypic diversity associated with oculocutaneous albinism.

### Authors:

S.K. Loftus<sup>1</sup>; J. Choi<sup>2</sup>; K. Brown<sup>2</sup>; L. Lundh<sup>3</sup>; NISC. Comparative Sequencing Program<sup>4</sup>; W. Oetting<sup>5</sup>; D. Adams<sup>3</sup>; W.J. Pavan<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 2) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 3) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, 10 Center Drive, MSC 1851, Bethesda, MD, 20892-1851, USA; 4) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 5) Department of Experimental and Clinical Pharmacology, University of Minnesota, Minneapolis, Minnesota 55455, USA

---

Oculocutaneous Albinism (OCA) is a recessive disorder characterized by reduction of skin, hair and eye pigmentation. OCA occurs with an average world-wide incidence of ~1 in 17,000. OCA individuals are at higher risk for UV-induced skin cancers and a complex set of developmentally-mediated visual impairments, and present with diverse clinical severity. While two definitive OCA mutation alleles are predominately identified in *TYR* (OCA types 1A and 1B) or *OCA2* (OCA type 2), a small number of cases can be attributed to mutations in *TYRP1*, *SLC45A2*, *SLC24A5*, *C10orf11/LRMDA*, or a linkage region on 4q24. Most critically, 20-50% patients remain without identification of biallelic mutations after traditional exon-based screening for these six known OCA genes. This suggests potential coding-variation in novel loci, or aberrant splicing or non-coding cis-regulatory variation at known OCA genes may contribute to the missing heredity and diverse clinical presentations observed in OCA patients. To better define the molecular mechanisms underlying OCA, we are sequencing over 400 OCA proband individuals and associated family members, at coding, intronic, and cis-regulatory regions of six OCA genes and 31 genes identified by pigmentation GWAS. We are using custom capture, short-read sequencing, in which we defined the non-coding cis-regulatory regions from a catalog of DNase1-HS data (derived from primary melanocytes, immortalized melanocytes, melanoma tumors, and retinal pigment epithelium) and from H3K27Ac ChIP cis-regulatory elements we have identified under conditions of MAPK activation and inhibition in 501mel melanoma cells. In addition, we have included regions encompassing SNPs in linkage disequilibrium with SNVs identified from our melanocyte expression quantitative trait (eQTL) dataset, generated from RNA-Seq analysis of 106 primary melanocytes. This dataset identifies eQTLs for 4997 genes expressed in the melanocyte lineage. Here we find over 73,000 eQTLs for 128 genes associated with visible pigmentation-related phenotypes, including 156 and 684 eQTLs for *TYR* and *OCA2*, respectively. Our family, trio-based analysis and interrogation of coding, and non-coding sequence, will better define the molecular spectrum of OCA mutations, facilitate assessment of non-coding DNA variation function, and allow for the subsequent systematic evaluation of the molecular mechanisms underlying both missing heritability and phenotypic variation observed among OCA patients.

## PgmNr 3190: ClinVar variants in Rhesus monkey (*Macaca mulatta*).

### Authors:

M. Raveendran; A. Harris; R.A. Gibbs; J. Rogers

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genome Sequencing Center, Baylor College Medicine, Houston, Texas.

---

Nonhuman primates provide unique and remarkably informative animal models for a wide range of human diseases and biomedical questions. In neuroscience, infectious disease, endocrine function and reproductive biology, developmental psychobiology and other aspects of biomedical research, nonhuman primates continue to play an indispensable role. Human genetic research has identified specific genes that influence a variety of human diseases, but for many pathogenic mutations, non-primate models fail to adequately model the human condition. We are documenting naturally occurring genetic variation in research colonies of rhesus macaques (*Macaca mulatta*), the most widely used nonhuman primate in biomedical research, and have produced Illumina whole genome sequence data for 537 animals, identifying more than 72 million unique single nucleotide variants (SNVs). Among those variants there are 339,699 unique missense variants affecting 19,929 protein coding genes, 94.4% of the genes annotated in the rhesus genome. This large pool of naturally occurring variation provides an outstanding and mostly untapped resource for the development and exploitation of spontaneous highly translational genetic models of human disorders. To evaluate the potential impact of this standing genetic variability, we downloaded the ClinVar database of human genetic variants and using the liftOver tool determined the orthologous base in the rhesus genome. We find 509 pathogenic or likely pathogenic variants from ClinVar that are also polymorphic in the macaques, pointing directly to hundreds of novel genetic models. Of those 509, 46 are pathogenic variants in autosomal dominant diseases (based on ClinVar records), making these immediate candidates for validation by Sanger sequencing and detailed phenotypic characterization of carriers. The dominant mutations include documented causes for cardiomyopathy, cancer, hypercholesterolemia, Angelman Syndrome and others. Given the significance of primates for studies of neurobiology, we investigated the potential relevance of the 509 shared pathogenic missense variants and found 65 are predicted to influence brain or neurological function. These analyses demonstrate that naturally occurring variation segregating in NIH funded colonies of nonhuman primates constitutes a unique, valuable and largely unexploited opportunity for rapid development of highly translatable animal models of numerous significant public health problems.

## PgmNr 3191: Human plasma protein levels associate to genetic and heterogeneous non-genetic factors including blood cell fractions.

### Authors:

B. Caron <sup>1,2</sup>; E. Patin <sup>3,4,5</sup>; B. Charbit <sup>6</sup>; B. Piasecka <sup>3,5</sup>; M.L. Albert <sup>7</sup>; L. Quintana-Murci <sup>3,4,5</sup>; D. Duffy <sup>6,8,9</sup>; A. Rausell <sup>1,2</sup>; The Milieu Intérieur Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Bioinformatics Lab, Imagine Institute, Paris Descartes University, Sorbonne Paris Cité, 75015 Paris, France; 2) INSERM UMR 1163, Imagine Institute, 75015 Paris, France; 3) Unit of Human Evolutionary Genetics, Pasteur Institute, 75015 Paris, France, EU; 4) Center of Bioinformatics, Biostatistics and Integrative Biology, Pasteur Institute, 75015 Paris, France, EU; 5) CNRS UMR 2000, 75015 Paris, France, EU; 6) Center for Translational Research, Pasteur Institute, Paris, France; 7) Insitro, South San Francisco, CA, USA; 8) Immunobiology of Dendritic Cell Unit, Pasteur Institute, Paris, France; 9) INSERM U1223, Pasteur Institute, Paris, France

Blood plasma proteins play an important role in immune responses, including cytokine signaling, the complement system and the acute-phase response. Variability in plasma protein levels is observed in the general population, and has been partly explained by both genetic and non-genetic factors, including age and sex. Quantitative Trait Loci (pQTLs) collectively involving more than 1,597 plasma proteins have been recently reported. However, previous studies of genetic determinants of plasma protein levels have lacked an exhaustive control of the concomitant blood cell composition which may act as a confounding factor.

Here we evaluated plasma protein levels across 400 unrelated healthy individuals of Western European ancestry stratified by sex, and by age across two decades (20-29 and 60-69 years) from the *Milieu Intérieur* population-based cohort. 297 proteins were quantified by antibody-based multi-analyte profiling using Luminex technology in a clinically certified laboratory. More than 5.2M single-nucleotide polymorphisms were assessed following genome-wide DNA genotyping and imputation. Non-genetic variables considered in the analysis, included cell counts of major immune cell types measured by blood tests and standardized flow cytometry as well as 188 lifestyle and biochemical factors.

A total of 704 significant associations involving 131 proteins and 75 non-genetic factors were identified. Consistent with previous studies, age, sex, body-mass index and glomerular filtration rate were among the factors with the highest number of associations. In addition, a total of 46 proteins were associated with specific blood cell fractions including erythrocytes, platelets, neutrophils, lymphocytes and eosinophils. After controlling for these covariates, we detected 101 and 12 pQTLs acting in *cis* and *trans* respectively, collectively associated to 93 plasma protein levels (False Discovery Rate adjusted p-value <0.05). These results replicate between 51% and 77% of the plasma proteins evaluated with QTLs previously reported. Moreover, we identified an additional set of 29 plasma proteins whose expression levels are under genetic control. Overall, genetic factors, followed by blood cell fractions, showed a larger contribution to variable plasma protein levels than age and sex. Our study highlights common human polymorphisms that may modulate immune phenotypes on

the basis of altered plasma protein levels.

## PgmNr 3192: Inference of target genes regulated by open chromosome regions in human brain.

### Authors:

R. chen <sup>1</sup>; Y. Jiang <sup>1</sup>; A. Shieh <sup>2</sup>; T. Goodmam <sup>2</sup>; Z. Shen <sup>3</sup>; H.M. Wang <sup>3</sup>; C. Chen <sup>4</sup>; C.Y. Liu <sup>2</sup>; B.S. Li <sup>1</sup>; PsychENCODE

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetic Institute, Department of Molecular Physiology & Biophysics, Vanderbilt University, Nashville, TN, USA; 2) Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY, USA; 3) Center for Human Genetics, The Brown foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA; 4) Center for Medical Genetics, School of Life Sciences, Central South University, Changsha, Hunan, China

The spatial and temporal gene expression involves in interaction between transcription factor (TF) and cis-regulatory element (CRE), abnormality of which plays a causal role in many human diseases. GWAS has identified thousands of SNPs related to complex diseases; however, most of them are in noncoding regions. Translating GWAS discoveries to biology requires linking CREs to their target genes. CREs in which GWAS loci are located are often distant to their target genes, making it nontrivial to link CREs to their target genes. Assay for Transposase-Accessible Chromatin (ATAC-seq) can assay genome-wide open chromatin profiling, providing opportunities for studying regulatory elements. Previous quantitative trait loci (QTLs) studies have indicated that accessibility of open chromatin regions and gene expression can be regulated genetically. Our preliminary data show overlaps of QTLs for open chromatin regions (chromatin accessibility QTLs - caQTLs) and gene expression (eQTLs), implying their co-regulation by same genetic variants, motivating us to identify open-chromatin-region-gene pairs leveraging the genetic sharing.

In the brainGVEX project of PsychENCODE, we assayed matched multi-omics data, including whole-genome sequencing (WGS), RNA-seq, ribosome profiling (Ribo-seq), ATAC-seq and proteomics, in frontal cortex samples from over 400 individuals. In this study, we first trained an elastic-net model to identify variants associated with a gene's expression, and then used the variants to predict ATAC-seq peaks in 200kb window centered the gene's transcription start sites (TSS). An ATAC-seq peak with non-zero coefficients for at least one variants, is considered to link to this gene. In the preliminary exploration, we focused on the 149 genes that have both eQTLs and caQTLs; these genes have a median of 27 ATAC-seq peaks and 46 variants to predict their gene expression. Among these genes, we identified 77 genes (47.7%) that have linked ATAC-seq peaks, of which a median of 6 variants was included in the predictive models. The median of R<sup>2</sup> is 0.18 for peak-gene pairs, comparing to the 0.02 of the non-relevant peaks. The preliminary results show the potential of our method to identify ATAC-seq peaks and their target genes, and we are using Hi-C data in public datasets to further validate our results. We will expand the analysis to whole genome and interpret their biological meanings in SCZ combing the data from GWAS.



## PgmNr 3193: Elucidating the role of mRNA splicing in disease pathogenesis.

### Authors:

A. Shah<sup>1</sup>; Y.I. Li<sup>2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL; 2) Section of Genetic Medicine, University of Chicago, Chicago, Illinois.; 3) Department of Human Genetics, University of Chicago, Chicago, IL

---

Up to 70% of disease-associated genetic variants do not have apparent effects on gene expression levels. This suggests that a large fraction of functional variants may disrupt gene regulation through alternate mechanisms, such as mRNA splicing. We have shown previously that mRNA splicing is a primary mechanism that links genetic variation to disease. Several studies have underscored this point by identifying splicing quantitative trait loci (sQTLs) in linkage disequilibrium with GWAS hits for a number of diseases, including Type 2 diabetes, Alzheimer's disease, and schizophrenia. This suggest that SNPs affecting splicing have the potential to be casual variants underlying a substantial number of disease-relevant GWAS hits. However, there exists no comprehensive analysis of the regulatory mechanisms by which genetic variation affects RNA splicing in a large number disease-relevant tissues. Therefore, we investigated the impact of genetic variation on mRNA splicing in 53 tissues from the Genotype-Tissue Expression (GTEx) project. Using LeafCutter, we mapped thousands of sQTLs in each tissue individually and quantified that 80% of splicing events and 95% of sQTL are shared across all tissues.

One limitation of sQTL analysis is that it is often difficult to disentangle the mechanisms by which SNPs affect splicing to contribute to disease. For example, SNPs can affect splicing by strengthening or weakening the sequence or recognition of splice sites and splicing regulatory elements, which can result in aberrant protein products and cause disease. Our hypothesis is that SNPs that increase disease risk through splicing likely do so by reducing splicing accuracy, resulting in the production of a number of non-functional mRNA transcripts. Therefore, we devised a novel strategy to detect non-functional transcripts produced as a result of error-prone splicing (i.e. 'noisy' splicing), mapped thousands of noisy splicing QTL (noisy sQTL) in disease-relevant tissues from GTEx, Depression Genes and Networks (DGN), and the CommonMind Consortium, and checked for associations between these non-functional transcripts and disease risk (e.g. for autoimmune, neurodegenerative, and cardiovascular diseases) by performing mRNA splicing imputation followed by splicing transcriptome-wide association studies (TWASs).

Our results provide the field with a better understanding of the molecular mechanisms underlying sQTL activity and effect size variability in the context of disease.

## **PgmNr 3194: Towards robust clinical genome interpretation: Developing a consistent terminology to characterise disease-gene relationships for harmonised curation - inheritance modes and disease mechanisms.**

### **Authors:**

A.M. Roberts <sup>1,2</sup>; N. Whiffin <sup>1,2,13</sup>; P. Theotokis <sup>1,2</sup>; M.T. DiStefano <sup>3</sup>; J. Amberger <sup>4</sup>; J.S. Berg <sup>3</sup>; E. Birney <sup>5</sup>; F. Cunningham <sup>5</sup>; S. Ellard <sup>3,6</sup>; H.V. Firth <sup>7,8</sup>; D.R. Fitzpatrick <sup>7,8</sup>; J. Goldstein <sup>3</sup>; A. Hamosh <sup>4</sup>; M.E. Hurles <sup>7,8</sup>; A. Marmignion <sup>9</sup>; C.L. Martin <sup>3</sup>; E.M. McDonagh <sup>10</sup>; S. Plon <sup>3</sup>; E.M. Ramos <sup>11</sup>; E.R. Riggs <sup>3</sup>; E. Tudini <sup>12</sup>; C.F. Wright <sup>8</sup>; H.L. Rhem <sup>3,13</sup>; J.S. Ware <sup>1,2,13</sup>; on behalf of the Gene Curation Coalition (GenCC)

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Imperial College London, UK; 2) Royal Brompton & Harefield NHS Trust, UK; 3) Clinical Genome Resource (ClinGen), USA; 4) Online Mendelian Inheritance in Man (OMIM), Johns Hopkins University School of Medicine, USA; 5) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom; 6) University of Exeter Medical School, Exeter, EX2 5DW, UK; 7) Transforming Genetic Medicine Initiative, Wellcome Genome Trust, England; 8) Deciphering Developmental Disorders/Gene2Phenotype, England; 9) Inserm-US14 - Orphanet, France; 10) Genomics England, Queen Mary University of London, Dawson Hall, Charterhouse Square London, EC1M 6BQ, UK; 11) National Human Genome Research Institute, National Institutes of Health, USA; 12) Australian Genomics Health Alliance; 13) Broad Institute of MIT and Harvard

---

The clinical application of genomic data is reliant upon robust, curated associations between locus, genotype, mechanism, and disease phenotypes.

Several groups provide resources for variant annotation to aid interpretation, but each of these use different terminology to describe inheritance, allelic requirement, and both structural and functional consequences of a variant. There is considerable discrepancy in the derivation and application of these terms.

The Gene Curation Coalition (GenCC), includes members of Clinical Genome Resource (ClinGen), DECIPHER, Gene2Phenotype (G2P), Transforming Genetic Medicine Initiative (TGMI), Genetics Home Reference (GHR), Genomics England PanelApp (PanelApp), Online Mendelian Inheritance in Man (OMIM), and Orphanet. Together, this group is working to harmonise approaches to ensure gene-level curated resources are comparable and interoperable, and use standardized terms for mode of inheritance, allelic requirement, genetic consequence, functional consequence, and mechanism of pathogenicity. This will allow these international groups to most effectively work together to provide consistent and useful resources for the community, including guiding the formation of gene panels for disease-targeted testing and genomic analysis.

Here we present an approach to standardize terminology for characterization of disease-gene relationships to facilitate harmonized curation, and support variant interpretation within the

ACMG/AMP framework.

Terminology used by all GenCC members and partners was reviewed. Terms for allelic requirement and inheritance coalesce around human phenotype ontology (HPO) terms, while terms for gene, transcript and functional consequence coalesce around sequence ontology (SO) terms. Use of these standardized terms will facilitate assessment of which variant classes are likely to be disease-relevant (e.g. to apply PVS1) and assessment of relevance of functional evidence, to determine which ACMG/AMP rules are applicable.

In addition to streamlining disease-gene curation and facilitating interpretation of variants in established disease-gene pairs, this standardized terminology will aid in assessment of novel potential disease-gene or disease-variant relationships. For a previously unseen variant, we usually understand the genetic consequence, but not (directly) functional effect. Similarly we may not know the precise mechanism for a disease or variant, but can interpret likely disease-relevant variant classes.

## PgmNr 3195: The functional impact of structural variation in the human brain.

### Authors:

R.A. Vialle <sup>1</sup>; K.P. Lopes <sup>1</sup>; J.F. Crary <sup>2,1</sup>; T. Raj <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Neuroscience, Loeb Center for Alzheimer's disease, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Pathology, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY

---

Structural variants (SVs) are an important source of genetic diversity and have been linked to many diseases. However, their contribution to molecular traits in the brain and its impact on neurodegenerative diseases remains unknown. The AMP-AD consortium provides an extensive collection of multi-omics data that allow us to identify SVs and characterize their functional impact. This includes deep whole-genome sequencing (WGS) from 1860 subjects from four aging cohorts (ROS/MAP, MSBB and Mayo Clinic). Here, we developed a rigorous pipeline integrating six different SV discovery tools plus merging and genotyping strategies to identify a total of 169330 high confidence SVs, present in at least one cohort, and 44919 SVs shared amongst all cohorts. While majority (~63k) SVs were also identified in 1KGP and gnomAD-SV, we also detected ~22k novel SVs in our dataset, suggesting potentially pathogenic variants. We identified previously known SV at *MAPT* locus and identified several novel associations to neuropathology traits. By integrating RNA-seq data from different brain regions, we mapped *cis* expression quantitative trait loci associated with SVs (SV-eQTLs). We identified a total of 945 SVs altering the expression of 1336 genes (FDR<0.05) in the dorsolateral prefrontal cortex (DLPFC) tissues from ROS/MAP, of which 5 SVs were tagged by GWAS SNPs related to neurodegenerative diseases. In Mayo Clinic and MSBB, we identified respectively 907 and 576 genes altered in at least one brain region, of which 6 and 10 SVs tagged by neurodegenerative GWAS SNPs like, for example, a known 157kb duplication in *KANSL1* (*MAPT* locus) that overlaps with a variant (rs2732703) associated with Alzheimer's Disease and a novel 56kb deletion affecting *HLA-DRA* and that overlaps with a variant (rs115641444) associated with Autism spectrum disorder and schizophrenia. Additionally, we found reproducible SV-eQTLs for 147 genes across all cohorts (including *C4A*, *ERAP2*, *GBP3*, *KANSL1*) that promoted widely varying levels of expression changes in the brain. In order to validate the SVs, we have generated targeted (*MAPT* locus) PacBio long-read sequencing data. We are now testing the impact of common and rare SVs on protein levels, splicing, histone acetylation, and methylation of DLPFC. Overall, we present the most comprehensive map of structural variation in aging cohorts, providing a valuable resource for understanding the functional impact of SVs in neurodegenerative and neuropsychiatric diseases.

## PgmNr 3196: Characterization of circulating cell-free miRNA signatures in cystic fibrosis.

### Authors:

J.E. Ideozu<sup>1,2,3</sup>; X. Zhang<sup>1,2,3</sup>; V. Rangaraj<sup>1</sup>; S. McColley<sup>1,3</sup>; H. Levy<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Pulmonary Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL; 2) Human Molecular Genetics Program, Stanley Manne Children's Research Institute, Chicago, IL; 3) Northwestern University Feinberg School of Medicine, Chicago, IL

---

**Introduction:** In cystic fibrosis (CF), dysfunction of the cystic fibrosis transmembrane conductance regulator (CFTR) protein, due to *CFTR* gene variants, results in a multisystem disease. Clinical heterogeneity seen among patients with the same *CFTR* genotype presents a major therapeutic challenge. Increasing evidence from transcriptome profiling studies suggests that several cellular microRNAs (miRNAs) act as regulators of gene expression and influence CF disease progression. However, their abundance and functional relevance in CF extracellular milieu remain poorly understood. We utilized microarray profiling to identify cell-free miRNAs differentially expressed between CF and non-CF subjects.

**Methods:** Plasma was extracted from peripheral blood of 5 CF and 5 non-CF subjects. Total RNA was then isolated from each sample and subjected to microarray profiling with Affymetrix miRNA 4.0 microarray. Labelling, hybridization, washing, and scanning was performed according to the manufacturer's recommendation. Microarray datasets were assessed for quality using Transcriptome Analysis Console. The Robust Multi-Chip Analysis (RMA) was utilized for background correction and normalization. The high-quality data were then mapped to miRbase mature miRNA annotation (v20) and differential expression analysis performed with Partek Gene-Specific Analysis algorithm. Cell-free miRNAs differentially expressed between CF and non-CF subjects with at least 2-fold difference (FDR<0.05) or more were prioritized for functional analysis.

**Results:** Analysis of the microarray dataset revealed that several miRNAs were abundantly expressed in CF. Among the key miRNAs significantly (FDR<0.05, log<sub>2</sub> FC<sup>3</sup> 2) differentially expressed between CF and non-CF subjects were members of the let-7, and 103/107. Using in silico methods we identified the top-ranked validated targets were genes involved mostly in miRNA biogenesis and gene expression, and signal transduction pathways were the most significantly (p<0.05) enriched pathways.

**Conclusion:** Cell-free miRNAs play a crucial role in several biological processes and are proven biomarkers for many diseases. Our results suggest their aberrant expression may be clinically in CF.

## PgmNr 3197: Carriers of rare protein-coding variants in *AS3MT* have reduced arsenic metabolism efficiency.

### Authors:

D. Delgado<sup>1</sup>; M. Chernoff<sup>1</sup>; L. Tong<sup>1</sup>; F. Jasmine<sup>1</sup>; M.G. Kibriya<sup>1</sup>; S.A. Cole<sup>2</sup>; K. Haak<sup>2</sup>; J. Kent<sup>2</sup>; J. Umans<sup>3</sup>; L.G. Best<sup>4</sup>; A. Navas-Acien<sup>5</sup>; M. Karagas<sup>6</sup>; H. Ahsan<sup>1,7,8,9</sup>; B.L. Pierce<sup>1,7,8</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Public Health Sciences, University of Chicago, Chicago, Illinois.; 2) Texas Biomedical Research Institute, San Antonio, Texas; 3) Georgetown-Howard Universities Center for Clinical and Translational Science, Washington, DC; 4) Missouri Breaks Industries Research, Inc., Timber Lake, South Dakota; 5) Mailman School of Public Health, Columbia University, New York City, New York; 6) Department of Epidemiology, Geisel School of Medicine, Dartmouth College, Hanover, New Hampshire; 7) Department of Human Genetics, University of Chicago, Chicago, Illinois; 8) Comprehensive Cancer Center, University of Chicago, Chicago, Illinois; 9) Department of Medicine, University of Chicago, Chicago, Illinois

---

Exposure to inorganic arsenic in drinking water is a global health hazard, affecting >200 million individuals worldwide. Common variation in the 10q24.32 region, containing the arsenic methyltransferase (*AS3MT*) gene, is associated with arsenic metabolism efficiency (AME) and arsenic-related diseases. Findings from familial aggregation studies suggest AME has substantial unexplained heritability, which may be attributable to rare variants. The key role of *AS3MT* in arsenic metabolism and the established association of 10q24.32/*AS3MT* common variants with AME makes *AS3MT* a strong candidate gene for rare variant studies of AME. We generated targeted sequencing data for the coding regions of *AS3MT* (and ~25 additional genes with posited roles in arsenic metabolism) for three arsenic-exposed cohorts: Health Effects of Arsenic Longitudinal Study (HEALS, n=2,343), Strong Heart Study (SHS, n=873), and New Hampshire Skin Cancer study of squamous cell carcinoma (NHSCC, n=666). We measured AME as the percentage of arsenic in urine that is dimethylarsinic acid (DMA), the end metabolite in a series of reduction and methylation reactions (catalyzed by *AS3MT*) that act to detoxify arsenic. We conducted gene-level association tests of the collective effects of rare protein-altering variants (MAF<0.01) in specific genes on DMA% using the SKAT-O R package. We applied MetaSKAT to calculate the combined effect of rare protein-coding variants on AME across all cohorts. All analyses adjusted for common SNPs in *AS3MT* known to be associated with DMA%. In HEALS, we identified three genes showing evidence of association with DMA%, *AS3MT* (P=0.005), *GSTO1* (P=0.05), and *SFR1* (P=0.04). In SHS, we observed associations for *AS3MT* (P=0.04) and *SLCOB1* (P=0.02), and in NHSCC, an association was observed for *GSTO1* (P=0.05). In the meta-analysis, only *AS3MT* (P=0.03) was associated with DMA%. Across all cohorts, the *AS3MT* gene-set primarily included singletons and doubletons, which were all missense mutations with the exception of one stop-gained mutation. Collectively, carriers of rare variants in *AS3MT* had low DMA% compared to non-carriers. Protein-coding *AS3MT* variant rs11191439 was excluded from gene-based tests due to high MAF (0.05, 0.15, and 0.07 for HEALS, SHS, and NHSCC, respectively). This variant showed an inverse association with DMA% across all cohorts (p<0.05) indicating it reduces AME. In summary, we provide evidence that rare variants in *AS3MT* affect AME in multiple populations.

## PgmNr 3198: Non-additive genetic effects on gene expression modulated by cell heterogeneity in whole blood.

### Authors:

A. Brown <sup>1,2</sup>; A. Viñuela <sup>2</sup>; A. Martinez-Perez <sup>3</sup>; N. Panousis <sup>2</sup>; A. Ziyatdinov <sup>3</sup>; M. Sabater-Lleal <sup>3,4</sup>; A. Hamsten <sup>4</sup>; J. Souto <sup>3</sup>; A. Buil <sup>5</sup>; J. Soria <sup>3</sup>; E. Dermitzakis <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Population Health and Genomics, University of Dundee, United Kingdom; 2) Genetic Medicine and Development, University of Geneva, Switzerland; 3) Instituto de Investigaciones Biomedicas Sant Pau, Barcelona, Spain; 4) Karolinska Institutet, Sweden; 5) Institute of Biological Psychiatry, Roskilde, Denmark

Thousands of genetic variants have been reported to influence blood cell counts, with implications for disease risk. However, our understanding of gene expression regulation in specific blood cells types and its relationship with disease remains unclear. To investigate the cell specific nature of gene expression and its additive and non-additive regulation, we quantified cell counts and RNAseq data in whole blood in 915 individuals from 35 pedigrees, including 390 parent-child trios and 180 parent-child duos. Almost all genes (99.8%) are significantly associated with at least one cell count, from 910 genes associated with basophils to 16,679 genes with lymphocytes. Next, we compared heritability ( $h^2$ ) estimates of gene expression before and after controlling for cell counts. Since cell type abundance is heritable, we expected cell counts to contribute to the genetic component of expression, decreasing  $h^2$ . However, we observed a significant increase (mean  $h^2 = 0.192$  to  $h^2 = 0.201$ ,  $P\text{value} = 4.36e-05$ ) suggesting most genetic effects on expression are mediated by other processes, with cell composition playing an important role in the environmental component.

To investigate non-additive regulation, we first looked genetic effects dependent of cell composition (cell-type x genotype interactions) and found 1,311 significant interactions in 636 genes across the 14 cell types. For 11 genes, we found evidence for multiple loci interacting with different cell types, suggesting multiple cell-specific regulatory elements. Secondly, we investigated whether parent-of-origin effects were also influenced by cell-type abundance. Focusing on known imprinting genes ( $N=195$ ), we looked for SNPs that would act in a parent-of-origin dependent manner (PO-eQTL) and also interact with cell-type composition. We detected 30 genes with PO-eQTLs and cell-types interactions. For 16 genes, the same or a nearby eSNP ( $R^2 > 0.8$ ) were significant for both effects, including genes such as *SNURF* (lymphocytes and neutrophils), *NOTCH3* (monocytes), and *CAST* (basophils).

In conclusion, we show the influence of genetic variation on gene expression in a specific tissue is combinatorially affected by factors such as cell composition and allelic inheritance: considering these factors together may be crucial to understanding the mediating role that gene expression plays between genetics and disease.

## PgmNr 3199: Predicting the effects of SNPs on transcription factor binding affinity.

### Authors:

S.S. Nishizaki <sup>1</sup>; N. Ng <sup>2</sup>; S. Dong <sup>3</sup>; C. Morterud <sup>3</sup>; C. Williams <sup>3</sup>; A.P. Boyle <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, Stanford University, Stanford, CA; 3) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI

---

GWAS have revealed that 88% of disease associated SNPs reside in noncoding regions. However, noncoding SNPs remain understudied, partly because they are challenging to prioritize for experimental validation. To address this deficiency, we developed the SNP effect matrix pipeline (SEMpl). SEMpl estimates transcription factor binding affinity by observing differences in ChIP-seq signal intensity for SNPs within functional transcription factor binding sites genome-wide. By cataloging the effects of every possible mutation within the transcription factor binding site motif, SEMpl can predict the consequences of SNPs to transcription factor binding. This knowledge can be used to identify potential disease-causing regulatory loci.



# PgmNr 3200: The molecular basis and genetic control of local gene co-expression.

## Authors:

D. Ribeiro <sup>1</sup>; A. Ramisch <sup>2,3,4</sup>; E. Dermitzakis <sup>2,3,4</sup>; O. Delaneau <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Department of Computational Biology, University of Lausanne, Lausanne, Switzerland; 2) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 3) Swiss Institute of Bioinformatics (SIB), University of Geneva, Geneva, Switzerland; 4) Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland

---

Gene regulation is an essential component of cellular function. Many Eukaryotic genes are not randomly distributed across the genome, but rather form clusters and tend to be expressed together. Indeed, nearby genes often show high expression correlation across individuals, defining co-expression domains (CODs). Recent studies highlight the existence of a complex network of structural and regulatory domains orchestrating the organised expression of nearby genes (Delaneau et al. 2019 Science 364(6439)). However, the full prevalence of local gene co-expression and their genetic control are yet far from being understood.

Here, we developed a novel unbiased method to identify groups of nearby genes (within 1Mb) that are co-expressed, and detect CODs genome-wide using RNA-seq data from large populations (e.g. >200 individuals). As a proof of concept, we identified 2556 high-confidence CODs (FDR 1%) using RNA-seq data from 358 human lymphoblastoid cell line (LCL) samples from the Geuvadis dataset. These CODs comprise 8885 (52.5%) out of 16907 protein-coding and lincRNA genes expressed in LCLs, demonstrating that local gene co-expression is highly prevalent.

Next, by integrating various ChIP-seq and Hi-C datasets, we functionally characterized CODs by studying their chromatin structure and local regulatory element activity. For instance, analysing Hi-C data for LCLs, we found that the region surrounding the transcription start sites of co-expressed gene pairs are often found interacting together, inclusive for genes that are >500Kb apart, evidencing the importance of chromatin 3D structure in gene co-expression.

Importantly, we further explored how co-expressed genes may be co-regulated by identifying expression Quantitative Trait Loci (eQTLs) for more than 70% of the LCL CODs, and investigating how eQTL effects propagate through the co-expressed genes. Notably, we applied our novel approaches to identify and compare thousands of CODs across 48 distinct human tissues present in the GTEx dataset, allowing us to study in-depth the tissue-specificity of gene co-expression and their co-regulation by genetic variants.

The systematic identification and characterisation of CODs and the study of the pleiotropic effects of eQTLs are of crucial importance to deepen our understanding of the regulatory mechanisms underlying gene expression and the functional impacts of eQTLs, ultimately allowing us to fine-map the effects of genetic risk factors involved in complex traits.

## PgmNr 3201: Enhancer domains predict gene pathogenicity and inform complex disease gene discovery.

**Authors:**

Wang<sup>1,2</sup>; D.B. Goldstein<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) Columbia University, New York, NY; 2) Institute for Genomic Medicine, Columbia University Medical Center, New York, NY

---

**Introduction:** Non-coding regulatory elements, such as transcriptional enhancers, are critical for the precise spatiotemporal regulation of gene expression. While enhancers have been shown to harbor common and rare variants associated with human disease, applying enhancer information to prioritize and discover disease genes remains a major challenge due to difficulties in identifying damaging non-coding variants and affected genes.

**Methods & Results:** Here, we demonstrate that the genes with larger cognate enhancer domains have greater importance in development and disease. This relationship is sufficiently strong that a machine learning classifier trained using only distal enhancer features predicts disease genes with accuracies comparable to or better than commonly-used metrics of gene constraint, including pLI, LOEUF and RVIS. Further, the relationship persists even after controlling for promoter conservation and genic constraint.

We investigated factors contributing to the biological basis for this relationship. We found that disease-relevant genes with large cognate enhancer domains are subject to both increased purifying selection and are buffered against disruptive regulatory variation due to enhancer redundancy. These genes are depleted for eQTLs and less affected by rare variants in their regulatory regions.

**Broader perspective:** To illustrate the practical consequences of this relationship on disease gene discovery, we focused on the problem of identifying causal disease genes from GWAS studies. We show that candidate causal genes at GWAS loci have larger enhancer domains, suggesting that causal disease genes may be harder to detect than bystander gene signals at the same loci. Collectively, these results provide insights into the identification of disease genes, as well as the disruption of gene regulation by regulatory variants.

## PgmNr 3202: Haplotype based expression quantitative trait loci mapping.

**Authors:**

R. Hofmeister; O. Delaneau

[View Session](#) [Add to Schedule](#)

**Affiliation:** Computational Biology, University of Lausanne, Lausanne, Waadt, Switzerland

---

Expression Quantitative Trait Loci (eQTLs) are genetic variations that affect the expression levels of genes. Considering that genes usually have multiples eQTLs, it is likely that nearby genetic variants contribute together to the variation in expression level, especially when affecting transcription factor binding sites or interactions between regulatory elements. Current studies mainly focus on the effects of individual genetic variants in the vicinity of genes but lack in identifying whether multiple variants can act synergistically as part of haplotypes.

In this work, we investigate whether using haplotype data instead of genotype data at specific variants can help to dissect eQTL signals. To this aim, we use a Hidden Markov Model (HMM) that clusters haplotypes into a small set of K 'founder' haplotypes and provides per-individual dosages for these that continuously vary along the genome in function of the underlying linkage disequilibrium structure. Then, we run linear regressions to test the resulting dosages for association with the expression at nearby genes (1Mb). This effectively builds a statistical test at the haplotype level that is run per-variant and that automatically accounts for the joint effects with surrounding variants. Finally, we used a Likelihood Ratio Test (LRT) to identify genes for which the underlying haplotypic structure better explains the variation in expression levels compared to genotypes alone, as typically done in standard eQTL analysis. This helps to unravel genes impacted by combination of variants (i.e. haplotypes) without having to define these combinations a priori.

Using the RNA-seq and genotype data for 358 individuals from the Geuvadis study, we first show that our clustering captures blocks of high linkage disequilibrium that closely match those defined in the HapMap project, thereby validating our clustering approach. Then, by running our association tests, we find evidences that ~20% of the eQTLs originally reported in the Geuvadis study can be better explained by haplotypic effects. Finally, we show how our approach can help to better characterize the genetic architecture of eQTLs by focussing on few particular genes for which we provide original and informative visualizations.

Overall, we believe that our method is well suited to determine whether an eQTL signal is of genotypic or haplotypic origin and constitutes a useful addition to the toolbox for fine-mapping eQTL effects.

## PgmNr 3203: Manual curation of homozygous predicted loss-of-function variants in 125,000 exomes reveals 1,752 human genes tolerant to inactivation.

### Authors:

E. Seaby<sup>1,2</sup>; M. Singer-Berk<sup>1</sup>; E. England<sup>1</sup>; D. Rhodes<sup>1,3</sup>; N. Watts<sup>1,2</sup>; M. Solomonson<sup>1,2</sup>; S. Baxter<sup>1</sup>; Q. Wang<sup>1,2,4</sup>; L. Francioli<sup>1,2</sup>; G. Tiao<sup>1,2</sup>; B. Cummings<sup>1,2,4</sup>; J. Alföldi<sup>1,2</sup>; K. Karczewski<sup>1,2</sup>; A. O'Donnell-Luria<sup>1,2,5</sup>; D. MacArthur<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 3) Centre for Translational Bioinformatics, William Harvey Research Institute, Barts and the London School; 4) Program in Biological and Biomedical Sciences, Harvard Medical School, Boston, MA, 02115, USA; 5) Division of Genetics and Genomics, Boston Children's Hospital, Boston, Massachusetts 02115, USA

With an increasing number of individuals undergoing diagnostic next-generation sequencing, there is a clear need to accurately delineate pathogenic variants from benign variation. Predicted loss of function (pLoF) variants, including stop gained, essential splice, and frameshift variants, are on average rare and deleterious, with many known to cause severe Mendelian disease through gene inactivation. On the other hand, some genes are tolerant to complete inactivation without severe phenotypic consequences; accurate curation of homozygous pLoF variants can expand our understanding into the biology of these genes. Given the rarity and deleteriousness of most true LoF variants, pLoF variants are enriched for mapping, genotyping and annotation errors; and therefore careful filtering and manual curation is required to remove such errors and identify probable LoF variants.

Using data from the aggregation of 125,748 exomes from human sequencing studies into the Genome Aggregation Database v2.1 (gnomAD), we extracted pLoF variants present homozygously in at least one individual. These variants were further filtered through the application of stringent Loss-Of-Function Transcript Effect Estimator (LOFTEE) criteria to distinguish high confidence pLoF variants from annotation artifacts. Additionally, variants in exons that do not show evidence of expression in adult tissue were removed (pext score < 0.1). To remove additional sequencing and annotation artifacts undetected by bioinformatic approaches, we applied deep manual curation to these 3,385 homozygous pLoF variants spanning 2,166 genes to identify high-confidence genes tolerant to human gene knockout, using a novel custom-built curation portal. We developed systematic criteria for pLoF variant filtering based on likely technical errors, rescue events and transcript errors.

This deep curation process resulted in 2,469 homozygous pLoF variants (73%) passing curation filters, resulting in a list of 1,752 high-confidence genes that are likely tolerant of human biallelic inactivation in the general population and that will be publicly released. This list of LoF tolerant genes was significantly depleted of essential genes (OR = 0.10, p-value =  $1.54 \times 10^{-17}$ ) and LoF constrained genes (p-value =  $4.48 \times 10^{-193}$ ). These results highlight a fraction of human coding genes that can

likely be deprioritized in disease gene discovery studies and highlight potential therapeutic targets tolerant of human inactivation.

## PgmNr 3204: Genetic influences and negative selection signatures on human blood metabolites in the Japanese population.

### Authors:

T. Iwasaki <sup>1</sup>; Y. Kamatani <sup>1</sup>; S. Kawaguchi <sup>1</sup>; K. Sonomura <sup>2</sup>; T. Sato <sup>2</sup>; F. Matsuda <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 2) Shimadzu Corporation, Japan

---

Blood metabolites are important biomarkers of aging and complex traits, including diseases. Genetic variations which influence their levels in the blood are known for some metabolites. In this study, we conducted a quantitative trait loci (QTL) analysis of targeted metabolome in the Japanese population to construct a comprehensive database of such QTL and to clarify the functional roles of associated genetic variants. A total of 131 soluble metabolites in plasma were measured by gas chromatography-mass spectrometry (Shimadzu GCMS-QP2010 ultra) for 8,314 healthy individuals enrolled in the Nagahama Study. Among them, 4,888 individuals had been genotyped and for those we imputed 10,499,485 genetic variants using 1000 Genomes Phase 3 v5 as a reference panel. These imputed genotypes were then used for the further analyses. Batch effects of metabolome measurements were corrected using a smoothing spline approach. We identified signals with genome-wide significance ( $p < 5.0 \times 10^{-8}$ ) in 132 chromosomal loci of which 52 were not previously reported. Eleven loci showed an association with more than one metabolite, and 24 loci showed pleiotropic associations with clinical phenotypes. Conditional analysis of each QTL showed that there were more than one association signals in 30 loci, which comprised a total of 51 independent QTL signals ( $p < 1.0 \times 10^{-5}$ ). The correlation of genome-wide genetic effects was different ( $p < 0.05$ ) between Japanese and Europeans for ten metabolites. We also found that significantly higher numbers of QTL genetic variants were classified as loss of function (4.0-fold,  $p = 2.3 \times 10^{-3}$ ) or non-synonymous substitution (5.0-fold,  $p = 2.6 \times 10^{-25}$ ), as compared with MAF- and chromosome-matched randomly chosen SNPs. This showed intriguing contrast with the results of disease GWAS in which susceptibility genetic variants are frequently located in regulatory regions. Finally, we tried to identify negative selection signatures of the genetic effects on metabolites by using a Bayesian mixed linear model (MLM) method. We found multiple metabolites showing significant selection signals, suggesting that these metabolites affect the fitness of individuals. In this context, they could be potential candidates for disease-related biomarkers.

## PgmNr 3205: Nuclear genetic control of human mitochondrial RNA modification and its role in complex disease.

### Authors:

A.T. Ali <sup>1</sup>; Y. Idaghdour <sup>2</sup>; A. Hodgkinson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, King's College London, London, UK, SE1 9RT; 2) Department of Biology, Division of Science and Mathematics, New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates

---

RNA modifications affect the stability and function of RNA species, regulating important downstream processes. Modification levels are often dynamic, varying between tissues and individuals, although it is not always clear what modulates this variation or what impact it has on biological systems. Here, we quantify variation in RNA modification levels at functionally important positions in the mitochondrial genome across 11,552 samples from 39 tissue/cell types and find evidence that modification levels impact mitochondrial transcript processing. We identify novel links between mitochondrial RNA modification levels in whole blood and genetic variants in the nuclear genome, including missense mutations in *LONP1* and *PNPT1*, as well as missense mutations in *MRPP3*, *SLC25A26* and *MTPAP* that associate with RNA modification levels across multiple tissue types. Genetic variants linked to modification levels are also associated with blood pressure and breast cancer, and we find that variation in both mitochondrial RNA modification and gene expression levels are associated with intermediate cardiovascular phenotypes, suggesting a role for these processes in complex disease.

## PgmNr 3206: Profiling the expression, regulation, and network connections of lncRNAs across human tissues.

### Authors:

O.M. de Goede <sup>1</sup>; N.M. Ferraro <sup>1</sup>; D.C. Nachun <sup>1</sup>; A.S. Rao <sup>1</sup>; F. Aguet <sup>2</sup>; A.N. Barbeira <sup>3</sup>; S. Castel <sup>4,5</sup>; F. Hormozdiari <sup>6</sup>; S. Kim-Hellmuth <sup>4,5</sup>; Y. Park <sup>7</sup>; C. Brown <sup>7</sup>; W. Wen <sup>8</sup>; T. Lappalainen <sup>4,5</sup>; H.K. Im <sup>3</sup>; K. Ardlie <sup>2</sup>; K. Kirkegaard <sup>1</sup>; S.B. Montgomery <sup>1</sup>; GTEx Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Stanford University, Stanford, CA; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) University of Chicago, Chicago, IL; 4) New York Genome Center, New York, NY; 5) Columbia University, New York, NY; 6) Harvard T.H. Chan School of Public Health, Boston, MA; 7) University of Pennsylvania, Philadelphia, PA; 8) University of Michigan, Ann Arbor, MI

---

Recent studies cataloguing lncRNA genes have expanded our understanding of this heterogeneous group. However, it is still difficult to distinguish functional lncRNAs from those that are byproducts of surrounding transcriptional activity. To identify potential hallmarks of biological function, we used GTEx v8 data to profile the expression, regulation, and gene network connectivity of lncRNAs across 49 tissues.

Compared to expression-matched coding genes, lncRNAs displayed highly tissue-specific patterns of both gene expression and expression quantitative trait loci (eQTLs). The testis, brain, skin, and blood had the highest numbers of tissue-specific expressed genes and eGenes. Fewer lncRNAs had eQTLs compared to expression-matched coding genes, at ~50% versus ~75%, respectively. However, out of all eGenes, similar proportions of lncRNAs and coding genes (~25%) had independent associations with >1 eVariant.

Outlier expression analysis identified 942 lncRNAs for which at least one individual was a multi-tissue outlier. Of these, 301 were tissue-specific lncRNAs; 110 of these were expressed in unexpected tissues at TPM >0.5 in the outlier individual. Of note is *AC025219.1*, for which there were multiple over-expression outlier individuals that all had the same rare (<1%) variant at the gene's transcription start site, highlighting the potential for rare variants to modulate lncRNA tissue specificity.

We built co-expression networks for each of the GTEx tissues. To interpret these networks, modules were annotated based on similarity to gene expression signatures of cell types and cell compartments. Overall, lncRNAs were less connected in their assigned modules than coding genes. However, there were examples of modules involving lncRNAs with high intra-module connectivity, such as epithelium modules in the lung, small intestine, colon, and vagina.

We assessed novel lncRNA-trait associations through colocalization of GWAS and lncRNA eQTLs. We found 4843 instances of significant lncRNA colocalization events (PP > 0.5). Of these, 3562 had a lncRNA PP at least 0.3 higher than any coding gene within +/-1Mb. Integrating network modules with these colocalization events identified compelling disease-associated lncRNAs, such as *C6orf3* for psoriasis and *LINC01475* for IBD.

This work provides a hypothesis-generating resource to interrogate individual lncRNAs, as well as to systematically discover and annotate novel functional lncRNAs with cell-type and human trait relevance.



## PgmNr 3207: Machine learning using nonlinear stacking method for predicting long non-coding RNA and enhancer transcription from genome sequence.

### Authors:

M. Koido <sup>1</sup>; C.C. Hon <sup>1</sup>; H. Kawaji <sup>1</sup>; Y. Murakawa <sup>1</sup>; Y. Kamatani <sup>1,2</sup>; P. Carninci <sup>1</sup>; C. Terao <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Kanagawa, Japan; 2) Graduate School of Frontier Sciences, the University of Tokyo, Tokyo, Japan; 3) Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan

---

Genome-wide association studies (GWAS) have found many non-coding single nucleotide variants (SNVs) associated with human complex traits. So far, catalogs of expression quantitative trait loci (eQTL) have provided plausible and novel interpretations of GWAS findings in a cell-type dependent manner; however, eQTL for lowly expressed non-coding RNAs including enhancer RNAs is poorly understood due to the lack of large-scale studies measuring them. Here we show that, by using only genome sequence surrounding transcriptional start site (TSS) as input data, machine learning (ML) can predict cell-type specific expressions of promoters and enhancers (as a binary “on-off” switch) in Cap Analysis of Gene Expression (CAGE). We designed ML models by combining deep convolutional neural networks (from sequence to epigenetic features, using publicly available pre-trained models) and binary classifier using gradient boosting trees (from epigenetic features to transcription probabilities). We leveraged the transcriptomes of 347 major human cell-types and tissues from the Functional Annotation Of the Mammalian genome (FANTOM5). We trained the ML models with the use of the autosomal promoter and enhancer expressions profiled by CAGE except for chromosome 8 and tested the accuracy using those on chromosome 8. As a result, ML models using  $\pm 100$ -kb sequences from TSS showed high accuracy to predict transcription (mean ( $\pm$  SD) area under the ROC curve (AUROC) =  $0.83 \pm 0.22$  (n = 347)). Notably, 295, 125, and 152 among the 347 models achieved AUROC > 0.7 for divergent-lncRNA, intergenic-lncRNA, and enhancer peaks, respectively. We further explored false positive signals and found that 31-70% of these CAGE peaks were actually expressed according to the profiles of nascent elongating RNAs. These findings indicated the uniqueness of our ML models to predict expressions of unstable, lowly expressed non-coding RNAs. Finally, we calculated effects of SNVs on proximal promoter transcription and found high concordance between the predicted effects on gene-level expressions (aggregated promoter-levels) and known eQTL effects (e.g., increased probability by 10% achieved ~90% of the direction accuracy in the blood model). Collectively, these findings indicate the potential use of our ML models to efficiently investigate genetic architecture for non-coding transcripts without additional population studies.

## PgmNr 3208: A map of single cell *cis*-eQTLs in human brain tissue.

### Authors:

K.P. Lopes; R.A. Vialle; T. Raj

[View Session](#) [Add to Schedule](#)

**Affiliation:** Departments of Neuroscience, Genetics & Genomic Sciences, Loeb Center for Alzheimer's disease. Icahn School of Medicine at Mount Sinai

---

Genome-wide association studies have identified thousands of genetic variants that are associated with several diseases. However, the causal genetic variants and underlying mechanisms for many of these associations remain unknown. In order to identify relevant variants that can impact in central nervous system (CNS) cell types, we performed expression quantitative trait loci (*cis*-eQTLs) analysis in single-cell RNA-seq (scRNA-Seq) data from prefrontal cortex tissue. The scRNA-seq data was previously generated from post-mortem human brain samples from the ROSMAP dataset (Mathys *et al.* 2019). To perform eQTL analysis, we leveraged scRNA-seq from 80,000 cells from 39 donors (21 with AD pathology, 18 controls) and genotypes called from Whole Genome Sequencing. Gene expression was measured by averaging the normalized expression per gene per donor and per major cell type (and also within subtypes). We identified from 12 to 316 significant genes ( $FDR < 0.1$ ) in eight cell types, of which excitatory neurons (230) and oligodendrocytes (316) had the most genes with *cis*-eQTLs followed by microglia (69). For the most part, the number of *cis*-eQTLs per cell types correlates well with number of cells per each cluster. Of these, 66% of *cis*-eQTLs were not detected at bulk tissue level, suggesting a substantial proportion are specific to a cell type. We subsequently performed *cis*-eQTL analysis on 34 subpopulations within major cell types. This resulted in several *cis*-eQTLs that were only present in specific subpopulations including *IFITM2* ( $P = 1.15 \times 10^{-7}$ ) and *MEF2C* ( $P = 1.23 \times 10^{-4}$ ) in microglia1, a cluster associated with AD and damage-associated microglia. Although we are underpowered, we identified one novel *cis*-eQTL (*FUS*) colocalized with both Alzheimer's and Parkinson's disease loci (rs1549299;  $P = 8.5 \times 10^{-5}$ ) and several *cis*-eQTLs (nominal  $P < 10^{-3}$ ) in microglia that colocalized with AD risk alleles (*FCER1G*, *PILRB*, *SLC24A4*, *MEF2C* and *MS4A4A*), consistent with emerging narrative of the role of microglia in AD. We also identified *cis*-eQTLs colocalized with AD risk alleles in astrocytes or oligodendrocytes, suggesting non-myeloid contribution to AD cells. We are in the process of replicating these eQTLs in an independent cohort ( $n = 107$ ) with isolated primary microglia. However, a larger dataset is needed to fully uncover effects of genetic variation on gene expression in subpopulations, which in turn, will help in elucidating the contribution of distinct cellular subsets in CNS diseases.

## PgmNr 3209: Intronic deletions that reduce the minimal distance between 5' splice-site and branchpoint define a pathogenic class of splice variant.

### Authors:

S.J. Bryen<sup>1,2</sup>; H. Joshi<sup>1</sup>; F.J. Evesson<sup>1,2,3</sup>; C. Girard<sup>4</sup>; R. Ghaoui<sup>1,2,5</sup>; L.B. Waddell<sup>1,2</sup>; A.C. Testa<sup>6</sup>; B. Cummings<sup>7,8</sup>; S. Arbuckle<sup>9</sup>; N. Graf<sup>9</sup>; R. Webster<sup>10</sup>; D.G. MacArthur<sup>7,8</sup>; N.G. Laing<sup>6</sup>; M.R. Davis<sup>11</sup>; R. Lührmann<sup>4</sup>; S.T. Cooper<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Kids Neuroscience Centre, The Children's Hospital at Westmead, Westmead, Sydney, NSW 2145, Australia; 2) Discipline of Child and Adolescent Health, Sydney Medical School, University of Sydney, Camperdown, Sydney, NSW 2006, Australia; 3) The Children's Medical Research Institute, 214 Hawkesbury Road, Sydney, NSW 2145, Australia; 4) Max-Planck-Institute for Biophysical Chemistry, Department of Cellular Biochemistry, Am Fassberg 11, 37077 Göttingen, Germany; 5) Royal Adelaide Hospital, Department of Neurology, Adelaide, SA 5000, Australia; 6) Centre for Medical Research, The University of Western Australia, Harry Perkins Institute of Medical Research, Perth, WA 6009, Australia; 7) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA; 8) Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142, USA; 9) Department of Histopathology, The Children's Hospital at Westmead, Sydney Children's Hospital Network, Sydney, NSW 2145, Australia; 10) Neurogenetics Clinic, The Children's Hospital at Westmead, Sydney Children's Hospital Network, Sydney, NSW 2145, Australia; 11) Neurogenetic Unit, Department of Diagnostic Genomics, PathWest Laboratory Medicine, QEII Medical Centre, Perth, WA 6009, Australia

---

A precise genetic diagnosis is the single most important step for families with genetic disorders to enable personalized and preventative medicine. In addition to genetic variants in coding regions (exons) that can change a protein sequence, abnormal pre-mRNA splicing can be devastating for the encoded protein, inducing a frame-shift or in-frame deletion/insertion of multiple residues. Non-coding variants that disrupt splicing are extremely challenging to identify. Stemming from an initial clinical discovery in two index Australian families, we define 25 families with genetic disorders caused by a class of non-coding splice variant due to intronic deletions. These pathogenic intronic deletions spare all consensus splice motifs, though critically shorten the minimal distance *between* the 5' splice-site (5'SS) and branchpoint. The mechanistic basis for abnormal splicing is due to biophysical constraint precluding U1/U2 spliceosome assembly, which stalls in A-complexes (that bridge the 5'SS and branchpoint). Substitution of deleted nucleotides with non-specific sequences restores spliceosome assembly and normal splicing; arguing against loss of an intronic element as the primary causal basis. Incremental lengthening of 5'SS-branchpoint length in our index *EMD* case defines 45 - 47 nt as the critical elongation enabling (inefficient) spliceosome assembly. The 5'SS-branchpoint minimal length mechanism, not currently factored by genomic informatics pipelines, is relevant to diagnosis and precision medicine across the breadth of Mendelian disorders and cancer genomics.

## PgmNr 3210: Tissue and individual variability of regulatory variant effects and the role of transcription factors.

### Authors:

E.D. Flynn <sup>1,2</sup>; S.E. Castel <sup>1,2</sup>; H. Zhang <sup>3</sup>; F. Aguet <sup>4</sup>; X. Li <sup>5</sup>; R. Agrawal <sup>5</sup>; A. Brown <sup>6</sup>; F. Hormozdiari <sup>7</sup>; GTEx Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Systems Biology, Columbia University, New York, NY; 2) New York Genome Center, New York, NY; 3) Department of Medicine, Columbia University Medical Center, New York, NY; 4) Broad Institute, Cambridge, MA; 5) Department of Biological Sciences, Columbia University, New York, NY; 6) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 7) Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA; 8) The Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, CA

---

The GTEx Consortium has identified thousands of genetic regulatory variants in *cis* (*cis*-eQTLs) that affect gene expression in multiple tissues. Over a third of GTEx eQTLs are active in most tissues, often with varying effect sizes, but the molecular mechanisms of cross-tissue eQTL effect size variability are poorly understood. Since eQTLs are enriched in transcription factor binding sites (TFBS), we developed a model of eQTL effect size as a function of transcription factor (TF) levels across tissues.

We investigated mechanisms of tissue variability of eQTL effect size in the GTEx v8 dataset (838 individuals with WGS, 15,201 RNA-seq across 49 tissues). We first examined the correlation of eGene expression with eQTL absolute effect size across tissues and identified 1,864 expression-correlated eQTLs at a 5% FDR, with 51% positive and 49% negative correlations. This shows that expression level and genetic regulatory effect are not independent phenomena. We then proceeded to correlate TF activity with eQTL effect size across tissues. We find that over a third of all tested eQTLs correlate with the activity of at least one TF, and we find an enrichment of the relevant TFBS overlap among the TF-correlated eQTLs.

Furthermore, we examined patterns of cross-individual variability of eQTL effect size within each GTEx tissue, using eQTL effect sizes measured from allelic expression data. We find that these patterns support our hypothesized cross-tissue TF regulators but also discover novel potential TF regulators that are not observable in the cross-tissue data. We validate our cross-tissue and cross-individual findings by analysis of allelic binding in ENCODE ChIP-seq data.

We highlight an interesting example of PU.1 regulation of a *LIPA* gene eQTL that colocalizes with a GWAS signal for coronary artery disease, and we provide data from functional experiments that support this hypothesis.

Our results provide a basis for identifying regulatory mechanisms of eQTL activity and for understanding eQTL effect size variability in the context of TF activity. This has major implications for future work investigating context-specific mechanisms of GWAS loci and their effects on human phenotype and disease.

## PgmNr 3211: Transposable element mediated rearrangements are prevalent in human genomes.

### Authors:

P. Balachandran <sup>1</sup>; J.N. Dayton <sup>1</sup>; D.C. Tafur <sup>1</sup>; G. Ananda <sup>1</sup>; C.R. Beck <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Jackson Laboratory for Genomic Medicine, Farmington, CT.; 2) Department of Genetics and Genome Sciences, UConn Health, Farmington, CT.

---

Transposable elements comprise more than half of the human genome, yet their roles in genomic instability remain poorly understood. Structural variants between two distinct repetitive sequences are termed Transposable Element Mediated (TEM) rearrangements, and differ from transposon insertion and deletion events. The vast majority of TEM events are between homologous repeats, and they are prevalent when examined throughout primate evolution. Due to the nature of their breakpoints, TEM events are challenging to detect using short-read whole genome sequencing. We implemented a multi-tool (Lumpy, Delly, GROM, and Manta) approach on three extensively characterized trios of individuals from the Human Genome Structural Variation Consortium to identify TEM events in short-read sequencing data. Overall, we found that ~17% of the rearrangements in these genomes were TEM events, in comparison to ~25% mediated by transposable element insertions and deletions. Orthogonal support for almost half of the TEM events was provided by PacBio data. Experimental validation confirmed the fidelity of the calls also present in long-read sequencing data, and our validation for non-supported calls will be used to develop a classifier to better predict true TEM events. Overall, we found 2321 distinct rearrangements across our cohort spanning 8 Mb of the human genome; 134 of these rearrangements may disrupt exons of protein-coding genes. Of the 2321 TEM events, 1475 were absent from the Database of Genomic Variants, and the number of events per trio was consistent with the evolutionary histories of the three populations (Yoruba, Puerto Rican and Han Chinese). We next investigated potential mechanisms of TEM rearrangements. Previous studies indicated selection against *Alu* elements that are separated by <1 Kb and on opposite strands (inverted *Alus*), and proposed that they contribute to genomic instability. Interestingly, the incidence of inverted *Alus* is significantly enriched around the breakpoints of TEM rearrangements. Furthermore, our yeast model of template switching indicates that inverted *Alu* pairs can increase the rate of genomic rearrangements. Our data elucidate a potential role for local genome sequence in the generation of double strand breaks that are then repaired with homologous substrates, show the importance of transposable elements in generating human genomic rearrangements, and implicate inverted repeats in driving local genomic instability.

## PgmNr 3212: Assessing functional assays: A case study of functional data use in variant interpretation using a structured evidence curation framework.

### Authors:

S.M. McNulty <sup>1</sup>; D.M. Kanavy <sup>1</sup>; M.K. Jairath <sup>1</sup>; S.E. Brnich <sup>1,2</sup>; C. Bizon <sup>3</sup>; B. Powell <sup>1</sup>; J.S. Berg <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 2) Curriculum in Genetics and Molecular Biology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; 3) Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

---

Functional assay data represents a powerful tool in the clinical interpretation of sequence variants, necessitating clear standards for the application of such evidence. The 2015 ACMG/AMP variant interpretation guidelines indicate “well-established” *in vitro* or *in vivo* functional studies can be applied as strong evidence for or against pathogenicity (PS3/BS3). However, differences in the application of functional data are a major contributor to discordance in variant interpretation between clinical labs, suggesting that current guidance is insufficient for the consistent evaluation of functional evidence. With the goal of identifying hallmarks of “well-established” assays, we assessed the use of functional data by experts in multiple disease areas through a systematic analysis of assays approved for PS3/BS3 application by six Variant Curation Expert Panels (VCEPs) associated with and approved by the Clinical Genome Resource (ClinGen). To aid in this process, we developed and employed a structured framework for the curation of primary literature cited by VCEPs in their pilot variant interpretation. Through our analysis, we observed variable specification of key experimental parameters: controls, replicates, thresholds, and validation, as well as variation in the satisfaction of these suggested parameters by individual assay instances cited as evidence. We further identified four recurring areas of discordance that may contribute to differences in functional evidence application: (1) methodology for estimating the predictive power of assays, (2) use of functional data from patient-derived experimental materials, (3) unclear recommendations for creation and interpretation of model organism evidence, and (4) limited guidance for conflicting evidence. Overall, this analysis highlights the many factors that require consideration in functional assay evaluation and suggests insufficient guidelines may act as a blockade to uniform curation, particularly in disease areas not yet addressed by VCEP recommendations. To advance the current guidelines, findings from this study are being used to inform the development of guidance for functional evidence application in variant interpretation to improve the accessibility of functional study analysis and to ensure minimum quality standards are met by studies cited as evidence. Furthermore, the framework established in this study is being used to develop a data model to structure functional data evidence in a computable form.

## PgmNr 3213: MicroRNAs that target the tripartite complex CARD14-BCL10-MALT1 and could potentially be targeted to downregulate a psoriasiform response.

### Authors:

J.M. Frost <sup>1</sup>; J.P. Sundberg <sup>2</sup>; J. Krueger <sup>3</sup>; A.M. Bowcock <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.; 2) The Jackson Laboratory, Bar Harbor, Maine, USA; 3) Investigative Dermatology, Rockefeller University, New York, New York, USA; 4) Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

---

We previously performed small RNA sequencing and identified 98 differentially expressed miRNAs and a handful of novel siRNAs in psoriatic versus healthy skin. We also discovered highly penetrant, gain-of-function, dominantly acting mutations within the human caspase recruitment domain family, member14 (*CARD14*) gene that led to the development of PS and PSA. *CARD14* mutations lead to enhanced NF-κB signaling and activation of a subset of psoriasis-associated genes in keratinocytes. This enhanced NF-κB signaling is due to a tripartite complex formed between CARD14-BCL10-MALT1. B cell lymphoma 10 (BCL10) and mucosa-associated lymphoid tissue lymphoma translocation protein 1 (MALT1). The introduction of the de-novo p.E138A alteration into the *Card14* gene of mice leads to features of psoriatic skin including a thickened epidermal barrier and immune cell infiltration into lesions. Moreover, the altered transcriptomes of the skin of p.E138A versus WT mice are very similar to that of human psoriatic skin versus that of healthy controls. Guided by these preliminary data we validated a set of psoriasis specific differentially expressed miRNAs in the skin of human PS patients and that of the p.E138A *Card14* mouse knockin with qRT-PCR. Two PS specific microRNAs, miR-26a-1-3p and miR-30e-5p, were predicted to target *BCL10* mRNA, one of the partners of CARD14 upon activation of NF-κB signaling. We confirmed the interaction of miR-30e-5p with the 3'UTR of *BCL10* mRNAs using luciferase assays (p=0.001) and hypothesize that this miRNA is a regulator of CARD14 signaling that can be targeted to downregulate a psoriasiform response.



## PgmNr 3214: Analysis of TFAP2A-binding events in a cellular model for nonsyndromic cleft lip +/- palate.

### Authors:

J. Welzenbach <sup>1</sup>; K.L.M. Ruff <sup>1</sup>; J. Fazaal <sup>1</sup>; F. Thieme <sup>1</sup>; E. Mangold <sup>1</sup>; M. Knapp <sup>2</sup>; K.U. Ludwig <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Institute for Medical Biometry, Informatics and Epidemiology (IMBIE), University of Bonn, Bonn, Germany

Non-syndromic cleft lip with or without cleft palate (nsCL/P) is a multifactorial malformation with a prevalence of about 1 in 1,000 livebirths. Despite recent success in the identification of underlying genetic risk factors, the functional effects of the risk variants are poorly understood. In the present study we hypothesize that transcription factor (TF) binding events, that occur in relevant cell types and might be modified in an allele-specific manner, contribute to a further understanding of the pathobiology of nsCL/P. To investigate this, we established cell culture and ChIP-Seq of human embryonic palatal mesenchyme cells (HEPM, ATCC CRL-1486), a cell line obtained from the facial region of a ten-week old female fetus which has been shown to represent a model for the developing and fusing secondary palate.

Based on expression profiling with 3'RNA-Seq, we identified several TFs that are expressed in HEPM cells and for which evidence for an involvement in craniofacial development has been presented. We focused our subsequent molecular work on TFAP2A that has been shown with a clear role in nsCL/P and evidence for which includes *in vivo* knockout models with craniofacial defects. Using a polyclonal TFAP2A antibody we performed ChIP-seq in HEPM, in three replicates. After quality control of sequence data (~40 Mio, incl. input control), data analysis was performed using an in-house pipeline. Overall, we observed about 7,000 peaks, with motif analysis confirming a strong enrichment for the TFAP2A consensus binding within the peaks summits ( $P < 2.04 \times 10^{-4}$ ). We next identified positional overlaps of results from our in-house GWAS within 865 high-confidence peaks, and extracted P-values for association with nsCL/P. We observed 657 SNPs within these regions, none of which showed genome-wide significance. We also observed five SNPs with  $P < 10^{-3}$  which might represent risk variants with lower effect sizes that have escaped detection in GWAS so far. This, however, requires further replication to determine the genuine association. Next we will explore preferential allele-specific binding, using heterozygous positions in HEPM and read coverage in the ChIP-Seq data. Our strategy will shed light on molecular events that are triggered by associated risk variants and help to identify regulatory networks in the developing palate. Furthermore, our study might contribute to an identification of novel risk loci associated with nsCL/P.



## PgmNr 3215: Investigation of microRNA-149-3p in neural crest cells and its role in orofacial clefting.

### Authors:

R. Hollstein <sup>1</sup>; L.G. Stüssel <sup>1</sup>; M. Laugsch <sup>2,3,4</sup>; F. Haeberlein <sup>5</sup>; L.M. Hochfeld <sup>1</sup>; J. Welzenbach <sup>1</sup>; J. Schröder <sup>1</sup>; F. Thieme <sup>1</sup>; S. Heilmann-Heimbach <sup>1</sup>; B. Odermatt <sup>5</sup>; E. Mangold <sup>1</sup>; C.P. Schaaf <sup>3,4</sup>; A. Rada-Iglesias <sup>2,6</sup>; K.U. Ludwig <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Human Genetics, University Hospital of Bonn, Bonn, Germany; 2) Center for Molecular Medicine Cologne (CMMC), University of Cologne, Cologne, Germany; 3) Institute of Human Genetics, University of Cologne, Cologne, Germany; 4) Institute of Human Genetics, University of Heidelberg, Heidelberg, Germany; 5) Institute of Anatomy, University Hospital of Bonn, Bonn, Germany; 6) Cologne Excellence Cluster for Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, German

---

Nonsyndromic cleft lip with/without cleft palate (nsCL/P) is a common facial malformation that occurs during the first weeks of embryonic phase. There is increasing evidence that human neural crest cells (hNCC) are major contributors to craniofacial development as this mesenchymal precursor cell population gives rise to the majority of cranial cartilage and bones. nsCL/P has a multifactorial etiology with a strong genetic component. The majority of associated risk loci, mainly identified through GWAS, map to non-coding regions of the genome. This suggests that a substantial fraction of nsCL/P pathomechanisms are mediated through regulatory effects on gene expression, one mechanism of which might be posttranscriptional gene regulation by micro RNAs (miRNAs). We first generated a catalogue of miRNAs expressed in hNCC, using array-based miRNA profiling of four hNCC samples that had been derived from human induced pluripotent stem cells (hiPSC). This revealed a set of 125 candidate miRNAs which was next integrated with in-house GWAS data on nsCL/P. We identified miR-149-3p as strong candidate for nsCL/P, based on consistent expression in hNCC across replicates and the presence of associated risk variants within its genomic region. To investigate the impact of miR-149-3p on neural crest development, we modified miR-149-3p abundance in hiPSC-derived hNCC through overexpression and inhibition. Expression profiling by RNA-Seq revealed differential expression of a number of candidate genes involved in craniofacial development and nsCL/P, such as *BMP7*. Moreover, using scratch assays, we found that the modification of miR-149-3p significantly affected migration of hNCC, which is one of the fundamental properties of hNCC. Future directions include the characterization of the molecular mechanisms of miR-149-3p target genes *in vivo*, using gene knockdown analyses in zebrafish. Overall, our data suggest that the integration of large-scale genetic data and expression patterns in relevant cell types can help to pinpoint regulatory mechanisms that are involved in craniofacial development and might be related to the etiology of nsCL/P.

## PgmNr 3216: An innovative approach to characterizing longitudinal changes in gene expression that underlie disease progression and translational relevance in a large clinical resource: The path to diabetes.

### Authors:

J.E. Below <sup>1</sup>; H.H. Chen <sup>1</sup>; L.E. Petty <sup>1</sup>; K.E. North <sup>2</sup>; J.B. McCormick <sup>3</sup>; S.P. Fisher-Hoch <sup>3</sup>; E.R. Gamazon <sup>1,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Vanderbilt Genetics Institute, Vanderbilt Univ Med Center, Nashville, Tennessee. Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TX, USA; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA; 3) The University of Texas Health Science Center at Houston (UTHealth) School of Public Health, Brownsville, TX, USA; 4) Clare Hall, University of Cambridge, Cambridgeshire, United Kingdom

---

Type 2 diabetes (T2D) is a complex and heritable disease, however the regulatory genetic mechanisms involved in progression to diabetes are largely unknown. We developed a two-stage approach to identify changes in gene expression associated with T2D transition and validate their clinical impact in a large biobank. In the transcriptome discovery stage, we studied longitudinal peripheral blood RNA sequence data from two time points in Hispanic study participants. All subjects had fasting blood glucose (FBG) <126 mg/dl (mean 106.7 mg/dl) at baseline. Twenty-four subjects' FBG transitioned to levels diagnostic of T2D ( $\geq 126$  mg/dl) at follow-up visit (cases) while the other 34 subjects maintained FBG <126 mg/dl (controls). We implemented a novel network methodology to identify genes with effect on transition status and evaluated its predictive performance, using an external comprehensive GWAS repository. In the biobank validation stage, we independently validated the identified set of genes in 23,000 participants from the large biobank at Vanderbilt University, BioVU. Associations of genetically determined expression of identified genes with diabetes-related traits in the electronic health record were used to establish their clinical significance. We identified 17 networks, including one comprising 822 genes with significant effect on transition and in which the connectivity of a gene was significantly correlated with the gene's association with T2D transition ( $p = 3.8 \times 10^{-29}$ ). This network is enriched for known metabolic trait genes identified by GWAS ( $p = 0.013$ ), but also includes novel associations. In this network, *WDFY3* accounted for the most variance, *STX3* exhibited the highest connectivity, and expression changes in *MIR3605*, *ASB9P1*, *NUDT16*, and *MKNK1-AS1* had the highest correlation with transition status. In BioVU, this network was significantly enriched for genes whose genetically determined expression is associated with T2D (enrichment  $p = 0.012$ ) and other metabolic traits in our biobank validation analysis. Importantly, we validated 43 novel gene-level associations with T2D, never reported by previous GWAS. Gene Ontology analysis underscored the importance of chronic inflammation and activated innate immunity in expression changes that occur during development of T2D. In this study we present a novel approach to identify key gene expression changes associated with T2D onset and demonstrate their translational relevance in a large clinical resource.

## PgmNr 3217: Maps of high-resolution chromatin interactions in human pancreatic beta-cells elucidate gene regulatory mechanisms at loci associated with type 2 diabetes.

### Authors:

J.M. Torres<sup>1</sup>; V. Nylander<sup>2</sup>; J. Telenius<sup>3</sup>; A.P. Morris<sup>1,4</sup>; D. Downes<sup>3</sup>; A. Mahajan<sup>1</sup>; M.I. McCarthy<sup>1,2,5</sup>; A.L. Gloyn<sup>1,2,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Wellcome Centre for Human Genetics, The University of Oxford; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford; 3) Weatherall Institute for Molecular Medicine, University of Oxford; 4) Institute of Translational Medicine, Department of Biostatistics, University of Liverpool; 5) NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK

---

Genome-wide associations studies (GWAS) have identified >400 signals at 250 loci associated with type 2 diabetes (T2D) risk. Yet resolving causal genes at these regions remains a challenge as the majority of T2D-associated SNPs are non-coding and map to regulatory elements across multiple tissues. To identify gene regulatory mechanisms, we mapped physical chromatin interactions in a human-derived pancreatic  $\beta$ -cell line (i.e. EndoC- $\beta$ H1) using next-generation capture-C (NGCC). This method enriches 3C libraries for loci of interest by a factor of 100,000 and yields the highest attainable resolution of physical interaction maps. Moreover, given the importance of  $\beta$ -cells to systemic glucose metabolism, we captured enhancer variants with likely regulatory functions in this cell type. To prioritise these variants, we performed a functional GWAS using summary statistics from the latest DIAGRAM GWAS meta-analysis (900K individuals of European ancestry) and a panel of islet epigenetic chromatin states derived from DNA methylation, open chromatin (ATAC-seq), and histone post-translational modifications (ChIP-seq). We integrated fine-mapping results from this analysis with maps of open chromatin in EndoC- $\beta$ H1 cells, islet eQTLs, and loci previously implicated in  $\beta$ -cell physiology to generate a set of nearly 300 “high priority” SNPs at 59 loci. We applied NGCC to capture enhancers encompassing these variants and identified ~600 significant chromatin interactions which were highly enriched for islet transcription factor (e.g. PDX1, FOXA2, NKX6.2) binding sites (p-values < 0.001). Moreover, 52% of all captured sites physically interact with transcription start sites; some of which corroborate GWAS signals colocalizing with islet eQTL associations (e.g. ADCY5, ZMIZ1). Furthermore, these interactions support putative novel T2D genes regulated by risk-associated variants such as KCNK16. These 3D chromatin interaction maps along with eQTL and epigenetic annotations implicate effector transcripts that promote T2D risk in  $\beta$ -cells and prioritise genes that may serve as therapeutic candidates.

## PgmNr 3218: A map of inflammatory-stimulated chromatin in human pancreatic islets reveals novel mechanisms of type 1 and 2 diabetes risk variants.

### Authors:

A. Aylward <sup>1,2</sup>; M. Okino <sup>2</sup>; P. Benaglio <sup>2</sup>; J. Chiou <sup>2,3</sup>; E. Beebe <sup>2</sup>; S. Huang <sup>2</sup>; J. Padilla <sup>2</sup>; K.J. Gaulton <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Bioinformatics & Systems Biology, University of California San Diego, San Diego, CA.; 2) Pediatrics, University of California San Diego, San Diego, CA.; 3) Biomedical Science, University of California San Diego, San Diego, CA.

---

The majority of genetic risk loci for type 1 and type 2 diabetes (T1D, T2D) occupy non-coding DNA and affect gene regulation in disease-relevant tissues such as pancreatic islets. Islets occupy a complex and dynamic microenvironment, and the responses of islets to environmental changes during diabetes progression are important mediators of disease etiology. Currently available epigenomic maps, however, do not effectively capture changes in the chromatin landscape of islets upon exposure to disease-relevant environmental conditions. In this study we used ATAC-seq and RNA-seq to map chromatin accessibility and gene expression from a total of 7 primary human islet samples exposed to pro- and anti-inflammatory treatment conditions relevant to diabetes pathophysiology, including glucocorticoids (dexamethasone) and cytokines (IL-1 $\beta$ , IFN $\gamma$ , TNF- $\alpha$ ) in multiple combinations and concentrations. We identified highly dose- and stimulus-dependent changes in chromatin accessibility in stimulated compared to untreated islets that mapped to 4,193 stimulus-responsive sites, transcription factor motifs driving these changes such as IRF/STAT for cytokines and glucocorticoid receptor (GR/NR3C1) for dexamethasone, and target genes and molecular pathways affected by stimulus-responsive chromatin. A majority of stimulus-responsive sites (3,643/87%) had evidence for differential chromatin accessibility on a gradient across glucocorticoid-treated, untreated and cytokine-treated islets. We observed enrichment of genetic variants influencing diabetes risk within stimulus-responsive islet chromatin, and identified fine-mapped diabetes risk variants in stimulus-responsive chromatin including at the *SIX3* and *NFATC2* loci. At *NFATC2*, rs3787186 is likely causal for both T1D and T2D risk, mapped in an islet chromatin site with a gradient of activity across stimuli, had stimulus-dependent allelic effects on islet chromatin in allelic imbalance and gene reporter assays, altered a binding motif for GR and other nuclear hormone receptors, was linked to the *NFATC2* promoter using single cell co-accessibility, and affected *NFATC2* expression in islet eQTLs. Our results provide an expanded catalog of chromatin sites in pancreatic islets upon exposure to diverse, disease-relevant environmental conditions through which we identify a greater breadth of genetic mechanisms and genes underlying diabetes risk.

## PgmNr 3219: Dysregulation of circulating microRNAs in Chinese type 2 diabetes patients with obesity.

### Authors:

C. He <sup>1,2</sup>; P. Chen <sup>1,3,4</sup>; X. Pan <sup>1,2</sup>; L. Chen <sup>5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics, College of Basic Medical Sciences, Jilin University, Changchun, Jilin Province, China; 2) Department of Molecular Biology, College of Basic Medical Sciences, Jilin University, Changchun, China; 3) The Key Laboratory of Pathobiology, Ministry of Education, College of Basic Medical Sciences, Jilin University, Changchun, Jilin Province, China; 4) Department of Pathology, College of Basic Medical Sciences, Jilin University, Changchun, Jilin Province, China; 5) College of Clinical Medicine, Jilin University, Changchun, Jilin Province, China

**Background:** The incidence of type 2 diabetes (T2D) has increased dramatically in recent decades. Differential expression of microRNAs (miRNAs) in adipose tissues from patients with obesity was reported. While obesity is closely related to T2D, the role of miRNAs in obesity under the physiological condition of T2D remained unclear. In this study, we assumed that circulating miRNAs are markers to measure tissue-specific miRNA expression and sought to identify miRNAs that are differentially expressed in the serum samples from T2D patients with obesity as compared to those without obesity.

**Methods:** This study included 58 Chinese T2D patients. We grouped patients by the standard of China National Nutrition and Health Survey data in 2002, body mass index (BMI)  $< 24$  (kg/m<sup>2</sup>) for non-obese and  $\geq 28$  (kg/m<sup>2</sup>) for obesity. We conducted a two-stage study, with T2D patients without retinopathy as discovery stage (12 obese v.s 20 non-obese) and T2D patients with retinopathy as replication stage (15 obese v.s 11 non-obese). Differential expression was quantified by next-generation sequencing and evaluated by DESeq2 adjusted by age and gender. Finally, we investigated the functional enrichment of target genes of the differentially expressed miRNAs.

**Results:** Three miRNAs were differentially expressed in both two stages ( $p$ -value  $< 0.05$ , LFC  $\geq 2$  or LFC  $\leq -2$ ), including miR-3909, miR-3960, and miR-3656. Their target genes were enriched in calcium signaling pathway, GABAergic synapse, type 2 diabetes mellitus, glycerolipid metabolism and adipocytokine signaling pathway ( $p$ -value  $< 0.05$ ).

**Conclusion:** Our results provided the evidence of three miRNAs as candidates for future mechanism studies. The indicated pathways may shed light on the drug development and clinical practice of weight management in T2D patients.

## PgmNr 3220: Epigenome-wide association study of change in blood metabolite levels from young- to middle adulthood in the Northern Finland Birth Cohort 1966.

### Authors:

H. Draisma<sup>1,2</sup>; I. Pupko<sup>3</sup>; L. Zudina<sup>4</sup>; V. Goffert<sup>5</sup>; M.D. Anasanti<sup>2</sup>; M. Wielscher<sup>6</sup>; M. Ala-Korpela<sup>7,8</sup>; S. Sebert<sup>2,8</sup>; M.-R. Järvelin<sup>6,8</sup>; N. Pervjakova<sup>9</sup>; M. Kaakinen<sup>1,2,10</sup>; I. Prokopenko<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Surrey, Guildford, Surrey, United Kingdom; 2) Section of Genomics of Common Disease, Department of Medicine, Imperial College London, London, United Kingdom; 3) Department of Medicine, Imperial College London, London, United Kingdom; 4) Department of Life Sciences, Imperial College London, London, United Kingdom; 5) University of Tartu, Tartu, Estonia; 6) Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom; 7) Baker Heart and Diabetes Institute, Melbourne, Australia; 8) Center for Life Course Health Research, University of Oulu, Oulu, Finland; 9) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia; 10) Centre for Pharmacology and Therapeutics, Department of Medicine, Imperial College London, London, United Kingdom

---

Several associations between metabolite levels and DNA methylation (DNAm) have been reported. However, the relationship between longitudinal changes in metabolite levels and differential DNAm is underexplored. We assessed associations between epigenome-wide blood DNAm and change in blood metabolomics-based variables. For 595 non-diabetic individuals from the Northern Finland Birth Cohort 1966 for whom nuclear magnetic resonance-based metabolomics data were available at both ages 31 (T1) and 46 (T2) as well as concurrent blood DNAm data at T2, we calculated for each of the 228 metabolomics-based variables the average change in level per year between T1 and T2. We used our methylSCOPA software, which enables both longitudinal and multi-phenotype epigenome-wide association studies (EWAS), for single-phenotype EWAS of change residuals – corrected for sex – for each metabolomic variable versus the degree of DNAm for 832,569 markers on the Illumina (San Diego, CA) MethylationEPIC BeadChip. We quality-controlled, residualized, and normalized the DNAm data, mapped genomic locations to CGCh37/hg19, and adopted  $1 \times 10^{-7} / 228 = 4.4 \times 10^{-10}$  as the Bonferroni-corrected  $P$ -value threshold to denote epigenome- and metabolome-wide statistical significance of association between an epigenetic mark and change over time in a particular metabolomic variable. We detected significant association between DNAm status at site cg16904859 (chr1:206,849,995;  $\beta=2.31$ ;  $SE=0.36$ ;  $P=3.2 \times 10^{-10}$ ) and change over time in the level of triglycerides in chylomicrons and extremely large very low-density lipoprotein particles (XXL-VLDL-TG). The cross-sectional association between this epigenetic mark and XXL-VLDL-TG levels at T2 was not significant ( $\beta=0.036$ ;  $SE=0.024$ ;  $P=0.13$ ). cg16904859 also showed epigenome-wide ( $P < 1 \times 10^{-7}$ ) but not epigenome- and metabolome-wide significance of association with the concentrations of total lipids, phospholipids, total cholesterol and free cholesterol in chylomicrons and extremely large very low-density lipoprotein particles as well as with the concentration of these types of particles. In conclusion, using a novel powerful methylSCOPA approach, we demonstrated that longitudinal changes in blood metabolite levels are associated with DNAm.



## PgmNr 3221: Multi-stage genome-wide methylation analysis identifies two novel CpG sites associated with obesity in Korean.

### Authors:

S. Moon; I.-U. Koh; N.-H. Choi; J. Kong; B.-J. Kim

[View Session](#) [Add to Schedule](#)

**Affiliation:** Division of Genome Research, Korea National Institute of Health, Cheongju-si, Chungchungbuk-do, Korea, Republic of

Obesity is a major risk factor of chronic diseases including type 2 diabetes and cardiovascular disease. Recently, with advances in genome-wide association studies (GWASs), more than hundreds of genetic variants associated with obesity-related traits have been identified. However, even though the great success of GWASs, most of the risk factors of obesity still remains unexplained. DNA methylation is a key epigenetic mechanism that is suggested to be associated with obesity and related metabolic mechanisms. Because excessive fat accumulation is caused by a combination of multi-factorial contributions, alternative approaches to finding non-genetic factors can be warranted to explain missing factors.

Here, we performed an epigenome-wide analysis of multi-stages to identify DNA methylation site with significant differences between obese and lean subjects. Furthermore, we examined common methylation sites that are significant between whole blood and adipose tissue. Subsequently, we examined differentially methylated site affecting the expression of the nearby gene. To this end, genome-wide methylation levels were measured in whole blood using the Illumina HumanMethylationEPIC array (EPIC) on 450 participants (phase I) with respective 377 (phase II), and 75 participants (phase III) from two independent cohorts. Furthermore, additional 75 pair of the EPIC array and Illumina HT-12 Gene Expression BeadChip data obtained from adipose tissue of the same participants as those of phase III was used for correlation analysis between CpG and gene expression and between tissue and whole blood.

From the results, 124 differentially methylated obesity CpG sites were reached the statistical significance with FDR <0.05. Particularly, 5 CpG sites of 3 genes, ABCG1, CPT1A and TXNIP related to the lipid metabolism, were also replicated two independent replication stages. Of 5 CpG sites, 2 sites were newly identified. Of the 124 CpG sites, differences in DNA methylation in 55 CpG sites were correlated with expression of 49 genes in fat tissue of phase III ( $|\text{spearman's correlation}| > 0.2$ ; 28 positive and 27 negative correlations). Among those, 19 genes of 22 CpG sites showed significant differences in phase I blood DNA methylation.

In conclusion, we identified two novel CpG sites annotated to CPT1A and TXNIP, protein-coding genes involved in lipid metabolism and homeostasis, providing the functional importance of obesity-specific DNA methylation in the Korean population.

## PgmNr 3222: Subcutaneous adipose transcriptomes reveal an environmentally responsive novel master *trans* regulator, TBX15, controlling a co-expression network with 39 obesity GWAS genes.

### Authors:

D.Z. Pan<sup>1,2</sup>; Z. Miao<sup>1,2</sup>; K.M. Garske<sup>1</sup>; S. Rajkumar<sup>1,3</sup>; A. Ko<sup>4</sup>; D. Kaminska<sup>1,5</sup>; J.S. Sinsheimer<sup>1,6</sup>; K.L. Mohlke<sup>7</sup>; M. Laakso<sup>8</sup>; J. Pihlajamäki<sup>5,9</sup>; P. Pajukanta<sup>1,2,10</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA.; 2) Bioinformatics IDP, UCLA, Los Angeles, California.; 3) Computational and Systems Biology Interdepartmental Program, UCLA, Los Angeles, USA.; 4) Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA.; 5) Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland.; 6) Department of Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles, USA.; 7) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA.; 8) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland.; 9) Clinical Nutrition and Obesity Center, Kuopio University Hospital, Kuopio, Finland.; 10) Institute for Precision Health at UCLA, Los Angeles, USA.

Gene-environment interactions contributing to the global obesity epidemic remain elusive. We hypothesized that transcription factors (TFs) context-specifically respond to environmental cues and alter co-expression networks in *trans* in obesogenic tissues. Using Weighted Gene Co-expression Network Analysis (WGCNA), we first discovered a waist-hip-ratio adjusted for BMI (WHRadjBMI)-associated co-expression network utilizing subcutaneous adipose RNA-sequence data (n=335) from the Finnish METSIM cohort. The network contains 39 obesity GWAS genes, including adipose transcriptional regulators, *PPARA*, *PPARG*, and *TBX15*. The *cis* regions ( $\pm$ 500kb from the TSS) of the network genes are enriched for SNPs explaining heritability of WHRadjBMI (enrichment=1.58,  $p=0.011$ , LD Score Regression). To test the significance of this network in response to weight loss, we RNA-sequenced subcutaneous adipose biopsies from morbidly obese Finns of the KOBS cohort both at the time of (n=254) and one year after bariatric surgery (n=168). Of the 347 genes in the WHRadjBMI network, 111 (32%) are differentially expressed (DE) in KOBS, including two TFs, *TBX15* and *HOMEZ*. Notably, *TBX15* has a context-specific *cis*-eQTL, rs1779445, exclusive to the extreme obesity state ( $p=6.7 \times 10^{-6}$ ), which is also a WHRadjBMI GWAS variant in the GIANT cohort (n=224,459). In line with this, previous studies have shown that *TBX15* inhibits adipocyte differentiation and reduces triglyceride accumulation. The context-specific *TBX15 cis*-eQTL rs1779445 is a *trans*-eQTL (nominal  $p<0.05$ ) for 27 (24%) of the 111 KOBS DE genes in the WHRadjBMI network, all of which also have a *TBX15* motif in their promoter (1kb upstream of TSS). Moreover, rs1779445 is associated with the expression principal component 1 (PC1) of these 27 genes in both METSIM and KOBS ( $p<0.015$ ). Using mediation analysis, we found that the rs1779445 effect on PC1 of the 27 KOBS DE genes is mediated through *TBX15* expression (32% total effect mediated,  $p=0.015$ ), context-specifically in the extreme obesity state. Finally, the adipose expression of two key drivers among the 27 genes, *CORO1C* and *MLX*, together with *TBX15* and rs1779445 explain a substantial amount (10.87%) of the variance in WHRadjBMI in METSIM. Taken together, we discovered *TBX15* as a novel adipose master *trans*



regulator that context-specifically mediates the effect of a WHRadjBMI GWAS variant, rs1779445, on a WHRadjBMI-associated co-expression network containing 39 obesity GWAS genes.

## **PgmNr 3223: Systematic characterization of the role of alternative splicing in adipocyte differentiation.**

**Authors:**

K. Farris; D. Sobreira; A. Joslin; M. Nobrega

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Human Genetics, University of Chicago, Chicago, IL

---

Alternative splicing is a key source of functional complexity in human tissues and is known to play an important role in tissue identity and development. A genome-wide analysis of alternative splicing across human adipocyte differentiation could therefore provide novel insights into both the regulation of adipose tissue development and mechanisms of disease. However, splicing is also a noisy process, which can result in many low abundance isoforms with little or no functional relevance. Therefore, while alternative splicing can have important biological consequences, there is also a need to determine which splicing events are biologically relevant. Here, we used the identification of differential splicing events across adipocyte differentiation to systematically characterize the role that alternative splicing plays in the process and to identify likely regulated splicing events that may be of particular biological importance. We used three replicates of RNA-seq data from four time points spanning from preadipocytes to mature adipocytes to identify differential splicing events between pairs of time points using LeafCutter. We identified 281 significantly differentially spliced clusters, and investigated the regulation of these clusters through the identification of enriched sequence motifs within the clusters that may be associated with splicing regulators. These results provide insights into the role of regulated splicing in adipocyte differentiation, as well as providing a resource to investigate the disruption of alternative splicing during adipocyte development as a potential mechanism contributing to obesity.

## PgmNr 3224: Pro-inflammatory cytokines induce circRNA over-expression in human pancreatic $\beta$ -cells.

### Authors:

S. Kaur <sup>1</sup>; A.H. Mirza <sup>1</sup>; C. Frørup <sup>1</sup>; T. Fløyel <sup>1</sup>; M.L. Colli <sup>2</sup>; D.L. Eizirik <sup>2</sup>; F. Pociot <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Steno Diabetes Center Copenhagen, Gentofte, Denmark; 2) ULB Center for Diabetes Research, Medical Faculty, Université Libre de Bruxelles (ULB), Brussels, Belgium; 3) Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

---

**Background:** Non-coding RNAs have emerged as essential players in the regulation of gene expression and their dysregulation contributes to various pathophysiological conditions. In the last decade, non-coding RNAs, including microRNAs and long non-coding RNAs have been implicated in pancreatic  $\beta$ -cell dysfunction and death in type 1 diabetes. Circular RNAs (circRNAs) are a newly discovered class of endogenous non-coding RNAs that are formed by non-canonical backsplicing events. In the present study, we profiled circRNAs in human  $\beta$ -cells exposed to pro-inflammatory cytokines to study their possible roles in  $\beta$ -cell dysfunction and type 1 diabetes pathophysiology.

**Methods:** Total RNA from human EndoC- $\beta$ H1 cells treated and non-treated with pro-inflammatory cytokines (INF $\gamma$ +IL-1 $\beta$ ) at 24 and 48h was extracted and circRNAs profiled using Arraystar Human circRNA Array V2. Before labeling and hybridization, total RNA was digested with Rnase R to remove linear RNAs and enrich the samples for circular RNAs. Quantile normalization of raw data and subsequent data processing was performed using limma package in R. Differentially expressed circRNAs were identified using a log2FC cutoff of 1 and adjusted p-value of 0.05. RT-qPCR was used to verify the differential expression of circRNAs. CircRNA/microRNA interactions were predicted using TargetScan & miRanda.

**Results:** A total of 74 and 50 circRNAs were differentially regulated by cytokines at 24 and 48h respectively. Most of the circRNAs were up-regulated (71/74 and 48/50 circRNAs) by cytokines at both time-points. RT-qPCR validation demonstrated that the expression levels of hsa\_circ\_0033184, hsa\_circ\_0000479, hsa\_circ\_0033191, hsa\_circ\_0002938, hsa\_circ\_0053955, hsa\_circ\_0053958 and hsa\_circ\_0002017 were consistent with the results from the microarray analysis. Highly up-regulated circRNAs from the *CRIM1* and *WARS* loci harbored multiple binding sites for miR-320 family (miR-320a, miR-320b, miR-320c and miR-320d). The miR-320 family is highly expressed in human pancreatic islets and one of their targets, p85, plays a critical role in cell growth by increasing Akt phosphorylation.

**Conclusion:** Based on our findings, we suggest that circRNAs play a role in the regulation of  $\beta$ -cell gene expression via modulation of host gene expression and miRNA levels. Further studies are ongoing to identify specific mechanisms by which circRNAs regulate  $\beta$ -cell function in human islets under inflammatory conditions.

## PgmNr 3225: Integrated genetic-epigenetic analysis identifies novel biological pathways for diabetic kidney disease.

### Authors:

Sheng <sup>1,2</sup>; C. Qiu <sup>1,2</sup>; J. Hsu <sup>4</sup>; C. Gluck <sup>1,2</sup>; J. He <sup>5</sup>; C. Hsu <sup>6</sup>; D. Sha <sup>4</sup>; M. Weir <sup>7</sup>; T. Isakova <sup>8</sup>; D. Raj <sup>9</sup>; R. Hernan <sup>10</sup>; H. Feldman <sup>4</sup>; R. Townsend <sup>3</sup>; H. Li <sup>4</sup>; K. Susztak <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Institute for Diabetes, Obesity and Metabolism, University of Pennsylvania, Philadelphia, PA, USA; 2) Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA; 3) University of Pennsylvania, Philadelphia, PA. Department of Medicine, Renal Electrolyte and Hypertension Division, University of Pennsylvania, Philadelphia, PA, USA; 4) Department of Epidemiology, Biostatistics and Informatics, University of Pennsylvania, Philadelphia, PA, USA; 5) Tulane University; 6) University of San Francisco; 7) university of maryland; 8) Northwestern University; 9) the Division of Kidney Diseases and Hypertension within the Department of Medicine at The George Washington University School of Medicine and Health Sciences in Washington, D.C; 10) Nephrology and Hypertension, Cleveland Clinic Main Campus

---

Poor metabolic control may induce epigenetic changes, which play an important role in the development and progression of diabetic kidney disease (DKD). Most prior studies, however, failed to integrate genotype information into the epigenome wide analysis studies (EWAS) to distinguish genetically and environmentally induced epigenetic changes.

We analyzed genome wide cytosine methylation changes in blood samples of 473 patients with diabetes from the Chronic Renal Insufficiency Cohort (CRIC) using the Illumina EPIC arrays. We adopted adjusted linear mixed effect models to identify methylation probes associated with kidney function. We have catalogued genotype-driven methylation changes by performing methylation quantitative trait (mQTL) analysis. Bayesian co-localization analysis was performed to identify causal genes and variations by integrating GWAS, EWAS, mQTL and eQTL datasets.

We found that methylation level of 1 probe significantly (FDR<0.05) associated with baseline kidney function. Methylation could be an important intermediate to mediate the effect of underlying sequence variations; as genetic variants could explain multiple EWAS associations and most GWAS signals resulted in methylation changes. Integration of genotype, methylation and gene expression data identified novel genes, associated with DKD pathogenesis and indicated an importance of underlying sequence variation.

## PgmNr 3226: Multiple testing correction for longitudinal epigenome-wide association studies (EWAS).

### Authors:

K. Leap<sup>1,2</sup>; S. Lent<sup>3</sup>; M.F. Hivert<sup>4</sup>; A. Cardenas<sup>4,5</sup>; K. Kleinman<sup>2</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Human Genetics, University of California Los Angeles, Los Angeles, CA; 2) Biostatistics and Epidemiology, University of Massachusetts Amherst, Amherst, MA; 3) Biostatistics, Boston University, Boston, MA; 4) Population Medicine, Harvard Medical School, Harvard Pilgrim Health Care Institute, Boston, MA; 5) Environmental Health Sciences, University of California Berkeley, Berkeley, CA

---

**Motivation:** Aberrant DNA methylation (DNAm) has been associated with adverse health outcomes and common diseases; however, unlike genetic markers that are fixed over the lifetime, DNAm can drift over time, varies with age, and is responsive to environmental changes. Longitudinal changes in DNAm are of broad interest, but given the ever increasing number of DNAm sites surveyed using microarrays, measuring them at different time points introduces multiple testing problems beyond those that exist in a static genetic context. We use as a motivating example the longitudinal effect of maternal glucose exposure in utero on DNAm levels in blood throughout childhood.

**Methods:** We simulated datasets of the same size as our motivating epigenome-wide association study according to four underlying longitudinal associations: 1) a persistent effect of the exposure over time, 2) an attenuated effect over time, 3) an effect that does not persist, and 4) no effect. We compared two methods of testing for persistent effects of the exposure: linear mixed effects model for each site (LME) or pre-selecting sites with a correlation between maternal glucose and DNAm at the first time point and then fitting LME models (PS).

**Results:** Across 4000 simulations of each setting with 560 true associations, we found that both methods controlled the false discovery rate (empirical FDR = 0.05). However, pre-selection reduced the number of sites identified as significantly associated by an order of magnitude (LME: 338 persistent, 218 attenuated; PS: 82 persistent, 83 attenuated). Furthermore, in the case of an effect that did not persist, pre-selection found an average of 73 sites significantly associated, while fitting all sites identified an average of 1.

Analysis of maternal glucose associations with DNAm in offspring found three sites at the first time point in cord blood (cg01696984, cg17077845, cg16578923), but none were identified using LME models incorporating data from all three time points (cord blood, ages 3 and 7).

**Conclusion:** If the research question hinges on whether an association with exposure persists over time, pre-selecting with a correlation test will not answer this question. We recommend that pairwise tests with the time point of interest be conducted separately from the longitudinal analysis in order to find both persistent effects and effects that resolve with time. Our data suggest that the effect of maternal glucose on DNAm in cord blood does not persist into childhood.

## PgmNr 3227: Genome-wide DNA methylation analysis for diabetic kidney disease in a UK-ROI population.

### Authors:

L. Smyth<sup>1</sup>; N. Sandholm<sup>2,3,4</sup>; C. Wooster<sup>1</sup>; J. Kilner<sup>1</sup>; C. Forsblom<sup>2,3,4</sup>; P.H. Groop<sup>2,3,4,5</sup>; A.P. Maxwell<sup>1,6</sup>; A.J. McKnight<sup>1</sup>; on behalf of the GENIE consortium.

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom.; 2) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Helsinki, Finland.; 3) Abdominal Center Nephrology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.; 4) Diabetes & Obesity Research Program, Research Program's Unit, University of Helsinki, Helsinki, Finland.; 5) Department of Diabetes, Central Clinical School, Monash University, Melbourne, Victoria, Australia.; 6) Regional Nephrology Unit, Belfast City Hospital, Belfast, Northern Ireland, United Kingdom.

---

Diabetic kidney disease (DKD), as characterised by progressive development of proteinuria and loss of renal function, is a common complication of diabetes. DKD is the most common cause of end-stage renal disease, wherein individuals require renal replacement therapy. Increasing evidence suggests that epigenetic alterations, including DNA methylation, are involved in the development and progression of DKD. This investigation compared methylation profiles between individuals from the British Isles with type 1 diabetes (T1D) and DKD to individuals with T1D and no evidence of renal failure, to identify potential methylation-based biomarkers of DKD.

Using the Zymo EZ DNA methylation kit to bisulphite treat the blood-derived DNA and the Infinium HD Methylation Assay, MethylationEPIC BeadChips (Illumina), the methylation status of >850,000 CpG sites on gene bodies, promoters and CpG islands were determined for 492 individuals with T1D. Of these, 241 individuals had DKD and 251 controls had no evidence of renal disease. Cases and controls were matched carefully for ethnicity, sex, age and duration of diabetes. DNA obtained from each individual was treated consistently, with standard quality control and bioinformatic analyses conducted.

Methylation data was analysed using Partek Genomics Suite v7.0. A total of 486 CpG sites had significantly different levels of methylation in cases compared with controls ( $p \leq 10^{-8}$ ). The top-ranked gene, *FKBP5* ( $p = 1.23 \times 10^{-23}$ ) contained several significantly associated CpG sites and has previously been linked with T1D, ageing and CKD. Additional significant genes included *BCL2* and *PSD3*, and both have been previously linked to diabetes. High concordance ( $r^2 = 0.994$ ) between duplicate samples ( $n = 7$ ) was observed. Pathway analysis revealed Notch signalling with the greatest enrichment score (16.96) where  $p = 4.29 \times 10^{-8}$ .

We previously reported results for methylation sites differentially regulated in individuals with T1D with and without renal disease using Illumina's HumanMethylation450K BeadChip and HumanMethylation27K BeadChip arrays. The *FKBP5* gene demonstrated different methylation levels in the 27K, 450K, and now EPIC arrays with a consistent direction of methylation change. This research confirmed earlier loci for DKD, adds new information on previously unexplored regions of the

methylome, and identified new associations demonstrating blood-derived methylation signatures may serve as minimally invasive biomarkers of DKD.

## PgmNr 3228: Whole-blood transcriptome analysis in longitudinal change in BMI.

### Authors:

M. Kuokkanen <sup>1</sup>; J. Kettunen <sup>2</sup>; M. Perola <sup>1</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Dept. of Public Health Solutions, National Institute for Health and Welfare, Helsinki, Finland; 2) Computational Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, Finland

---

Body mass index (BMI) is the most commonly used measure to evaluate body fat content and further obesity. BMI is associated with numerous comorbidities including type 2 diabetes, fatty liver disease and cardiovascular disease. We performed a whole-blood transcriptome analysis to detect expression changes in two time points in relation changes in BMI. We used population based subsamples of the Finnish Dietary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) (n = 518 in 2007, n = 297 in both 2007 and 2014). At first, we performed a whole transcriptome association analysis for BMI at baseline using 518 samples. We observed that 186 transcripts were significantly (The Benjamini and Hochberg false discovery rate, FDR < 0.01) associated with BMI. We continued analysis choosing individuals (297) who had follow up data from BMI and transcriptome measurements in 2014. We observed that 19 transcripts of those 186 associated with the alteration of BMI. Those 19 transcripts contained among other things *PLEK2*, *RANBP10* and *COX7C* and *CA1* genes involving PI3K/AKT signaling, agents acting on the renin-angiotensin system, AMPK enzyme complex pathways, O<sub>2</sub>/CO<sub>2</sub> exchange in erythrocytes including metabolism, respectively. Our results provide novel insights which genes is actually regulated when change in BMI is observed.



## PgmNr 3229: Identification of *CFTR* *cis*-regulatory variants.

### Authors:

M. Collobert <sup>1</sup>; K. Rouault <sup>1,2</sup>; C. L'Hostis <sup>1,3</sup>; M.P. Audrézet <sup>1,2</sup>; C. Férec <sup>1,2</sup>; S. Moisan <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France; 2) Laboratoire de Génétique Moléculaire et d'Histocompatibilité, CHRU de Brest, Bretagne, Brest, France; 3) Association Gaétan Saleun, Brest, France

---

Although, more than 2000 mutations have been discovered in *Cystic Fibrosis Transmembrane conductance Regulator (CFTR)* gene, some patients with cystic fibrosis or *CFTR*-related disorders (*CFTR*-RD) have incomplete genotypes or present extreme phenotypes. Development of chromatin conformation study techniques has identified several long-range regulatory elements as involved in this control expression. The objective of this project is to study the involvement of 'cis-ruption', that is the dysfunction of a *cis*-regulatory element, in cystic fibrosis and *CFTR*-RD (*Congenital Bilateral Absence of Vas Deferens*, CBAVD).

In a homogeneous group of 25 CBAVD patients carrying only one F508del mutation, 17 *cis*-regulatory regions of *CFTR* gene were sequenced. By comparing to European population, some variants display a frequency significantly different. In particular, one variant located in *cis*-regulatory region of intron 21 encompassing an important transcription factor binding site HNF1a, is 40 times more frequent in this group. Enhancer tests are realised to measure the effect of the intron 21 region on the activity of *CFTR* promoter in intestinal and airway cells. By combining the enhancer of intron 21 and the enhancer of intron 11 (strong enhancer described in intestinal cells), a strong cooperative effect is observed on the *CFTR* promoter activity in intestinal cells. These two enhancers have common transcription factors binding sites. The insertion of the variant of interest into intron 21 showed a strong decrease of *CFTR* promoter activity in intestinal cells. This variant inhibits the cooperative effect between enhancers of intron 11 and 21, and could affect the recruitment of regulatory elements, such as HNF1a or EP300.

In conclusion, 6 potential regulatory variants have been identified in CBAVD patients carrying a F508del mutation. In particular, one variant strongly reduces *CFTR* promoter activity the cooperative effect between enhancers of intron 11 and 21 and could affect the recruitment of regulatory elements in intestinal cells.

## **PgmNr 3230: Encode and single cell transcriptomics analysis of the developing pituitary gland: Insight into cell fate specification.**

### **Authors:**

A.Z. Daly; LYM. Cheung; M.L. Brinkmeier; L. Dudley; K.P. Toolan; S.A. Camper

[View Session](#) [Add to Schedule](#)

**Affiliation:** Human Genetics, Michigan, Ann Arbor, Michigan.

---

The ENCODE project has discovered regulatory elements in more than 250 cell lines and 150 tissues. This ambitious project integrates histone modification patterns, DNA accessibility, DNA methylation, transcription factor binding and RNA expression data to identify regulatory elements. This information is critical for understanding development and disease pathogenesis. The pituitary gland has not yet been analyzed by ENCODE, and it is essential for producing hormones that regulate growth, homeostasis, and reproduction. We have concentrated on mapping elements that emerge as pituitary progenitors develop into thyrotropes, the cells that secrete thyroid stimulating hormone (TSH), which regulates thyroid gland function. These cells represent only 2-5% of the cells in the adult pituitary gland, which presents a challenge for analysis. We used two immortalized cell lines that represent pituitary progenitors and TSH-expressing thyrotropes. To identify the set of transcription factors, cis-regulatory elements, and epigenetic marks that differ between these lines, we used the Assay for Transposase-Accessible Chromatin with sequencing (ATAC-seq) and Cleavage Under Target and Release Using Nuclease (CUT&RUN) for histone marks and binding of the lineage-specific transcription factor POU1F1. We identified putative regulatory elements for the hormone subunit genes of TSH and critical thyrotrope-signature genes. We validated them by transient cell transfection and confirmed a subset in transgenic mice. Motif enrichment analysis implicated several consequential transcription factors in thyrotrope function. To relate these findings to cells in intact pituitary glands, we collected pituitaries from newborn mice (postnatal day 7) with genetically-marked thyrotropes. We used fluorescence activated cell sorting to enrich for thyrotropes and conducted single-cell transcriptomics using the 10X genomics platform. This analysis revealed transcription factors that were concordant with the cell line analyses, suggesting their role in thyrotrope maintenance, and factors specific to the cell line analyses, suggesting their role in thyrotrope development. This data extends the ENCODE analysis to an organ that is critical for growth and metabolism, and it could be valuable for understanding genetic causes of pituitary dysfunction.

## PgmNr 3231: Genome-wide identification of epigenetic changes associated with 17 $\alpha$ -ethynylestradiol (EE2) exposure in the fathead minnow.

### Authors:

W. Huang <sup>1</sup>; D. Bencic <sup>2</sup>; R. Flick <sup>2</sup>; J. Fetke <sup>2</sup>; A. Biales <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) US Environmental Protection Agency, Research Triangle Park, NC 27709; 2) US Environmental Protection Agency, Cincinnati, OH, 45268

---

17 $\alpha$ -ethynylestradiol (EE2) is widely used estrogenic chemical present almost ubiquitous in aquatic environments throughout the United States and some other countries. Acting as an endocrine disrupting compound, EE2 can disrupt hormonal homeostasis, leading to developmental disorders, cancer and other diseases. Exposure to EE2 is known to induce expression of vitellogenin, a precursor protein of egg yolk normally only expressed in female fish, in male fish. However, the underlying epigenomic changes associated with vitellogenin induction and phenotypic changes (e.g. feminization), are not well understood. This study was designed to gain insights into such underlying epigenetic regulation mechanisms by evaluating genome-wide DNA methylation changes in CpG sites using the reduced representation bisulfite sequencing (RRBS). Taking advantage of the new and highly contiguous genome reference assembled in house, we used the fathead minnow as the model organism to identify CpG methylation changes before and after exposure to EE2. In our experiment, two groups of male fathead minnow fish, each with 16, were exposed 2.5ng/L and 10ng/L EE2 for 48 hours, respectively. In addition, another two groups not exposed to EE2 were used as control groups: one male group for negative control and the other female group for positive control. We then obtained RRBS data of both liver and brain tissues of all individuals. We assessed and compared CpG methylation changes immediately after EE2 exposure in both male liver and brain tissues. We also compared methylation CpG patterns after 7-day and 14-day depuration of EE2 to find if these methylation changes are temporary or potentially long-lasting. This was done separately for liver and brain tissues. Overall, we found that a limited number of CpG regions were subjected to significant methylation changes after 48-hr EE2 exposure in male liver tissue, and the number of affected CpG sites and their methylation level changes were even smaller in male brain tissue. Such epigenetic changes also exhibit dose-dependent effects with larger changes associated with a higher exposure dose. Our initial results show that CpG-sites methylation were changed quickly after EE2 exposure and stayed for a long time after depuration of EE2, indicating EE2 effects could be potentially long lasting. We are getting more detailed results including specific genes/regions associated with methylation changes, which will also be presented.

## PgmNr 3232: DNA methylation associated with kidney function identifies shared and ethnic-specific loci.

### Authors:

A. Batorsky<sup>1</sup>; M.K. Lee<sup>5</sup>; S.J. London<sup>5</sup>; A.J. Marron<sup>2</sup>; E.A. Whitsel<sup>2</sup>; N. Franceschini<sup>2</sup>; C.E. Breeze<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biostatistics, UNC Chapel Hill, Chapel Hill, North Carolina.; 2) Epidemiology, UNC Chapel Hill, Chapel Hill, North Carolina.; 3) Altius Institute for Biomedical Sciences, Seattle, USA; 4) UCL Cancer Institute, University College London, UK; 5) Epidemiology Branch, National Inst of Environmental Health Sciences, National Insts of Health, Dept of Health and Human Services, Research Triangle Park, NC

---

Epigenetic alterations such as changes in DNA methylation (DNAm) influence gene expression and regulation, and may contribute to chronic kidney disease (CKD) risk. CKD, defined by chronically low levels of estimated glomerular filtration rate (eGFR), is more common among individuals of Hispanic and African ancestry in the U.S, though the reasons for these differences are unknown. We use data from 2705 participants (833 African, 433 Hispanic, 1439 European ancestry) from the multi-ethnic Women's Health Initiative to examine associations between kidney function (eGFR) and DNAm patterns in whole blood. DNAm beta values obtained using Illumina 450K beadchip arrays were BMIQ-normalized and combat-adjusted for technical covariates. Epigenome-wide association studies (EWAS) were run in ethnic-stratified data using a linear model with robust standard errors adjusting for age, smoking, and study-specific variables, principal components of ancestry, and cell type composition. We meta-analyzed results across ethnicities using inverse-variance weighted methods. At a Bonferroni corrected significance level of  $p < 10^{-7}$  we found 28 differentially methylated positions (DMPs) across ethnicities, 18 of which are located in gene-coding regions and 10 of which are located in non-coding regions. Analyses within ethnicity for African, Hispanic, and European ancestry resulted in 14, 21 and 12 unique DMPs respectively. Enrichment analysis of the top 333 eGFR-associated DMPs across all ethnicities ( $p < 10^{-4}$ ) using eFORGE shows significant association of these sites with H3K4me1, an enhancer-associated mark, across a range of tissues, including mesenchymal embryonic stem cells and IPS cells. Further, these DMPs show a significant association with transcription factors (TF) motifs, *PAX1*, *HNF1B*, *POU2F1*, *SOX7*, *MAZ*, and *CDX2*, which are strongly implicated in developmental pathways including kidney. eFORGE-TF analysis shows that 12 of the top 333 DMPs overlap or are located close to *MAZ* motifs ( $q = 0.021$ ). In summary, DMPs found to be significantly associated with levels of eGFR, a measure of kidney function, show significant overlap with TF motifs involved in kidney development, and dysregulation of these sites may in turn mediate disease risk. Further examination of ethnicity-specific DMPs may elucidate differences in kidney disease risk across ancestries.

## PgmNr 3233: Assessment of DNA methylation on kidney function traits by developing and applying an EWAS workflow.

### Authors:

A. Teumer<sup>1,2</sup>; P. Schlosser<sup>3</sup>; F. Grundner-Culemann<sup>3</sup>; P. Matias Garcia<sup>4</sup>; A. Weihs<sup>1,5</sup>; A. Tin<sup>6</sup>; A. Köttgen<sup>3,6</sup>; on behalf of the CKDGen Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany; 2) DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany; 3) Institute of Genetic Epidemiology, Freiburg, Germany; 4) Helmholtz Zentrum München, Munich, Germany; 5) Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Germany; 6) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

The estimated glomerular filtration rate (eGFR) is the most common measure of kidney function and used to define and stage chronic kidney disease (CKD). Results of recent GWAS on kidney function implicated gene regulatory mechanisms. To get insight into the link between DNA methylation and traits related to kidney function, we conducted fixed-effects inverse-variance meta-analyses of epigenome-wide association study (EWAS) based on Illumina arrays with DNA methylation assessed from whole blood using samples of up to 35 cohorts from diverse ancestries.

The discovery stage of this project included up to 22,347 individuals for the EWAS on eGFR, 11,458 samples for the urinary albumin-to-creatinine ratio (UACR), and 12,479 individuals for serum urate levels. Associations that reached a 5% alpha level after Bonferroni correction for the number of sites tested per trait ( $p < 1.1E-07$ ) were taken forward for replication in up to 11,258 independent samples. Analyses were conducted using linear regression of the proportion of methylated DNA (beta values) on the trait adjusting for sex, age, white blood cell composition, technical factors, and known trait-specific correlates.

We developed and established a workflow for standardization and quality control of EWAS including automated harmonized generation of variables for pre-analysis data checks, an intra- and inter-study quality control pipeline of the EWAS results that systematically checks both methylation data and association results for consistency. By applying the BACON method, no inflation or bias of the test statistics was observed. In the replication stage, 72 out of the 117, one out of the 12, and 104 out of the 140 CpG sites for eGFR, UACR and serum urate, respectively, passed a  $FDR < 0.05$ . Nine of the 18 associations of a former EWAS on eGFR passed that FDR in the current EWAS. Novel sites passing the replication stage FDR for eGFR included *ZFHX3* and *WDR73*, two loci which were associated with kidney function in a recent GWAS of 1 million individuals. Subsequent analyses will include a detailed annotation of the associated sites, assessment of differentially methylated regions, and analysis of causality between DNA methylation and the associated traits.

These results will improve our understanding of the regulatory mechanisms of kidney function traits by highlighting novel pathways on the epigenetic level, and thus open new avenues on prediction and

potential treatment of CKD.

## PgmNr 3234: Genome-wide DNA methylation profiling identified genes involved in COPD severity.

### Authors:

S. Baek <sup>1</sup>; W. Kim <sup>2</sup>; K. Kim <sup>1</sup>; W. Chang <sup>1</sup>; J. Kim <sup>1</sup>; K. Yoo <sup>3</sup>; J. Lee <sup>1</sup>; D. Yoon <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Allergy and Chronic Respiratory Disease, Center for Biomedical Sciences, National Institute of Health, Korea Centers for Disease Control and Prevention, Osong, Republic of Korea; 2) Department of Internal Medicine and Environmental Health Center, Kangwon National University Hospital, School of Medicine, Kangwon National University, Chuncheon, Republic of Korea; 3) Division of Pulmonary, Allergy and Critical Care Medicine, Department of Internal Medicine, Konkuk University School of Medicine, Seoul, Republic of Korea

---

Chronic Obstructive Pulmonary Disease (COPD), a heterogeneous disease of the lungs, is the third leading cause of death worldwide. Multiple factors including genetic and epigenetic changes have been associated with the development of COPD. Numerous research groups have been studied altered DNA methylation profiling in COPD versus normal controls, but little is known about the epigenetic mechanisms underlying the DNA methylation changes according to the severity of COPD. In this study, we profiled genome-wide methylation differences between 50 patients with non-severe (GOLD 0 ~ II) COPD and 53 patients with severe (GOLD III, IV) COPD using Infinium MethylationEPIC BeadChip. Raw methylation data were quantile normalized with the R package (minfi). CpG sites with detection p value > 0.01 were regarded as failed and were assigned as missing values. The RnBeads was used for the analysis of methylation array data and DMP (Differentially Methylated Probe) identification. The Homer (v.4.7) software was used for the annotation of probes and assessment of the distribution of methylation probes across genomic features. In total, 425 significant DMPs were identified (Beta value Difference > 5%,  $P < 5.00 \times 10^{-2}$ ) in severe-COPD, and the majority of these DMPs were located in the promoter, 5'UTR, and exonic regions. There were 198 hyper-methylated and 227 hypo-methylated DMPs. Among them, a previously reported COPD genes, *TBX5*, *NOTCH4*, and *PTEN*, were observed. Previously hyper-methylation of *TBX5*, a member of a phylogenetically conserved family of genes involved in the regulation of development, was reported to be associated with COPD. Similarly, we observed hyper-methylation near *TBX5* in severe-COPD compared to those of non-severe group. We then used the DAVID database v6.8 to examine the functional annotation of genes. Genes with DMPs were enriched in KEGG pathways such as AMPK signaling pathway ( $P = 3.28 \times 10^{-4}$ ; *PPP2R1B*, *HNF4A*, *ACACA*, *CREB3L2*, *PFKP*, *PPP2R5E*, *PPP2R2C*, and *RPTOR*), and PI3K-Akt signaling pathway ( $P = 1.00 \times 10^{-3}$ ; *PPP2R1B*, *SGK2*, *TNXB*, *FGF7*, *CREB3L2*, *ITGB5*, *PPP2R5E*, *THBS2*, *PTEN*, *PPP2R2C*, *BRCA1*, and *RPTOR*). Taken together, our findings provide insights for understanding methylation patterns in severe COPD and novel predictive methylation markers for COPD severity.

## PgmNr 3235: HIV-induced epigenetic regulation: Analysis of methylation and RNA-seq data revealed apoptosis-related genes involved in HIV infected T lymphocyte.

### Authors:

Zeng <sup>1</sup>; S.K. Tsui <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Huazhong Agricultural University, Wu Han, China; 2) The Chinese University of Hong Kong, Hong Kong, China

---

It has been shown that both the methylation and gene expression play roles in HIV-host interplay. In this study, to further investigate the relationship between HIV infection and patterns of genome-wide methylation and gene expression, we performed MeDIP-seq and RNA-seq for five T cell lines and each cell line contained HIV+ and control samples. Within differentially expressed genes (DEGs) between HIV+ and control samples, the top enriched pathways included “cell apoptosis”, “primary immunodeficiency” and “inflammatory response”. Eighteen up-regulated and eight down-regulated DEGs were shared by four cell lines. Moreover, on average, 1,736 differentially methylated regions (DMRs) were found between HIV+ and control samples for each cell line. DMRs were significantly enriched in promoter and exon regions. Moreover, transcription factor binding motifs were also found to be significantly associated with methylation alterations, suggesting that DNA methylation might influence the binding of transcription factors with DNA during HIV infection. The top enriched pathways of DMR-associated genes were “ubiquitin-dependent protein catabolic process” and “GTPase activator activity”. Besides, 635 genes were also involved in the crosstalk of methylation and expression, out of which *ATP1B1*, *CAMK2D*, *GRIN2A*, *MAPK10*, *CACNA1C* and *F2R* were shared by more than two cell lines and enriched in “cAMP signaling pathway” and “cell apoptosis”, which affects HIV replication and infection. To validate our results, we selected the well-known apoptosis related genes for further analysis by qPCR, western blot and bisulfite sequencing. Results showed that all of them were successfully validated. Consistent with sequencing result, luciferase assay showed that the activity of *ATP1B1* and *GRIN2A* promoter were significantly decreased in the presence of Tat or Rev. Result of Annexin V assay revealed that knockdown of *CAMK2D* and *MAPK10* could reduce the apoptotic cells, implying their roles in apoptosis during HIV infection. This study provided comprehensive picture about changes at either transcriptional or epigenetic level for T cells in response to HIV infection. The identified candidate genes in this study may provide insights in treatment and diagnosis of AIDS in the future. This study broadens our understanding on the mechanism of HIV/AIDS and provides a resource about dynamic changes at the transcriptome and epigenetic levels during HIV infection.



## PgmNr 3236: A novel role for a long noncoding RNA in STAT signaling contributing to allergic asthma.

### Authors:

G.J. Kwon <sup>1,2</sup>; M. Yurieva <sup>1</sup>; A. Williams <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Jackson Laboratory, Farmington, CT; 2) Department of Genetics and Genome Sciences, University of Connecticut Health, Farmington, CT

---

Allergic asthma is characterized by airway hyperreactivity to a type 2 immune response; release of cytokines (IL-4, IL-13) lead to STAT6 phosphorylation & airway inflammation. Airway epithelial cells undergo distinct morphological changes following type 2 cytokine exposure, however the molecular events regulating these alterations are not well understood. Long noncoding RNAs (lncRNAs) are non-protein-coding transcripts with an increasingly reported diverse set of functions and cell type-specificity. Despite evidence of immune regulation, few lncRNAs have been functionally identified in the airway. In this study, we aimed to identify differentially expressed lncRNAs following IL-13 stimulation and assess their potential functional role in airway epithelial cells and type 2 immunity. Cultured primary human bronchial epithelial cells (HBECs) under air-liquid interface (ALI) conditions stimulated with vehicle or IL-13 from 8 individual donors were analyzed via RNA-seq. We identified a significantly induced lncRNA, *WFDC21P*, previously reported to prevent STAT3 dephosphorylation in dendritic cells (DC) and regulating consequent DC function. Additional exposure to IL-6 in HBECs, the prototypical cytokine to induce STAT3 phosphorylation, led to a reduction of *WFDC21P* expression back to basal levels. Similar to DCs, we found *WFDC21P* localization is cytoplasmic in HBECs via subcellular fractionation and FISH. Based on these preliminary observations, we posited *WFDC21P* regulates STAT phosphorylation and signaling in HBECs. To test this hypothesis, we used a bronchial epithelial cell line (HBEC3-KT) and dCas9 technology to overexpress *WFDC21P*. Flow cytometry assessed levels of phosphorylated (P)STAT. Overexpression of *WFDC21P* led to increased (P)STAT3 basally and with IL-13 stimulation. Interestingly, *WFDC21P* overexpression led to mildly decreased (P)STAT6 levels following IL-13 (30 minutes), a previously unreported finding. RNA-seq of HBEC3-KTs overexpressing *WFDC21P* following IL-13 stimulation (24 hours) identified robust changes in expression of known STAT6 and STAT3 target genes, such as *CCL26* and *SOCS1*. Current studies are focused on analyzing this RNA-seq data and characterizing *WFDC21P*'s effect on differentiated HBECs and airway cell-type specificity via single-cell RNA-seq. Our studies reveal a novel and cell type-specific role for *WFDC21P*, with further analysis leading to a better understanding of mechanisms governing allergic asthma and airway homeostasis.

## PgmNr 3237: Allele-specific expression changes dynamically during T cell activation in HLA and other autoimmune loci.

### Authors:

M. Gutierrez-Arcelus<sup>1,2,3,4</sup>; Y. Baglaenko<sup>1,2,3,4</sup>; J. Arora<sup>5</sup>; S. Hannes<sup>1,2,3,4</sup>; Y. Luo<sup>1,2,3,4</sup>; T. Amariuta<sup>1,2,3,4,6,7</sup>; N. Teslovich<sup>1,2,3</sup>; D.A. Rao<sup>3</sup>; J. Ermann<sup>3</sup>; A.H. Jonsson<sup>3</sup>; C. Navarrete<sup>8</sup>; P.K. Gregersen<sup>9</sup>; T. Esko<sup>2,10</sup>; M.B. Brenner<sup>3</sup>; S. Raychaudhuri<sup>1,2,3,4,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA; 2) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA.; 3) 3Department of Medicine, Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.; 4) 4Center for Data Sciences, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA.; 5) 5Max Planck Institute for Evolutionary Biology, Plön 24306, Germany.; 6) Graduate School of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA.; 7) Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA.; 8) Division of Infection and Immunity, University College London, UK.; 9) The Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, Northwell Health, Manhasset, NY, USA.; 10) Estonian Genome Center, Institute of Genomics, University of Tartu, Tartu, Estonia.; 11) 11Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

---

Genetic studies have revealed that susceptibility variants for autoimmune diseases are over-represented in regulatory elements of memory CD4<sup>+</sup> T cells. Understanding how genetic variation affects gene expression in T cells at key physiological states is essential for deciphering the mechanisms that lead to autoimmunity. Studies have analyzed genetic regulatory variation genome-wide in resting T cells and in a limited number of genes and cell states in activated T cells. However, an understanding of the dynamics of *cis* regulatory variation along multiple physiological states during T cell activation is lacking. Here we characterized the dynamics of genetic regulatory effects at eight time points during memory CD4<sup>+</sup> T cell activation with high depth RNA-seq in healthy individuals. We identified 186 genes with dynamic allele-specific expression, where the balance of alleles changes over time. These genes were four fold enriched in autoimmune loci. We found pervasive dynamic regulatory effects within six HLA genes, particularly for a major autoimmune risk gene, *HLA-DQB1*. Each *HLA-DQB1* allele had one of three distinct transcriptional regulatory programs. Using CRISPR/Cas9 genomic editing we demonstrated that a single nucleotide variant at the promoter is causal for T cell-specific control of *HLA-DQB1* expression. Our study in CD4<sup>+</sup> T cells shows that genetic variation in *cis* regulatory elements may affect gene expression in a lymphocyte activation status-dependent manner contributing to the inter-individual complexity of immune responses. It underscores the importance of studying the function of HLA class II genes in T cells, and ascertaining the evolutionary and disease risk aspects of regulatory variation of HLA genes independent of their protein-coding variation. Overall, we show a successful strategy that can be used in other biological systems for uncovering new context-specific regulation of genes with key roles in complex disease.

## PgmNr 3238: Defining regulatory variants for SLE susceptibility using an integrative post-GWAS functional genomic framework.

### Authors:

J.E. Molineros <sup>1</sup>; L.L. Looger <sup>2</sup>; C. Sun <sup>1</sup>; S.K. Nath <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Arthritis and Clinical Immunology, OMRF, Edmond, Oklahoma.; 2) Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA, USA

---

Multiple association studies have uncovered numerous genetic associations for systemic lupus erythematosus (SLE), an autoimmune disease with a strong genetic component. However the identity of the true causal variants linking their target gene(s) are largely unknown. The objectives of this study were to identify SLE susceptibility variants and their linked susceptibility genes by leveraging expression, epigenetic and 3D interaction data from five immune cell types.

We performed a comprehensive analysis using reported SLE loci, immune cell-specific gene expression and 3D chromatin interaction datasets to define predisposing variants and their target genes. We collated 346 reported genomewide significant SLE variants and 14,483 correlated ( $r^2 > 0.7$ ) variants to define 122 independent signals. We identified 91 deleterious coding SNPs (48 independent signals; CADD > 12.37), including SNPs at *PDHB* and *PXK* that significantly affected expression of *RPP14* in B cells, and further 10,132 eQTLs correlated with 879 genes in monocytes, neutrophils, CD4+T, CD8+T, and B cells. We identified eQTLs in gene promoters, and in enhancers (from promoter-enhancer interactions) for 204 genes. Overall, cell-specific target genes were overrepresented in 169 immune related pathways including Interferon gamma signaling ( $P = 7.09 \times 10^{-15}$ ). Additionally, 14 target genes from 12 independent signals exhibited aberrant isoform ratios, including *SYNGR1* where rs61616683 risk allele increased 7-fold expression of non-sense mediated decay transcript.

By leveraging immune cell expression, epigenetic information, and 3D interaction data, we identified functional SNPs and SLE genes at 77 independent SLE signals. The 90.3% of functional SNPs collocated with enhancers, followed by promoters, and by deleterious SNPs. Together, we identified 204 target susceptibility SLE genes from 69 loci expressed in immune cell types that were overrepresented in immune pathways.

## PgmNr 3239: Hypomorphic *TNIP1* expression in SLE: A cumulative effect of all risk variants on the *TNIP1* SLE risk haplotype.

### Authors:

S. Pasula <sup>1</sup>; K.L. Tessneer <sup>1</sup>; Y. Fu <sup>1</sup>; J. Gopalakrishnan <sup>1,2</sup>; P.M. Gaffney <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genes and Human Disease Research Program, Oklahoma Medical Research Foundation, Oklahoma city, Oklahoma.; 2) Department of Pathology, University of Oklahoma Health Science Center, Oklahoma City, OK, USA

### Background:

GWAS, large-scale replication, and fine-mapping studies identified *TNFAIP3 interacting protein 1* (*TNIP1*) as a strong SLE susceptibility locus shared across multiple ethnicities. Fine mapping identified two independent *TNIP1* SLE risk haplotypes associated with reduced *TNIP1* expression. In this study, we aimed to functionally characterize the SLE risk variant(s) driving the hypomorphic *TNIP1* expression in the SLE risk haplotypes. *TNIP1* rs10036748 risk allele was previously implicated in other autoimmune diseases and cancers. We investigated the potential role of rs10036748 and 10 additional SLE-associated *TNIP1* SNPs predicted to have functionality based on RegulomeDB and ENCODE data.

### Methods:

We evaluated allelic effects on nuclear factor binding and enhancer function using EMSAs and luciferase transactivation assays, respectively, in immune cells stimulated with or without PMA/Ionomycin (P/I). Affinity DNA pull-down assays, Western blotting (WB), and ChIP-qPCR were used to identify proteins with allele-specific binding to the rs10036748 SNP.

### Results:

Bioinformatic data identified EGR1, DEC1, and CREB1 as nuclear proteins predicted to bind rs10036748. EMSAs revealed reduced nuclear protein binding to the rs10036748 risk allele, relative to non-risk, in Jurkat, EBV B, and THP1 cells irrespective of P/I stimulation. Affinity DNA pull-down and ChIP-qPCR suggest that impaired binding was due, in part, to reduced DEC1 binding; EGR1 and CREB1 binding trended but were not significant. Despite reduced nuclear protein binding, the rs10036748 risk allele enhanced luciferase transactivation suggesting an enhancer, rather than a repressive effect on *TNIP1* expression.

To identify additional variants that may explain the repressive phenotype of the *TNIP1* haplotype, we strategically screened the remaining 10 SLE risk variants for functional relevance using EMSAs and luciferase transactivation assays. We observed a complex locus with several SLE risk variants contributing to often conflicting cell type-specific and/or stimuli-dependent changes in nuclear factor binding and enhancer activity.

### Conclusion:

Functional analyses of risk alleles in the *TNIP1* SLE risk haplotypes suggest a complex regulation at the *TNIP1* locus, exhibiting cell-type specificity, stimulation dependency, and allele-specific binding and enhancer activity. Decreased expression of the *TNIP1* risk haplotype may be the result of a cumulative effect of multiple functional risk variants.

## PgmNr 3240: Chronic lymphocytic leukemia (CLL) risk is mediated by multiple enhancer variants.

### Authors:

H. Yan <sup>1</sup>; S. Tian <sup>1</sup>; G. Kleinstern <sup>1</sup>; E. Braggio <sup>2</sup>; S. Slager <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Health Sciences Research, Mayo Clinic, Rochester, Minnesota.; 2) Department of Medicine, Mayo Clinic Arizona

---

Over 90% of the disease-associated single-nucleotide polymorphisms (SNPs) identified by genome-wide association studies (GWASs) are located in noncoding regions of the genome. It is increasingly recognized that noncoding risk variants play regulatory roles in mediating target gene expression, which often involves the alteration of local chromatin accessibility, histone modification, transcription factor (TF) binding affinity, and chromatin interaction. In CLL, we and others have identified 41 risk SNPs by GWAS. However, for the vast majority of the risk loci, the casual variants and the mechanism of casualty remain unknown. This study systematically analyzed 15 CLL risk loci located across 10 different chromosomes. To first map the gene regulatory regions, we uniformly processed chromatin accessibility data (n=24), ChIP-seq data for histone modifications (n=210), TFs and other DNA-binding proteins (n=99). These public epigenetic data were generated in lymphoblastoid cell lines, various blood cell types, lymphoma patient biopsies and cell line. We also sequenced 14 ChIP-seq libraries for H3K4me3, H3K4me1, H3K27ac and H3K27me3 in CLL. Comparing the 15 index SNPs from the CLL risk loci and their correlated variants ( $r^2 \geq 0.5$  in the EUR ethnic group) to the regulatory regions annotated above revealed that candidate casual variants were all located in enhancer regions, including eight in super-enhancers. Using the reference epigenomes generated by the Roadmap Epigenomics Project in other cell and tissue types, we found in nine of the cases that the H3K27ac signals were highly enriched in immune cells. At 13 of the risk loci, we identified strong evidence for the association of candidate casual variants with allelic imbalance in ChIP-seq signal, alteration of TF binding motifs, or both. Furthermore, enhancers covering four of the risk loci were found to be transcribed into noncoding enhancer RNAs (eRNAs) in primary B cells and lymphoblastoid cell line GM12878; eRNAs are proposed to play a role in modulating promoter-enhancer interactions. Finally, using public Hi-C, capture Hi-C and ChIA-PET data in lymphoblastoid cell lines and/or 17 primary blood cell types, we identified 15 target genes for all the risk loci, including seven that encode TFs. Notably, five of the genes are known to play a role in CLL and another seven genes in other lymphomas. Together, our analyses have illustrated multiple mechanisms underlying the pathogenic roles of risk variants in CLL.

## PgmNr 3241: Studying enhancer activity of hQTLs in human SLE by massively parallel reporter assay.

### Authors:

Y. Fu <sup>1</sup>; J. Kelly <sup>1</sup>; R. Pelikan <sup>1</sup>; P. Gaffney <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genes and Human Disease Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Pathology, University of Oklahoma School of Medicine, Oklahoma City, OK

**Introduction:** Genome-wide association studies have identified > 100 risk loci associated with the genetic susceptibility of systemic lupus erythematosus (SLE); however, identifying and functionally characterizing the causal variants involved in disease pathogenesis has been a laborious, inefficient, and resource-intensive task. Our lab previously integrated epigenetic and genotypic data from SLE patient-derived EBV-transformed B cells to identify and prioritize variants that induce allelic imbalance in the magnitude of histone post-translational modifications, i.e. histone quantitative trait loci (hQTLs). In this study, we applied a massively parallel reporter assay (MPRA) to confirm the allele-specific transcription regulation attributed to hQTLs.

**Methods:** We designed a 67,034 MPRA library that contains ~ 35,000 SNPs. We included: 1) the most significant hQTLs from our previous study; 2) non-hQTLs located on shared haplotypes; 3) autoimmune disease index SNPs; 4) eQTLs that interact with the hQTLs via 3D loops; 5) strong proxies of each SNP; 6) location controls; 7) random controls. 19bp indexes were added by PCR. The final MPRA library was transfected into 3 SLE EBV B cell lines, two replicates per cell line. After 24h culture, RNA was extracted to synthesize cDNA. Sequencing libraries were constructed by PCR with Illumina adapters and sequenced using Novaseq S2 flow cell.

**Results:** 61,446 (91.7%) oligos were tagged with at least one unique molecular index. 7933 (13.17%) SNPs demonstrated transcription regulatory activity. 280 (3.68%) SNPs demonstrated differential allelic expression between the reference and alternate alleles; the majority were hQTLs or SNPs looping to hQTLs. Importantly, hQTLs had the highest proportion (7.66%) of variants demonstrating allelic specificity relative to all other variants (0-4%). Among the most significant allelic specific hQTLs were known SLE risk genes: BLK, LRRRC16A, and IKZF1.

**Discussion:** We confirmed the transcriptional regulatory activity of hQTLs in SLE EBV B cells by MPRA. This study facilitates functional prioritization of hQTLs that demonstrate the most convincing and reproducible allele-specific influences on gene expression. Future study will focus on repeating the MPRA in cell lines of different lineage. We will also perform 3D chromatin topology assays to identify allele-dependent effects of hQTLs on chromatin regulatory networks to understand how hQTLs modify gene expression and SLE pathogenesis.



## PgmNr 3242: Altered DNA methylation profile in placentas from pregnant women with sickle cell disease.

### Authors:

G. Gil <sup>1</sup>; G. Ananina <sup>1</sup>; M. Maschietto <sup>2</sup>; S. Soares Lima <sup>3</sup>; S. Costa <sup>1</sup>; L. Baptista <sup>1</sup>; F. Costa <sup>1,4</sup>; M. Costa <sup>1,5</sup>; M. Melo <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas-UNICAMP, Campinas, SP, Brazil; 2) Boldrini Children's Center, Campinas, SP, Brazil; 3) National Cancer Institute (INCA), Rio de Janeiro, RJ, Brazil; 4) Hematology and Hemotherapy Center, University of Campinas, Campinas, SP, Brazil; 5) Center for Integral Attention to Women's Health (CAISM), University of Campinas, Campinas, SP, Brazil

**Introduction:** Sickle Cell Disease (SCD) constitutes a group of genetic pathologies that have as common characteristic the presence of hemoglobin S. The pregnancy of women with SCD is accompanied by an increased incidence of pain episodes and several systemic complications. Furthermore, placentas from these pregnancies present a series of abnormalities and dysfunctions, but little is known about genetic regulation in these tissues. Studies have shown that the placenta may suffer changes in its DNA methylation profile in the presence of maternal diseases. Thus, we have investigated if the DNA methylation profile is altered in placentas from pregnant women with SCD compared with placentas from healthy pregnant women.

**Methods:** We analyzed pregnant women: 8 with sickle cell anemia (SCA), 8 with hemoglobinopathy SC (HbSC) and 7 healthy individuals (control group). After childbirth, the placental tissue was collected and submitted to DNA extraction, bisulfite conversion and hybridization in the methylation BeadChip. We used *minfi* package for data normalization and technical correction. Probes located in SNPs and in the Y and X chromosomes were excluded. To analyze the differentially methylated positions (DMPs) we considered  $\beta$ -values  $\geq 15\%$  and applied Linear Models together with Empirical Bayesian approach implemented in *limma* package; the p-values were adjusted (adjP) by the Benjamini-Hochberg. The same criteria were applied to select differentially methylated regions (DMRs).

**Results:** The comparative analysis between SCA and control groups revealed 57 DMRs, of which 50 different genes were identified. The analysis between HbSC and control groups detected 106 DMRs, of which 87 different genes were identified. Among the most significant DMRs, some genes can be highlighted in the SCA group: *PTGFR*, *LHCGR*, *HOXD-AS1* and *GALR2*, which are involved in apoptotic process, cellular response to gonadotropin, cell proliferation and multicellular organism development. Considering the HbSC group, we highlight the *HLA-L*, *SEC31B*, *MARC2* and *GPR83* genes, which are involved in immune system process, protein transport, oxidation-reduction process and signal transduction.

**Conclusions:** SCA and HbSC pregnant women demonstrated altered DNA methylation profile in the placental tissue in comparison to the control group, suggesting that SCD can affect DNA methylation in placentas and potentially could contribute to placental dysfunction or fetal diseases.



## PgmNr 3243: Studying chromatin accessibility in human SLE at single-cell resolution.

### Authors:

R.C. Pelikan<sup>1</sup>; Y. Fu<sup>1</sup>; C.A. Lareau<sup>2</sup>; J.A. Kelly<sup>1</sup>; S. Ma<sup>3</sup>; J. Buenrostro<sup>3</sup>; P.M. Gaffney<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genes and Human Disease Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma.; 2) Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; 3) Department of Stem Cell and Regenerative Biology, Harvard University, Boston, Massachusetts

---

**Introduction:** Systemic lupus erythematosus (SLE) is a complex autoimmune disease arising from genetic and environmental factors. While the influence of many genetic risk loci on SLE has been well-described, the mechanisms responsible for SLE pathogenesis are poorly understood. Epigenetics provides a route through which genetic variation can influence chromatin accessibility and affect downstream gene expression. To understand how context-specific epigenetic changes are displayed between SLE patients and healthy controls, we applied a novel technique, single-cell combinatorial indexing assay for transposase accessible chromatin with sequencing (sciATAC-seq), to peripheral blood mononuclear cells (PBMCs). This enables us to study SLE-specific chromatin accessibility changes within the context of individual immune cell compartments.

**Method:** PBMCs isolated from 45 SLE patients and 50 controls were fixed using 0.1% paraformaldehyde and treated with transposase to attach specific adaptor index pairs. Eight samples were pooled and re-distributed into six 96-well plates with 20 cells per well. A unique PCR primer pair was added to each well. Following amplification, DNA from each cell was tagged with unique barcode combinations, then sequenced on the Illumina NextSeq platform. Cell-specific ATAC reads were demultiplexed and quantified by custom software.

**Results:** On average, 1,346 cells were sequenced per sample with a total of 398,973 accessibility sites detected. Reference bulk ATAC assays were used to assign 12 different blood cell types. Chromatin accessibility was significantly altered by SLE status at 19,944 locations dispersed throughout the genome. The proportion of differentially accessible (DA) regions unique to a given cell type ranged from 12% (CD8<sup>+</sup> T-cells) to 50% (monocytes), demonstrating high cell-type specificity. Genotyping data from our cohort was integrated to identify 7,308 genetic variants with significant allelic imbalance (aiQTLs), 74% of which are present in DA chromatin regions. Genes nearest to aiQTLs are enriched for genes related to interferon- $\gamma$  response, prioritizing these loci as genetic interfaces to the epigenetic circuitry in SLE autoimmunity.

**Discussion:** Our ability to differentiate subpopulations of immune cell types through sciATAC-seq allows us to determine changes in chromatin accessibility as a result of SLE, and identify genetic influences which may contribute to this epigenetic modulation, in a cell-type-specific manner.

## PgmNr 3244: Epigenetic regulation of *BST-2* expression levels and the effect on HIV-1 pathogenesis.

### Authors:

V. Ramsuran<sup>1,4,6</sup>; R. Singh<sup>2,3</sup>; V. Naranbhai<sup>4,5</sup>; K. Mlisana<sup>2</sup>; B. Walker<sup>5</sup>; T. Oliveria<sup>1</sup>; S. Karim<sup>4</sup>; M. Carrington<sup>6</sup>; T. Ndung'u<sup>2,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), School of Laboratory Medicine, University of KwaZulu Natal, Durban, South Africa; 2) HIV Pathogenesis Programme, Doris Duke Medical Research Institute, Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa; 3) Department of Microbiology, National Health Laboratory Services, KZN Academic Complex, Inkosi Albert Luthuli Central Hospital, Durban, South Africa.; 4) Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu-Natal, Durban, South Africa; 5) The Ragon Institute of MGH, MIT and Harvard University, Cambridge, MA 02139, USA.; 6) Cancer Inflammation Program, Laboratory of Experimental Immunology, Leidos -Frederick, Incorporated, Frederick National Laboratory for Cancer Research, Frederick, United States of America.; 7) Max Planck Institute for Infection Biology, Chariteplatz, D-10117 Berlin, Germany

---

HIV-1 must overcome host antiviral restriction factors for efficient replication. We hypothesized that elevated levels of bone marrow stromal cell antigen 2 (*BST-2*), a potent host restriction factor that interferes with HIV-1 particle release in some human cells and is antagonized by the viral protein Vpu, may associate with viral control. Expression levels of *BST-2* mRNA were measured using a real-time PCR assay and protein levels were validated by Western blotting. The expression level of *BST-2* showed an association with viral control within two independent cohorts of Black HIV infected individuals ( $r=-0.53$ ,  $p=0.015$  and  $r=-0.62$ ,  $p=0.0006$ ). DNA methylation was identified as a mechanism regulating *BST-2* levels, where increased *BST-2* methylation results in lower expression levels and associates with worse HIV disease outcome. We further demonstrate the ability to regulate *BST-2* levels using a DNA hypomethylation drug. Our results suggest *BST-2* as a factor for potential therapeutic intervention against HIV and other diseases shown to involve associate with *BST-2*.

## PgmNr 3245: DNA methylation features predict HIV frailty and mortality.

### Authors:

C. Shu <sup>1,2</sup>; X. Zhang <sup>1,2</sup>; A. Justice <sup>3</sup>; K. Xu <sup>1,2</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry, Yale School of Medicine, New Haven, CT.; 2) Connecticut Veteran Healthcare System, West Haven, CT; 3) Department of Internal Medicine, Yale School of Medicine, New Haven, CT

---

### Introduction:

Veterans Aging Cohort Study Index (VACS Index) is a component measure that is associated with disease progression and mortality in HIV-infected population. DNA methylation features associated with VACS index may be biomarkers for predicting HIV clinical frailty and mortality.

### Methods:

DNA methylome (DNAm) in peripheral blood from 1,150 HIV-infected men was profiled using the Infinium Human Methylation 450K BeadChip (450K) (training set) or the Infinium Human MethylationEPIC BeadChip (EPIC) (testing set). All individuals were grouped as a high frailty group (VACS index score  $\geq 50$ , N= 228), and low frailty group (VACS index group  $< 50$ , N= 916). In the training set, we pre-selected DNAm probes associated with VACS index for the feature selection process. We applied four machine learning methods (random forest, XGBoost, Support Vector Machines and GLMNET) to develop prediction models for the VACS index. In each method, 10-fold cross validation was performed. The best performing model was evaluated in the testing set using Area Under the Curve (AUC) and balanced accuracy, which is a better accuracy measurement in an unbalanced sample. To test whether the selected probes were collectively associated with mortality, we constructed a DNAm score by the sum of DNAm beta values weighted by effect size. DNAm score for each individual was normalized by using averaging DNAm score within the low frailty group. All individuals were assigned into high or low DNAm score groups. Cox proportional hazards model on ten-year mortality was estimated by comparing high versus low DNAm score groups.

### Results:

A panel of 113 DNAm features was selected from the training set to predict VACS index. The 113 probes showed AUC 0.835 (95%CI: 0.792, 0.879) and a balanced accuracy of 0.693 in the testing set. Among four models, random forest outperformed the other three methods. Furthermore, DNAm score (high vs. low) developed using this panel was significantly associated with mortality with hazard ratio of 2.04 (95%CI: 1.37, 3.02), adjusting for age, race, viral load, HIV treatment, smoking and CD4 count.

### Conclusion and General Impact:

We identified a panel of DNA methylation features that is predictive for HIV frailty and mortality. The DNAm profile of the VACS index may provide new insights into the mechanisms of HIV progression.

## PgmNr 3246: Epigenome-wide association study of immunoglobulin E levels using high-resolution DNA methylation profiling.

### Authors:

F. Demenais <sup>1</sup>; L. Baumier <sup>2</sup>; S. Chanoine <sup>2</sup>; W. Cheung <sup>3,4</sup>; A. Malpertuis <sup>2</sup>; E. Bouzigon <sup>1</sup>; D. Aïssi <sup>2</sup>; E. Boulrier <sup>3</sup>; X. Shao <sup>3</sup>; M. Munter <sup>3</sup>; T. Pastinen <sup>3,4</sup>; M. Lathrop <sup>3</sup>; W.O.C. Cookson <sup>5</sup>; M.F. Moffatt <sup>5</sup>; P. Jeannin <sup>1</sup>; I. Pin <sup>2</sup>; E. Grundberg <sup>3,4</sup>; V. Siroux <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) INSERM UMR-1124, Team of Genetic Epidemiology and Functional Genomics of Multifactorial Diseases, Université de Paris, Paris, France; 2) INSERM UMR-1209, Team of Environmental Epidemiology, Institute for Advanced Biosciences, Université Grenoble Alpes, Grenoble, France; 3) Department of Human Genetics, McGill University and Génome Québec Innovation Centre, Montréal, Québec, Canada; 4) Center for Pediatric Genomic Medicine, Children's Mercy Kansas City, Kansas City, MO; 5) Section of Genomic Medicine, National Heart and Lung Institute, London, United Kingdom

Asthma displays important phenotypic heterogeneity. Allergy features, mediated by Immunoglobulin E (IgE) levels, define a major endophenotype of asthma. The identification of differentially methylated CpG sites (DMC) or regions (DMR) associated with IgE levels in asthmatics may bring further insight into asthma heterogeneity.

High-resolution methylome data in regulatory elements of immune cells obtained by MethylC-Capture Sequencing (2.8 million CpGs after quality control) were analyzed in association with IgE levels in 599 blood samples from asthmatic subjects of the Epidemiological study on the Genetics and Environment of Asthma (EGEA). A binomial mixed model was fitted for each CpG, considering the proportion of methylated reads weighted for sequence read coverage as the dependent variable and log(IgE) as predictor while adjusting for age, sex, smoking and proportions of leucocytes. Associations reaching a  $P$ -value  $\leq 10^{-4}$  ( $N=317$  DMCs) were followed-up in 194 subjects from the Medical Research Council for Eczema (MRCE) study. The results from the two datasets were meta-analyzed using the inverse variance-weighted average method.

Meta-analysis of EGEA and MRCE results identified a DMC at chr21:45628025 associated with IgE at genome-wide significance ( $P=8.5 \times 10^{-9}$ ). The DMC maps to an immune-cell specific distal regulatory element 20kb downstream of *ICOSLG* gene. In addition, we found suggestive associations ( $P < 5 \times 10^{-6}$ ) at 13 other loci. Further investigation of DMRs, based on combining spatially correlated DMC  $P$ -values, showed that out of 15 DMRs detected in EGEA, the one replicating in MRCE included *ICOSLG*.

This study identified a novel IgE-associated DMC, nearby *ICOSLG*, a gene encoding the ligand for the T-cell specific cell surface receptor ICOS, which acts as a costimulatory signal for T-cell proliferation and cytokine secretion. This gene is of major biological relevance for allergy but has not yet been reported by genome-wide studies of allergic diseases.

Grant support: PHRC-CL-20131610-EVADA and ANR-15-EPIG-004-05-RESET-AID

## PgmNr 3247: Epigenetic signatures of sensitization and disease in beryllium-exposed individuals.

### Authors:

I. Konigsberg <sup>1</sup>; P. Mroz <sup>2</sup>; L. Li <sup>2</sup>; K. MacPhail <sup>2</sup>; T. Fingerlin <sup>3</sup>; L. Maier <sup>2</sup>; I. Yang <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Biomedical Informatics & Personalized Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO.; 2) Division of Environmental & Occupational Health Sciences, National Jewish Health, Denver, CO.; 3) Center for Genes, Environment, & Health, National Jewish Health, Denver, CO.

---

Beryllium exposure can lead to the development of chronic beryllium disease (CBD), an occupational lung disease characterized by granulomatous inflammation. Exposure first leads to T cell sensitization detected in the peripheral blood of genetically susceptible individuals; in some of these sensitized (BeS) individuals, beryllium-specific oligoclonal CD4+ T cells migrate to the lung and drive alveolar inflammation and CBD, while others do not develop this response. Despite well characterized genetic and environmental risks, the progression from BeS to CBD is poorly understood. Epigenetic mechanisms may explain variable disease phenotypes among those with similar exposure and genetic profiles. Our group has profiled DNA methylation and gene expression in whole bronchoalveolar lavage (BAL) cells and BAL-isolated CD4+ T cells from CBD and BeS patients. Cells were isolated from BAL collected from 30 CBD and 30 BeS individuals. DNA and RNA were extracted and assayed using Illumina HumanMethylation EPIC BeadChips and Agilent SurePrint G3 microarrays. Differentially methylated regions (DMRs) and differentially expressed genes (DEGs) were assessed using linear models adjusting for age, sex, and lymphocyte percentage. CD4+ T cells were isolated from BAL from a further 13 CBD and 13 BeS patients to confirm signals from whole BAL. Methylation was assayed again using EPIC BeadChips, while RNA was polyA-enriched and sequenced using the Illumina NovaSeq 6000.

89 DMRs and 121 DEGs were identified between CBD and BeS, and were found to be enriched for chemokine receptor activity, tumor necrosis factor receptor binding, T cell aggregation, cell-cell adhesion and other signatures of abnormal T cell physiology. Signatures of cytotoxicity, such as upregulation of multiple granzyme genes (GZMA, GZMM, GZMH), unexpected in Th1 cells seen in disease, indicate a possible functional role for DNA methylation in driving the inflammatory response observed in disease. These DNA methylation/gene expression changes will next be tested for association with disease in CD4+ BAL cells. Top hits in promoter and enhancer-associated CpG islands will be validated and functionally altered through dCas9-mediated epigenetic editing to observe effect on cellular phenotype. Multiple genes and pathways that are dysregulated in CBD have been identified and offer future potential targets for biomarker development, therapies, and further study.

## PgmNr 3248: Genetic variation in cis-regulatory domains and trans-regulatory hubs of immunity.

### Authors:

G. Rey <sup>1</sup>; A. Ramisch <sup>1,2</sup>; O. Delaneau <sup>1,2</sup>; E.T. Dermitzakis <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 2) Department of Computational Biology, University of Lausanne, Lausanne, Switzerland.

---

Genome-wide association studies have identified a large number of genetic variants that are associated with common diseases. However, the phenotypic effects of genetic variants are generally small and determining the causal effect at the molecular level may be challenging. Recently, several research groups, including our own, have used functional genomics data from a population to infer the effect of genetic variation on the structural and functional units in the human genome. In particular, we recently reported the discovery of cis-regulatory domains (CRDs) in LCLs from the inter-individual correlation among regulatory elements (Delaneau et al., Science 2019).

Building on this approach, we set out to investigate the cell-type specificity of CRDs using data from the BLUEPRINT Consortium, which include histone ChIP-Seq and RNA-Seq data from up to 200 individuals in neutrophils, monocytes and T cells (Chen et al., Cell 2016). Using this framework, we were able to call CRDs across the three immune cell types and found that a majority of them were cell-type specific. When looking at fraction of CRDs under genetic control (i.e. CRD-QTLs), we found that more than 60% of CRDs in monocytes and neutrophils were genetically controlled both with respect to their overall activity but more interestingly to their interactions (structure). We next integrated the CRD annotation with gene expression and investigated the regulation of gene expression via CRDs. We found that up to 40% of the CRDs were associated with at least one gene in a given cell type. Moreover, CRD-gene associations were generally specific to a single cell type, as only about a third was shared from one cell type to another. To characterize the impact of CRDs on gene regulation, we studied co-expression of gene pairs and their relative position to associated CRDs. We found that the fraction of co-expressed genes associated with the same CRD was strongly enriched, suggesting that CRDs may be functional units of coordination of gene regulation. In addition, we also noticed that co-expressed gene pairs associated with a CRD were mostly located within the boundaries of the CRD itself. Current analysis extends to trans regulatory hubs (TRHs) and the degree of tissue specificity of these trans interactions. Together, this study highlights the generalized role of CRDs and TRHs across 3 immune cell types and reveals mechanisms of cell-type specificity based on interactions between genes and regulatory domains.

## PgmNr 3249: eQTL fine mapping by CRISPR-Cas9 genome editing with single-cell clones and CROP-seq.

### Authors:

R. Tian <sup>1</sup>; Y. Pan <sup>2</sup>; T. Etheridge <sup>2</sup>; G. Bao <sup>2</sup>; C. Lee <sup>3</sup>; G. Gibson <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Integrative Genomics, Georgia Institute of Technology, Atlanta, GA; 2) Department of Bioengineering, Rice University, Houston, TX; 3) APC Microbiome Ireland, University College Cork, Ireland

---

Precision identification of causal variants that are responsible for disease risk is one of the major challenges in interpretation of the molecular basis of genome-wide association signals, the majority of which are located in the non-coding regions of loci and are thought to influence phenotypes by altering gene expression. We describe two approaches to fine map causal variants within credible intervals using moderate-throughput CRISPR-Cas9 mediated gene editing in monoclonal HL60/S4 cells. The first approach uses Fluidigm nano-scale PCR to profile individual clones with confirmed short deletions of candidate eSNPs in 8 autoimmune-disease associated eQTL peaks. Although significantly decreased expression was detected at the *CISD1* locus, substantial alterations in expression of non-targeted genes was observed in the same clones. Power calculations using a linear mixed model simulation study based on observed sources of experimental variance indicates that 80% power to detect at least 1 standard deviation unit after disruption of one of ten SNPs in an eQTL credible interval requires 8 replicates of each clone; with more complex models and relaxed sources of technical error, this approach was judged to be infeasible to fine map most eSNPs. Subsequently, we have begun to evaluate pooled lenti-CRISPR screening by CROP-seq on the 10X Genomics single-cell RNA sequencing platform. We targeted 67 SNPs, approximately 20 each in *CISD1*, *DAP* and *PARK7* eQTLs, along with controls, and contrasted GFP and puromycin-based selection of transduced cells. A novel targeted PCR method was used to obtain nearly 100% assignment of guide RNAs to cells, also revealing unexpectedly high multiplicity of infection that is not apparent from the raw scRNA-seq traces. This multiplicity facilitated fitting of a multivariate linear regression model contrasting all targeted SNPs at each locus, leading to identification of at least one SNP in each gene with significant alteration of the target expression. Variation in cell cycle and metabolic state that arises in the genetically uniform cells in the 10 days after transfection affects gene expression that must be controlled for in the analysis. Efforts to replicate this approach in primary lymphocytes also will be reported, and we will discuss the potential utility of CROP-Seq for fine mapping causal variants in eQTL credible intervals.



## PgmNr 3250: Transcriptional signatures of psychosocial experiences reveal GxE effects in leukocyte gene expression of children with asthma.

### Authors:

J.A. Resztak <sup>1</sup>; A.K. Farrell <sup>2</sup>; H.E. Mair-Meijers <sup>1</sup>; A. Alazizi <sup>1</sup>; S. Zilioli <sup>2</sup>; R.B. Slatcher <sup>2</sup>; R. Pique-Regi <sup>1</sup>; F. Luca <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan.; 2) Department of Psychology, Wayne State University, Detroit, MI

Environmental effects such as psychosocial experiences have been demonstrated to affect the epigenome and transcriptome and ultimately affect human health. Psychosocial factors include intrapersonal processes like emotionality, interpersonal social relationships, and broader structural environments like neighborhoods and socioeconomic status. Psychosocial stress affects cortisol levels that in turn affect gene expression in immune and inflammation-related pathways, which can exacerbate the severity of symptoms of diseases such as asthma. Allergic asthma is a chronic disease characterized by inflammation of the airways and hyperreactivity to allergens, with higher severity and morbidity rates in populations with health disparities associated with their environments. Genetic studies have identified transcriptional dysregulation in immune cells in blood circulation as an important component of underlying asthma heritability.

To identify transcriptional signatures of psychosocial experiences that may affect asthma symptoms, we analyzed 23 psychosocial variables and 16 asthma traits in leukocytes from an ethnically-diverse sample of 119 youth from Detroit, MI. We identified significant transcriptional signatures of 9 psychosocial and 4 asthma variables. We found considerable sharing of transcriptional signatures between asthma and psychosocial experiences: e.g, the transcriptional signature of self-disclosure was significantly associated with those of several measures of asthma: lung functioning (normalized FEV1,  $r=.30$ ,  $p=.001$ ), nightly asthma symptoms ( $r=.22$ ,  $p=.02$ ) and asthma severity ( $r=.38$ ,  $p<.001$ ). Gene co-expression network analysis identified 12 modules significantly associated with both measures of psychosocial experiences and asthma symptoms (10% FDR), including modules enriched for genes in the type I IFN, AP1, TNF and interleukin 10, 4 and 13 signaling pathways.

Using eQTL mapping on an expanded sample of 149 individuals, we identified 4,669 eGenes (FDR = 10%), 1,355 of which have not previously been detected in GTEx blood data. Using the transcriptional signatures as environmental proxies, we identified 117 interaction eQTLs, including 93 with psychosocial environments, thus demonstrating that negative psychosocial experiences in humans can contribute to interindividual variation in gene expression.

These results demonstrate that social genomics approaches in humans can uncover potential molecular mechanisms underlying health disparities.



## PgmNr 3251: Linking candidate GWAS (IBD) non-coding variants to genes using chromatin conformation capture assays.

### Authors:

S. Ghosh <sup>1</sup>; C. Espinoza <sup>1</sup>; F. Rahimov <sup>1</sup>; T. Mansfield <sup>1</sup>; J. Waring <sup>1</sup>; H. Jacob <sup>1</sup>; J. Lazar <sup>1</sup>; C. Miller <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomics Research Center, AbbVie, North Chicago, Illinois.; 2) Functional Genomics, GSK, Upper Providence, PA

---

An emerging challenge in the post-GWAS (genome-wide association studies) era is determining the function of disease-causing single nucleotide polymorphisms (SNPs) and linking them to their target genes. This is confounded by the fact that the majority of these GWAS-implicated SNPs (>90%) are in non-coding regions of the genome. Importantly, >60 % of these non-coding variants overlap with regulatory elements like enhancers suggesting that gene regulatory changes contribute to SNP function and disease risk. It is known that enhancers act in a cell and tissue-specific manner and can regulate genes tens or hundreds of kilobases away mediated by chromatin looping and long-range interactions. Thus, our goal is to intersect these disease-associated regulatory SNPs with 3D-genome interactions to identify disease-relevant target genes. From publically-available data sources, we identified and prioritized the most significant IBD (Inflammatory Bowel Disease) GWAS SNPs by integrating statistical significance, replicability, and epigenetic marks such as H3K27Ac, H3K4me3 and CTCF binding, along with DNaseI Hypersensitivity Site maps in relevant cell types. By using multiple chromatin conformation capture (3C) assays such as 4C-seq, PLAC-Seq, EnChIP and Hi-C we show how one regulatory SNP (rs12585310) interacts with the *CDX2* gene nearly ~1 Mb away. We also find that the interaction is cell-type specific and is only present in colonic epithelial cells (Caco-2 and immortalized primary human colonic epithelial cells) and not in other cell types (Jurkat, HS27). We further confirmed the interaction by using the *CDX2* locus as the viewpoint and showing reverse 4C interaction at rs12585310. *CDX2* is a homeobox transcription factor and is known to play a role in gut epithelial cell differentiation and is also a downstream target of inflammatory cytokines. We hypothesize that rs12585310 modulates the expression of *CDX2* leading to changes in epithelial lining of the colon which could alter responses to infection and inflammation. Indeed, single cell Caco2 clones harboring mono-allelic deletion of a 5 Kb region surrounding rs12585310 show reduced expression of *CDX2*. Further validation using CRISPR HDR to generate Caco2 clones with rs12585310 point mutations is in progress.

## PgmNr 3252: Optimal strategies for linking disease-associated SNPs to genes.

### Authors:

S. Gazal<sup>1,2</sup>; O. Weissbrod<sup>1,2</sup>; J. Nasser<sup>2</sup>; F. Hormozdiari<sup>1,2</sup>; C. Fulco<sup>2</sup>; J.M. Engreitz<sup>2</sup>; A.L. Price<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts.; 2) Broad Institute of MIT and Harvard, Cambridge, Massachusetts.

---

Although genome-wide association studies (GWAS) have identified thousands of disease-associated common SNPs, these SNPs generally do not implicate the underlying genes. Many strategies have been developed to systematically link regulatory SNPs to the genes that they regulate, but it is currently unclear which of these strategies should be prioritized.

We developed a framework, SNPs to genes (S2G), for evaluating and combining different strategies for linking SNPs to genes. The main idea of S2G is that an accurate linking of SNPs to genes should preferentially link SNPs with large causal effect sizes to constrained genes (high pLI; Lek et al. 2016 *Nature*), which are known to be enriched for common disease heritability. We construct genomic annotations based on SNPs linked to high-pLI genes by each respective strategy, and use stratified LD score regression to assess the contribution to disease heritability of each annotation; we assess this using standardized effect size ( $\tau^*$ ), defined as the proportionate change in per-SNP heritability associated with a 1 s.d. increase in the value of the annotation (conditional on exon, promoter, and MAF annotations) and meta-analyzed across traits. We identify strategies with jointly significant  $\tau^*$  values, which can then be used to weight these strategies.

We applied S2G to GWAS data from 12 independent autoimmune diseases and blood traits (average  $N=345K$ ), evaluating strategies incorporating functional annotations in blood. Top strategies included ABC enhancers (Fulco et al. *bioRxiv*) ( $\tau^*=2.05$ ), ATAC peak-gene expression correlation (Ulirsch et al. 2019 *Nat Genet*) ( $\tau^*=0.89$ ), Roadmap enhancer-gene linking (Liu et al. 2017 *Genome Biol*) ( $\tau^*=0.83$ ), fine-mapped eQTL (Hormozdiari et al. 2018 *Nat Genet*) ( $\tau^*=0.82$ ), and promoter capture Hi-C (Javierre et al. 2016 *Cell*) ( $\tau^*=0.57$ ) ( $P=10^{-15}$ - $10^{-6}$ ; all 5 strategies Bonferroni-significant in joint model). Simple strategies based on 100kb windows around each gene or on closest gene were less informative ( $\tau^*=0.30$  and  $\tau^*=0.06$ , respectively; both non-significant in joint model). The combined S2G strategy increased the strength of cell-type-specific signals in partitioned heritability analyses of the 12 traits using differentially expressed gene sets (Finucane et al. 2018 *Nat Genet*) ( $\tau^*=0.98$  vs.  $\tau^*=0.34$  for the default strategy based on 100kb windows around each gene) ( $P=10^{-33}$  vs.  $P=10^{-25}$ ). Our results highlight the benefits of combining different strategies for linking SNPs to genes.

## PgmNr 3253: DNA methylation changes are associated with particulate matter 2.5 exposure in SLE patients.

### Authors:

O. Solomon <sup>1</sup>; C. Lanata <sup>2</sup>; M. Dall'Era <sup>2</sup>; J. Yazdany <sup>2</sup>; P. Katz <sup>2</sup>; L. Trupin <sup>2</sup>; K. Taylor <sup>2</sup>; J. Nititham <sup>2</sup>; B. Rhead <sup>2</sup>; L. Criswell <sup>2</sup>; L. Barcellos <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetic Epidemiology and Genomics Laboratory, University of California, Berkeley, Berkeley, CA; 2) University of California, San Francisco, San Francisco, CA

---

Systemic lupus erythematosus (SLE) [152700] is an autoimmune disease with heterogeneous clinical manifestations. Epigenetic changes, including DNA methylation, have been implicated in SLE; specifically, differences in methylation have been associated with auto-antibody and lupus nephritis status. A prior study showed that closer residential proximity to highways was associated with hypomethylation of 3 CpG sites in the *UBE2U* gene in SLE patients. In this study, we examine the association of particulate matter 2.5 (PM<sub>2.5</sub>) levels on the day of blood draw with DNA methylation in a cohort of SLE patients from the California Lupus Epidemiology Study (CLUES). PM<sub>2.5</sub> levels were estimated by Sonoma Technology from air pollution concentrations based on geocoded residential locations. Concentrations of PM<sub>2.5</sub> fell between 1.8 and 25.9 µg/m<sup>3</sup> (SD = 3.3) with all subjects' exposure classified as either "Good" or "Moderate" according to the Air Quality Index (AQI). DNA methylation was measured using the Illumina HumanMethylationEPIC BeadChip for 271 unique subjects of White, Hispanic, African American, and Asian ethnicities. Noob background subtraction with dye-bias correction as well as quantile normalization were conducted in *minfi*. CpG sites with high detection p-values, cross-reactive probes, and CpG sites potentially measuring SNPs were removed for a total of 748,793 CpG sites for analysis. We first conducted an analysis for differentially methylated position (DMPs) using *limma* linear models with empirical Bayes variance shrinkage, adjusting for age, sex, current smoking status, current medication use, estimated cell-type proportions from *ReFACTor*, and genetic ancestry. Preliminary results showed 17 DMPs (after FDR correction) associated with PM<sub>2.5</sub> exposure, 16 of which were hypomethylated. These CpG sites mapped to genes including *TTC17*, *SGCB*, *CCDC26*, *AFF3*, *KCP*, *VGLL4*, *CNTN2*, *NPHP4*, and *GAPVD1*. We then conducted an analysis of differentially methylated regions (DMRs) using *bumphunter*. One region located 900bp upstream of *KLRC3* was significantly differentially methylated after FDR correction. Given the previous findings in the *UBE2U* gene, we also conducted a candidate DMP analysis for the 21 CpG sites on the EPIC chip which map to *UBE2U*; however, no CpG sites reached statistical significance. These results suggest for the first time that there is differential methylation with increased PM<sub>2.5</sub> exposure in SLE patients.

## PgmNr 3254: Interleukin-6 genetic polymorphisms correlate to DNA methylation and cytokine expression in Kawasaki disease.

### Authors:

H. Chi<sup>1,2,3,4</sup>; N.C. Chiu<sup>1,2,3,4,5</sup>; H.W. Yang<sup>3,6</sup>; M.R. Chen<sup>1,3,4,7</sup>; K.D. Yang<sup>1,4,6,8</sup>; T.N.D. Huang<sup>1,2,3</sup>; Y.I. Huang<sup>1,2,3</sup>; L. Chang<sup>1,2,3</sup>; F.Y. Huang<sup>1,2,3,4</sup>; Y.J. Lee<sup>1,6,9,10,11</sup>

[View Session](#)
[Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, MacKay Memorial Hospital, Taipei, Taiwan; 2) Department of Pediatric Infectious Diseases, MacKay Children's Hospital, Taipei, Taiwan; 3) MacKay Junior College of Medicine, Nursing, and Management, Taipei, Taiwan; 4) Department of Medicine, Mackay Medical College, New Taipei City, Taiwan; 5) Department of Pediatric Neurology, MacKay Children's Hospital, Taipei, Taiwan; 6) Department of Medical Research, MacKay Memorial Hospital, New Taipei City, Taiwan; 7) Department of Pediatric Cardiology, MacKay Children's Hospital, Taipei, Taiwan; 8) Department of Pediatric Allergy and Immunology, MacKay Children's Hospital, Taipei, Taiwan; 9) Department of Pediatric Endocrinology, MacKay Children's Hospital, Taipei, Taiwan; 10) Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; 11) Institute of Biomedical Sciences, Mackay Medical College, Taipei, Taiwan

**Background:** Kawasaki Disease (KD) is a febrile disease of children complicated with coronary artery lesions (CALs). Its incidence in Taiwan is 69/100000 under age 5 and ranks the third after Japan and Korea worldwide. Higher incidence rates in Japanese, Koreans, Han Chinese, and even Asian Americans indicate that genetic factors correlate to KD susceptibility. Activation of numerous immunologic factors such as T cells, cytokines and adhesion molecules, and delayed apoptosis of leukocytes characterize the acute stage and may contribute to the pathogenesis of KD.

Histopathological investigation of the coronary arteritis has revealed that activated T cell-dependent processes characterized by transmural infiltration of activated T cells occur with accumulation of CD8<sup>+</sup> T cells in vascular lesions. Because abnormal immune activation and inflammatory cytokines might contribute to KD, exploration of associated molecules using genetic approaches might shed light on the underlying pathogenic mechanisms. We hypothesized that the etiology of KD might be multifactorial, programmed by genes of inflammasome, immune brake, and cytokines.

**Methods:** We genotyped 670 KD patients (150 with CALs) by using the iPLEX sequenom MassARRAY assay including BCL2, IL6, IL10, PDCD1, PDL-1, Siglec5, Siglec9, STAT3, STAT5b, TNFSF10. Distributions of genotypes and alleles were compared to those of 1517 controls released from Taiwan Biobank using  $\chi^2$  test. P values were corrected for multiple comparisons. Plasma cytokine levels in KD patients were measured using ELISA kit.

**Results:** Alleles rs2069840C ( $P=0.024$ , OR 1.50) and rs1800796C ( $P=0.03$ , OR 1.20) of the IL6 gene conferred susceptibility to KD. Alleles rs2069840C ( $N=520$ ,  $P=0.009$ , OR 1.61) and rs1800796C ( $N=514$ ,  $P=0.015$ , OR 1.23) were more frequent in KD patients without CALs compared with controls. Plasma IL6 levels after IVIG treatment were lower in patients with allele rs2068840C than those with allele rs2069840G.

**Conclusions:** We detect significant association between the IL6 gene and KD. Alleles rs2069840C and rs1800796C conferred risk of KD. Plasma IL6 levels after IVIG therapy were associated with the

allele types of the IL6 gene. IL6 rs1800796 genotypes have been reported to be related to DNA methylation in IL6 promoter region. We speculated that methylation may play a role in KD susceptibility and plasma IL6 levels.

## PgmNr 3255: 3D promoter-open chromatin connectomes in human immune cells yield insight into cell-specific gene regulatory architectures and the genetic basis of multiple autoimmune diseases.

### Authors:

C. Su<sup>1</sup>; L. Cook<sup>9,10</sup>; P. Sharma<sup>3</sup>; E. Manduchi<sup>1,2</sup>; R.M. Thomas<sup>3</sup>; M.E. Leonard<sup>1</sup>; S. Lu<sup>1</sup>; K.M. Hodge<sup>1</sup>; C. Le Coz<sup>4</sup>; L. Shi<sup>4</sup>; J.A. Pippin<sup>1</sup>; M.E. Johnson<sup>1</sup>; A. Chesi<sup>1</sup>; K.E. Suillivan<sup>4</sup>; M. Levings<sup>9,10</sup>; N. Romberg<sup>4,5</sup>; S.F.A. Grant<sup>1,5,6,8</sup>; A.D. Wells<sup>3,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Institute for Biomedical Informatics, University of Pennsylvania, Philadelphia, PA; 3) Department of Pathology, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Division of Allergy Immunology, The Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 6) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 7) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 8) Division of Diabetes and Endocrinology, The Children's Hospital of Philadelphia, Philadelphia, PA; 9) Department of Surgery, The University of British Columbia, BC, Canada; 10) Children's Hospital Research Institute, Vancouver, BC, Canada

---

Genome-wide association studies (GWAS) have identified ?hundreds of autoimmune disease-associated loci; however, localization of causal variants at these loci, plus the corresponding genes and tissues impacted, have not been determined. To date, there has been a lack of systematic analyses across multiple disease-relevant immune cell types to give a comprehensive view of autoimmune disease etiology. To address this unmet need, we performed high-resolution, promoter-focused Capture C coupled with ATAC-seq to characterize the genome-wide open chromatin-promoter connectomes of 9 primary human immune cell types, including monocytes, naïve CD4+ T cells, naïve B cells, germinal center B cells, T helper-1 (Th1), Th2, Th17, regulatory (Treg) and follicular helper T cells. We found that comparative open chromatin landscapes across these immune cell types largely reflect their hematopoietic origin, indicating that development and differentiation are major forces shaping immune gene regulatory landscapes. Intersecting these open chromatin landscapes with maps of genome-wide promoter interactions revealed the full *cis*-regulatory architectures of >20,000 genes in each cell type. Comparison of these connectomes across cell types identified lineage-specific gene regulatory architectures, revealing the structural basis of cell-specific control on gene transcription. Exploiting these connectomes for variant-to-gene mapping successfully connected 979 putative causal variants at 65% of autoimmune GWAS loci to 1,291 putative effector genes enriched for functional categories such as interferon signaling and T helper cell differentiation. We found that open regions harboring a single GWAS locus commonly contact different genes in distinct cell types, suggesting combinatorial diversity in disease-associated gene regulatory networks. For example, the open proxies at T1D locus 'ITGB7' contact *ITGB7*, *AAAS* and *PFDN5* in most immune cells, while they connect to *PRR13* specifically in germinal center B cells. In addition, some effector genes (*CSF2*, *REL*, *IKZF3*) implicated by open variants through 3D regulatory connectomes were shared by multiple

autoimmune diseases, while others (*FOSL1* for Crohn's, *STAT6* for psoriasis) were specific to a disease. These open chromatin-promoter connectomes reveal both shared and lineage-specific gene regulatory architectures across immune cells, and provide a rich resource for variant-to-gene mapping to understand the genetic basis of autoimmune diseases.

# PgmNr 3256: The transcriptional mosaic of escape from X-inactivation across immune cell-types, individuals, and twin pairs.

## Authors:

A. Zito; A.L. Roberts; J.S. El-Sayed Moustafa; K.S. Small

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Twin Research & Genetic Epidemiology, King's College London, London SE1 7EH, UK

---

X-chromosome inactivation (XCI) balances the chrX dosages between 46,XX and 46,XY mammals. In a female diploid cell, one X is silenced (Xi) while the other is active (Xa). In humans, over 15% of genes escape XCI, and escape has been linked to sexual dimorphism in gene expression, autoimmunity and cancer incidence, and phenotypes in X-aneuploidies. Heterogeneity of escape across individuals and disease-relevant cells may modulate disease predisposition. The extent of variability in escape and the influence of genetics and environment on escape are not well known. Here, we use RNAseq and DNAseq to study escape in LCLs (N=136), fat (N=14), skin (N=30) samples, and purified monocytes, B and T-CD8<sup>+</sup> cells from a monozygotic twin pair. All samples exhibited skewed XCI, as this enables to distinguish silent from escape genes. We quantified escape (EscScore) by combining the gene's allele-specific expression with the sample's XCI-skew; values range from 0 (no escape) to 1 (fully escape). In line with the average Xi/Xa expression ratio (Tukiainen *et al*, 2017), we classified genes with EscScore  $\geq 0.35$  and  $< 0.35$  as escapees and silenced, respectively. EscScore discriminated between different XCI classes (Balaton *et al*, 2015). The fraction of genes escaping XCI varied from 15% in LCLs to 34% in fat and 44% in skin. 21 genes exhibited tissue-specific escape like *DMD*, *FHL1*, *CD99L2*. Some genes show complex patterns; *DDX3X*'s escape is highly variable in LCLs across individuals, but not in other tissues, indicating a strong individual and tissue-specific component in the regulation of escape. 50 genes exhibited consistent escape limited to one tissue (e.g *IRAK1* in LCLs, *JPX* in fat), and 60 genes variable escape across both individuals and tissues (e.g *KDM6A*, *TIMP1*). Overall level of escape in an individual varied between immune cell types ( $P < 10^{-15}$ ), with highest escape in T-CD8<sup>+</sup> cells. We identified 16 novel immune cell type-specific escapees like *FMR1* and *MAGT1* (T-CD8<sup>+</sup>), *MTCP1* (B-cells). We did not detect escape in monocytes. Thus, Xi regulation differs between myeloid and lymphoid lineages and between lymphoid cells in an individual. We found higher rates of discordant escape in dizygotic (30%) than monozygotic (24%) twins, suggesting that escape has heritable and environmental components. Within an individual, XCI escape generates a complex transcriptional mosaicism whose regulation may modulate the immune system and impact translational approaches of immuno-therapies.



## PgmNr 3257: Single-cell and neuronal network alterations in the fragile X syndrome.

### Authors:

F. Kooy<sup>1</sup>; A. Moskalyuk<sup>2</sup>; S. Van De Vijver<sup>2</sup>; P. Verstraelen<sup>3</sup>; W.H. De Vos<sup>3</sup>; M. Giugliano<sup>2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium; 2) Molecular, Cellular, and Network Excitability Lab, University of Antwerp, Antwerp, Belgium; 3) Laboratory of Cell Biology and Histology, University of Antwerp, Antwerp, Belgium; 4) International School for Advanced Studies (SISSA), Trieste, Italy

---

Our cognitive abilities depend to a large extent on the normal development and functioning of the neuronal networks in the brain. Genetic abnormalities that disturb the establishment of physiological neuronal connectivity lead to a variety of disorders, of which Fragile X syndrome is an important example. Fragile X syndrome is a frequent cause of autism and intellectual disability and it is caused by the absence of the Fragile X mental retardation protein (FMRP), mostly due to promotor hypermethylation. The function of the underlying gene has been extensively studied, and FMRP quickly became recognized for its pivotal role in many cellular processes, including translation, RNA transport, and stability. However, despite over 20 years of research, it still remains elusive how the loss of FMRP affects brain connectivity. Abnormalities in both excitatory and inhibitory transmission have been reported in Fragile X syndrome and an altered balance between excitation and inhibition has been hypothesized to underlie the clinical consequences of absence of the protein. Over the past decades, FMRP research focused on abnormalities both in glutamatergic and GABAergic signaling. However, direct proof of this theory has not been provided as yet. Using *Fmrp* knockout mice, we studied an in vitro model of cortical microcircuitry and observed that the loss of FMRP largely affected the electrophysiological correlates of network development and maturation but caused less alterations in single-cell phenotypes. The loss of FMRP also caused a structural increase in the number of excitatory synaptic terminals. Using a mathematical model, we demonstrated that the combination of an increased excitation and reduced inhibition describes best our experimental observations during the ex vivo formation of the network connections. Our study demonstrates that loss of *Fmr1* in neurons results in impairments at multiple levels of circuit organization, largely involving both excitatory and inhibitory transmission and to a much lesser extent single-cell excitability. These alterations manifested themselves in complex forms at the network level, indicating that any therapeutic intervention might require timely and precise pharmacological modulation of synaptogenesis and of synaptic transmission during network formation in order to recover the WT phenotype in vitro.

## PgmNr 3258: Dysregulation of the chromatin remodeling-associated proteins in *Fmr1* KO2 mouse testes: A role of *Fmrp* in mouse spermatogenesis.

### Authors:

R. Zong; K. Shelly; D. Nelson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas

---

Fragile X syndrome (FXS) results from loss of *FMR1* function. In addition to cognitive impairment, men with FXS also have enlarged testes (macroorchidism) and produce sperm with only premutation length alleles despite the presence of full mutations in other tissues. *Fmr1* KO male mice show macroorchidism but are fertile. The mechanism(s) responsible for these findings remain elusive, but selection in sperm precursor cells may explain the absence of sperm carrying full mutation alleles. We utilized two mouse models, *Fmr1* KO2 and mosaic *Fmr1* deletion (produced by incomplete conditional KO using tamoxifen-inducible Cre recombinase in cells marked with Rosa-Tomato/GFP), to identify dysregulated *Fmrp* targets in testis and study spermatogenesis.

In testes from *Fmr1* KO2 mice, we found reduced levels of histone H3meK9 along with reduction of the H3K9 methyltransferase, EHMT1, mRNA for which has been found in *Drosophila* to be an FMRP target. We also found enhanced levels of HDAC1 and developmental stage dependent alterations in Protamine1 levels. Several proteins altered in brain were unchanged in testis (MAP1B, MMP2, MMP9), while GluA2 levels were reduced in both testis and brain. Staining testes with hematoxylin and DAPI showed unusual DNA staining in elongated spermatids in the absence of FMRP, and PSA cytoplasmic staining under light microscopy also suggests abnormal morphology. Analysis of mosaic testes and the sperm produced is ongoing and will allow quantification of these effects and determination of the cell specificity for presence of FMRP.

Our results are consistent with a model in which dysregulation of chromatin remodeling factors in the absence of FMRP plays a role in both the enlarged testes and the unusual sperm production in full mutation males through altered histone H3 methylation. This role for FMR1 in testis has potential implications for its function in CNS cells and suggests potential pathways and targets that might be targeted with small molecule therapies.

## PgmNr 3259: Molecular diagnosis of Prader-Willi syndrome and Angelman syndrome by MS-MLPA.

**Authors:**

B. Kim; J.A. Lee; M.W. Seong; S.S. Park

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Laboratory Medicine, Seoul National University Hospital, Seoul, Republic of Korea

---

Prader-Willi syndrome (PWS) and Angelman syndrome (AS) are rare genetic disorders which are caused by a deletion on 15q11-q13, uniparental disomy of chromosome 15, or an imprinting defect. However, the methylation analysis, a standard method for diagnosis of PWS/AS cannot distinguish these entities. The aim of the study was to confirm the utility of methylation specific multiplex ligation-dependent probe amplification (MS-MLPA) as a diagnostic tool.

We performed MS-MLPA with DNA samples from 93 patients (45 PWS, 24 AS, 24 patients with negative results) requested for methylation-specific polymerase chain reaction (MS-PCR) for the diagnosis of PWS/AS. We compared the results of both assays.

MS-MLPA showed 100% sensitivity, specificity, and concordance rate to diagnose PWS and AS. Among 45 PWS patients, 26 patients (57.8%) had a deletion of 15q11-q13 and the others (42.2%) had maternal UPD 15 or imprinting defect. Seventeen out of 24 AS patients (70.8%) had a deletion of 15q11-q13. The other 7 patients (29.2%) had paternal UPD 15 or imprinting defect.

MS-MLPA has clinical significance for the diagnosis of PWS/AS and it is superior to MS-PCR in that it can classify the mechanism of the diseases.

## PgmNr 3260: Cryptic transcription is associated with age in mammalian stem cells.

**Authors:**

B. McCauley <sup>1</sup>; L. Sun <sup>1</sup>; W. Dang <sup>1,2</sup>

[View Session](#)[Add to Schedule](#)**Affiliations:**

1) Huffington Center on Ageing, Baylor College of Medicine, Houston, TX.; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

---

Aging is a multifaceted process that challenges organisms with stresses resulting from the dysregulation of cellular processes. Unsurprisingly, given how tightly regulated it is under normal conditions, transcription is one of the key pathways disrupted during aging. Indeed, dysregulation of transcription contributes to the activation of transposable elements, the loss of cellular identity, and decreased stem cell potency with age. Our previous work identified intragenic cryptic transcription (CT) as a novel type of age-associated transcriptional dysregulation that limits the lifespan of yeast and worms. Continuing this work, we show for the first time that CT increases with age in mammalian stem cells. Increased CT is associated with disrupted chromatin structure, particularly with the reduction of H3K36me3, a histone modification known to inhibit CT throughout eukaryotes. We propose that an age-associated reduction in H3K36me3 in actively transcribed gene bodies drives disruption of chromatin structure in these regions, resulting in an open chromatin state. This open chromatin state is permissive for the entry of RNA Pol II, which can then initiate transcription from within the gene body. These aberrant cryptic transcripts may contribute to the pathological load of mammalian aging.

# PgmNr 3261: Horizontal transfer of whole mitochondria restores tumorigenic potential and dihydroorotate dehydrogenase function after respiratory recovery of mtDNA deficient cancer cells.

## Authors:

L. Dong <sup>1</sup>; J. Kovarova <sup>2</sup>; M. Bajzikova <sup>2</sup>; A. Coelho <sup>2</sup>; S. Boukalova <sup>2</sup>; J. Rohlena <sup>2</sup>; M. Berridge <sup>3</sup>; J. Neuzil <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) School of Medical Science, Griffith University, Gold Coast, QLD, Australia; 2) Institute of Biotechnology, Czech Academy of Sciences, Prague, Czech Republic; 3) Malaghan Institute of Medical Research, Wellington 6012, New Zealand.

Recently, we showed that p cancer cells devoid of mitochondrial DNA (mtDNA) recover tumour formation ability after acquisition of host mtDNA (1). In order to restore respiration, mtDNA moves between cells in whole mitochondria (2). This process is gradual, and mitochondrial respiration is recovered at tumour formation onset. While it is now clear that recovery respiration is essential for tumour formation, there are several unanswered questions, such as: i) what happens in the grafted cells before tumour starts to form; ii) why respiration recovery is needed for tumour formation and progression, and iii) what is the functional link of respiration to tumour growth.

To further explore the mechanisms underlying mitochondrial transfer between cells, we prepared cell lines from 4T1 p cells on different days after grafting, and found that at the time when tumour starts to form, respiration was largely restored. Using our unique sublines, we documented the complex machinery needed for respiration recovery. This includes gradual increase in mtDNA level of homoplasmic host polymorphism, followed by binding of mtDNA-processing enzymes to its regulatory domain, replication and transcription of mtDNA, and increased expression of components of respiratory complexes, resulting in full respiration recovery. In addition, we found that pyrimidine biosynthesis, supported by the respiration-linked enzyme dihydroorotate dehydrogenase (DHODH), is required to overcome cell cycle arrest observed in non-/low-respiring cells. The reason being that respiration recovery, which is necessary for tumour cell proliferation allowing for tumour formation and progression, is associated with efficient *de novo* pyrimidine synthesis. This pathway is dependent on the mitochondrial inner membrane enzyme DHODH, implying the role of complex III and complex IV in coenzyme Q redox-cycling. We propose that re-activation of DHODH, a rate-limiting enzyme in the *de novo* pyrimidine synthesis, is the key event for triggering tumour growth following horizontal transfer of mitochondria into mtDNA-compromised cancer cells and that it is intimately linked to mitochondrial respiration (3).

In conclusion, we propose that DHODH is a critical link between *de novo* pyrimidine synthesis and respiration, and that it is a promising target for a broad-spectrum cancer therapy.

1. Tan A et al (2015) Cell Metabolism 21, 81-94.
2. Dong L et al (2017) eLife 15, e22187.

3. Bajzikova et al (2019) Cell Metabolism 29, 399-416.

## PgmNr 3262: Clinical and genotypic analysis of 62 cases mitochondrial epilepsy in children.

### Authors:

Han; F. Fang; H. Li; Z. Liu; L. Dai; J. Deng; C. Ding; X. Ren; J. Lv; H. Wu

[View Session](#) [Add to Schedule](#)

**Affiliation:** Neurology Department, Beijing Children's Hospital, Capital Medical University, Beijing, China

---

**Objective:** To summarize the clinical and genotypic characteristics of mitochondrial epilepsy in children, and to improve the understanding of it.

**Methods:** Retrospectively analyzing and following up the mitochondrial epilepsy of genetic diagnosis from October 2011 to December 2018 in the Department of Neurology Beijing Children's Hospital Affiliated to Capital Medical University.

**Results:** A total of 62 cases. 42(67.7%) focal motor seizure, 20(32.3%) generalized tonic clonic seizure, 16(25.8%) status epilepticus, 14(22.6%) myoclonic seizure, 33(53.2%) had two or more seizure types. 29 cases(46.8%) MELAS, 11 cases(17.7%) Leigh syndrome, 6 cases combined oxidative phosphorylation deficiency, 5 cases MERRF, 4 cases Alpers syndrome. 26 cases(41.9%) MT-TL1 gene mutation(m. 3243A>G), 8 cases(12.9%) mitochondrial complex ? gene mutation, 7 cases(11.3%) mitochondrial ammonia acyl tRNA synthase gene mutation, 4 cases mitochondrial DNA polymerase gene mutation. 24(55.8%) were medically refractory epilepsy in 43 followed up.

**Conclusion:** Focal motor seizure is the most common seizure type. MELAS is the most common clinical phenotype. M.3243A>G is a hotspot mutation. The genes that encoding mitochondrial complex ?, mitochondrial ammonia acyl tRNA synthase and mitochondrial DNA polymerase are major disease genes.

## PgmNr 3263: Exploring the relationship between mitochondrial DNA copy number and gene expression.

### Authors:

S.Y. Yang <sup>1,2</sup>; V.K. Pillalamarri <sup>1,2</sup>; D.E. Arking <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Predoctoral Training Program in Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD

Mitochondria are important for cellular function, playing critical roles in energy metabolism, signaling, and apoptosis. Mitochondrial DNA copy number (mtDNA-CN) is a proxy for mitochondrial function and has been associated with cardiovascular disease (CVD), overall mortality, and other morbidities. However, the molecular mechanisms that underlie these associations are still unknown. Using the Genotype Tissue Expression Project (GTEx), we aimed to interrogate the biochemical pathways that link mtDNA-CN and disease by examining the relationship between mtDNA-CN measured in whole blood and gene expression.

We estimated mtDNA-CN from whole genome sequencing data in GTEx (N=139), adjusting for neutrophil count, hematopoietic cell count, age, and cohort with a linear model. Effects for age and neutrophil count on mtDNA-CN were consistent with published literature, supporting the reliability of our mtDNA-CN metric.

We then tested for associations between adjusted mtDNA-CN and gene expression in blood, correcting for genotyping principal components, sex, and surrogate variables generated from transcript counts. Although no associations passed Bonferroni correction, the top hits included genes linked to mitochondrial function, such as mitochondrial phosphate carrier SLC25A23 ( $p=2.43e-05$ ) and coenzyme Q binding protein COQ10A ( $p=9.72e-05$ ). Additionally, 36 out of 37 mitochondrially encoded genes were positively associated with mtDNA-CN, of which 14 were nominally significant ( $p<0.05$ ). Gene ontology enrichment analysis of associated genes revealed enrichment for CREB phosphorylation ( $p=1.90e-07$ ), glucagon signaling in metabolic regulation ( $p=6.82e-05$ ), and muscle contraction ( $p=6.59e-07$ ).

After expanding this analysis to 47 additional tissues, several genes were nominally significant in multiple tissues, most notably brain, muscle, and thyroid tissue. These genes include the mitochondrial transporter SLC25A411, the pseudogene RPS2P46, and the scaffolding protein DLG3. DLG3 and RPS2P46 have not been previously associated with mitochondrial function, and thus may identify novel functional pathways. These results strengthen the link between mtDNA-CN and mitochondrial function. They also suggest that mtDNA-CN measured in blood may reflect processes occurring in other, more disease-relevant tissues, strengthening the rationale for measuring mtDNA-CN in blood as a biomarker for complex aging-related diseases.



## PgmNr 3264: Heterozygous *de novo* KAT5 variants cause a neurodevelopmental syndrome with overlapping facial dysmorphisms, sleep disturbances and brain malformations.

### Authors:

G. Lemire <sup>1</sup>; J. Humbert <sup>2</sup>; P. Makrythanasis <sup>3</sup>; J. Rousseau <sup>4</sup>; S. Ehresmann <sup>4</sup>; T. Garcia <sup>4</sup>; R. Alasiri <sup>5</sup>; S. Salian <sup>4</sup>; A. Bottani <sup>6</sup>; E. Beaver <sup>7</sup>; J. Heeley <sup>7</sup>; A.C.M. Smith <sup>8</sup>; S.I. Berger <sup>9</sup>; S.E. Antonarakis <sup>3</sup>; X.J. Yang <sup>5</sup>; J. Cote <sup>2</sup>; P. Campeau <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, Medical Genetics Division, Sainte-Justine Hospital, University of Montreal, Montreal, Canada; 2) St. Patrick Research Group in Basic Oncology, Laval University Cancer Research Center, Centre de Recherche du Centre Hospitalier Universitaire de Québec-Axe Oncologie, Québec City, Canada; 3) Department of Genetic Medicine and Development, University of Geneva Medical School and Geneva University Hospitals, Geneva, Switzerland; 4) Sainte-Justine Hospital Research Center, University of Montreal, Montreal, Canada; 5) Rosalind and Morris Goodman Cancer Research Centre, Department of Medicine, McGill University, Montreal, Canada; 6) Service of Genetic Medicine, Geneva University Hospitals, Geneva, Switzerland; 7) Mercy Kids Genetics, St-Louis, MO, USA; 8) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 9) Children's National Health System, Washington, DC, 20010 USA

---

The contribution of epigenetic regulation by histone acetylation is being increasingly recognized as a cause of genetic diseases. Notably, variants in several lysine acetyltransferase genes, such as *KAT6A* and *KAT6B*, have been identified in individuals with neurodevelopmental disorders characterized by intellectual disability and malformations. *KAT5* (previously named *TIP60*) encodes a lysine acetyltransferase involved in gene expression, DNA repair, chromatin remodeling, apoptosis and cell proliferation, but it remains unclear whether this gene is mutated in any genetic diseases. Here, we describe three individuals with *de novo* likely pathogenic missense variants in *KAT5* that affect three highly conserved residues with one at the chromodomain (p.Arg86His in the longest isoform) and two at or near the acetyl-CoA binding site (p.Cys402Ser and p.Ser446Ala). All three individuals have cerebral malformations, seizures, global developmental delay or intellectual disability, and a severe sleep disorder. Histone acetylation assays with purified mutant proteins demonstrated that the variants decrease or abolish the ability of native *KAT5* protein complexes (NuA4/TIP60) to acetylate the histone H4 tail in chromatin. In conclusion, *KAT5* variants cause histone acetylation deficiency with likely transcriptional dysregulation of multiples downstream genes, thereby leading to a neurodevelopmental syndrome with sleep disturbance and facial dysmorphisms.

## **PgmNr 3265: Implementation of clinical genomic DNA methylation testing for patients with neurodevelopmental presentations and congenital anomalies.**

**Authors:**

B. Sadikovic

[View Session](#) [Add to Schedule](#)**Affiliation:** London Health Sciences Centre, Western University, London, Canada

---

Conventional genetic testing of patients with neurodevelopmental presentations and congenital anomalies (ND/CA), i.e., the analysis of sequence and copy number variants, leaves a substantial proportion of the patients unexplained. Some of these patients have been shown to carry DNA methylation defects at a single locus (epi-variants), while others can exhibit syndrome-specific DNA methylation changes across multiple loci (epi-signatures). Here, we investigate the clinical diagnostic utility of genome-wide DNA methylation analysis of peripheral blood in unresolved ND/CA patients and present early data from clinical testing in three clinical laboratories in North America and EU. We generated a computational model enabling concurrent detection of 20+ syndromes, and 30+ genes using peripheral blood DNA methylation data with full accuracy. We demonstrate the ability of this model to resolve patients with uncertain clinical diagnoses, some of whom had VUS's in the related genes. We show that the provisional diagnoses can be ruled out in many of the cases, some of whom are shown by our model to have other diseases initially not considered. We demonstrate by applying this model to ND/CA patients without a previous diagnostic assumption and separate assessment of rare epi-variants in this cohort, we can identify patients with syndromic genetic disorders, imprinting and trinucleotide repeat expansion disorders, and patients with rare epi-variants, a portion of which involved genes clinically or functionally linked to the patients' phenotype. This study demonstrates that genomic DNA methylation analysis can facilitate the molecular diagnosis of unresolved clinical cases, and highlights the value of epigenomic testing in the routine clinical assessment of patients with ND/CA.

## PgmNr 3266: Comparative whole genome DNA methylation profiling in Costa Rican nonagenarians with and without dementia: Further evidence for a role of *PM20D1* .

### Authors:

G. Chavarria-Soley <sup>1,2</sup>; C. Coto-Vílchez <sup>1,2</sup>; L. Mora-Villalobos <sup>2</sup>; A.D. Genis-Mendoza <sup>3</sup>; J.J. Martínez-Magaña <sup>3</sup>; H. Nicolini <sup>3</sup>; H. Raventós <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Escuela de Biología, Universidad de Costa Rica, San José, Costa Rica; 2) Centro de Investigación en Biología Celular y Molecular, Universidad de Costa Rica, San José, Costa Rica; 3) Instituto Nacional de Medicina Genómica, México

---

Alzheimer's disease (AD) is a complex disorder in which different genetic and non-genetic factors have been shown to play a role. Multiple genetic variants can increase the risk of developing AD, but there are also protective variants which play a role in the genetic architecture of the disease. Individuals of advanced age without signs of dementia are very useful in the search for protective variants, but such samples can be hard to find. The aim of this study is to identify differentially methylated regions (DMRs) in the genomes of a sample of cognitively healthy individuals and a sample of individuals with dementia, with the distinctive feature that all individuals are over 90 years of age. We studied 32 nonagenarians from Costa Rica, 11 with dementia consistent with AD (mean age 95, SD=3.4), and 21 cognitively healthy individuals (mean age 95, SD=3.4). We compared whole genome DNA methylation profiles from blood of both samples, using the Infinium MethylationEPIC BeadChip. Two different approaches were used in order to identify differentially methylated regions (DMRcate and Bumphunter). We found a total of 11 DMRs between both groups, in which 7 genes are located. Five genes were hypomethylated in the sample with dementia, and 2 were hypermethylated. These genes are involved in cell cycle regulation, embryogenesis, synthesis of ceramides, and migration of interneurons to the cerebral cortex. The gene *PM20D1* was found to be hypermethylated in the sample with dementia, which is consistent with previously reported results for AD. An association of *PM20D1* with AD has been reported in the literature, where the gene was found to be a methylation and expression QTL. The hypothesis is that a higher expression of *PM20D1* can protect against AD and that loss of this expression is a risk factor for AD. This idea is supported by our results, where the group with dementia has potentially lost the *PM20D1* protection present in the cognitively healthy nonagenarians, due to promoter hypermethylation.

## PgmNr 3267: Variant-to-gene mapping for Alzheimer's disease using 3D genomics in an activated microglia human cell line.

### Authors:

A. Chesi<sup>1</sup>; M. Argenziano<sup>1</sup>; E. Manduchi<sup>1,2</sup>; C. Su<sup>1</sup>; A. Hippen-Anderson<sup>2</sup>; S. Lu<sup>1</sup>; J.A. Pippin<sup>1</sup>; S.A. Anderson<sup>1,2</sup>; L. Wang<sup>2</sup>; G.D. Schellenberg<sup>2</sup>; A.D. Wells<sup>1,2</sup>; S.F.A. Grant<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) The Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Philadelphia, PA

---

Genome wide association studies (GWAS) have revealed many loci for Alzheimer's disease (AD). However, GWAS report genomic signals and not the precise localization of effector genes, with eQTLs making strong inferences to only a subset of such loci. Chromatin conformation-based techniques that detect contacts between distant regions of the genome offer an opportunity to understand GWAS signals that principally reside in non-coding regions, thus likely influencing regulatory elements. AD, as many other neurological disorders, is characterized by chronic activation of microglia, the resident immune cells of the brain. We leveraged a massively parallel, high resolution Capture C based method to simultaneously characterize the physical genome-wide interactions of all human promoters in a human microglia cell line (HMC3). We first generated ATAC-seq open chromatin maps in the same cell line to 'fine-map' informative (i.e. open) proxy single nucleotide polymorphisms (SNPs) ( $r^2 > 0.7$ ) for each of 45 AD independent sentinel SNPs reported to date, yielding 25 candidate SNPs for 13 of these loci. By querying our Capture C promoter 'contactome' data, we observed loops to "open" promoters relevant to two of the original GWAS loci (*ABCA7* and *CASS4*). Interestingly, none of these contacts involved the nearest gene, which is commonly assigned to a locus by the GWAS community. Next, we performed a similar set of experiments in the interferon- $\gamma$  activated microglia cell line, a more relevant context for AD. Activation was confirmed by increased expression of both CD45 and CD68 and a change in morphology. ATAC-seq revealed 4401 differentially open chromatin regions, ~17% of which resided near promoters of coding genes. Pathway analysis indicated that these genes were involved in pathways relevant to immune activation, such as death receptor signaling, antigen presentation, TREM1 and interferon signaling. While chromatin interactions at the two previous loci were preserved, we also observed loops for 3 new loci (*CELF1*, *FERMT2* and *KAT8*) in the activated microglia. Interestingly, the *FERMT2* gene was recently implicated in brain  $\alpha\beta$  deposition. In conclusion, we identified putative candidate effector genes for five AD GWAS loci in a human microglial cell line. Follow-up functional studies are required to validate these findings, and further efforts in other relevant cell types could shed light on additional loci.

## PgmNr 3268: Epigenome-wide association study across multiple cognitive domains identifies a role of DNA methylation patterns in *CLDN5* on cognitive decline.

### Authors:

A. Huels<sup>1</sup>; K.N. Conneely<sup>1</sup>; R. Edgar<sup>2</sup>; P.L. De Jager<sup>3,4</sup>; D.A. Bennett<sup>5</sup>; C. Robins<sup>6</sup>; A.P. Wingo<sup>7,8</sup>; M.P. Epstein<sup>1</sup>; T.S. Wingo<sup>1,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University, Atlanta, Georgia, USA; 2) Centre for Molecular Medicine and Therapeutics, BC Children's Hospital Research Institute, and Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 3) Cell Circuits Program, Broad Institute, Cambridge, MA, USA; 4) Center for Translational and Computational Neuroimmunology, Department of Neurology, Columbia University Medical Center, New York, NY, USA; 5) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, USA; 6) Department of Neurology, Emory University School of Medicine, Atlanta, GA, US; 7) Division of Mental Health, Atlanta VA Medical Center, Decatur, GA, USA; 8) Department of Psychiatry, Emory University School of Medicine, Atlanta, GA, US

---

Cognitive decline is the clinical hallmark of dementia; however, the brain epigenetic signature of cognitive decline is unclear. Hence, we performed a brain epigenome-wide association study of cognitive decline in 636 participants from the Religious Order Study and Rush Memory and Aging Project (ROS/MAP). We repurposed DNA methylation (Illumina HumanMethylation450) profiles of the dorsal lateral prefrontal cortex to assess decline in five cognitive domains (episodic memory, perceptual speed, perceptual orientation, semantic memory, and working memory). To analyze epigenetic associations across these five cognitive domains in a powerful way, we used the Gene Association with Multiple Traits (GAMuT) test, which is a gene-based association test of multivariate phenotypes that exploits potential pleiotropy and accounts for trait correlations. We found an epigenome-wide significant association (Bonferroni threshold:  $3.44 \times 10^{-6}$ ) between methylation sites in the *PI4KA* region and cognitive decline ( $p\text{-value} = 1.32 \times 10^{-6}$ ), which was robust to adjustment for cell type proportions ( $p\text{-value} = 2.38 \times 10^{-6}$ ). Further inspection revealed this association was primarily driven by association with declines in episodic ( $p\text{-value} = 1.62 \times 10^{-6}$ ) and working memory ( $p\text{-value} = 9.49 \times 10^{-8}$ ). Fine-mapping of the *PI4KA* region identified *CLDN5* as the most significant gene and the gene most likely to be functional. Within *CLDN5*, the strongest association was found with the CpG site cg16773741. Higher levels of methylation in *CLDN5*, which corresponded with lower gene expression levels, were associated with faster cognitive decline, suggesting a protective role of *CLDN5* expression on cognitive decline. Methylation in this region was also associated with the degree of pathology in Alzheimer's disease assessed by Braak stage in ROS/MAP and in 66 independent frontal cortex samples.

## PgmNr 3269: Expanded GGC repeats in human-specific *NOTCH2NLC* gene modulate the neuronal toxicity associated with Alzheimer's disease.

### Authors:

H.E. Kong <sup>1</sup>; Y. Li <sup>1</sup>; Y. Tian <sup>2</sup>; E. Allen <sup>1</sup>; B. Tang <sup>2</sup>; L. Shen <sup>2</sup>; P. Jin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, Emory University School of Medicine, Whitehead Research Building, Room 305A, 615 Michael Street, Atlanta, GA 30322, USA; 2) 2Department of Geriatrics, Department of Neurology, and National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, 410008, China

---

Neuronal intranuclear inclusion disease (NIID) is a slowly progressing neurodegenerative disease characterized by eosinophilic intranuclear inclusions in the nervous system and multiple visceral organs. GGC repeat expansion within the human-specific *NOTCH2NLC* gene was recently shown to be the genetic cause of NIID-related disorders. Given the phenotypic overlap between NIID and other neurodegenerative disorders, such as Alzheimer's disease (AD) and Parkinson's disease, we screened a cohort of AD patients and identified multiple affected individuals with expanded GGC repeats, suggesting that the expansions in *NOTCH2NLC* could also contribute to AD pathogenesis. *NOTCH2NLC* is one of the three human-specific NOTCH2-related genes (*NOTCH2NLA*, *NOTCH2NLB* and *NOTCH2NLC*) in 1q21.1, highly expressed in the brain and thought to be involved in the evolutionary expansion of the human brain. We analyzed the expression of these three paralogs and found that *NOTCH2NLC* displays the highest expression in the brain among these NOTCH2 paralogs. Interestingly, while the expression of *NOCTH2NLB* only displays modest changes in human prefrontal cortex during aging, the expression of *NOTCH2NLC* increased significantly with age. Thus, the increased expression of expanded GGC repeats during aging could cause neurodegeneration. To further determine whether expanded GGC repeats could contribute to AD pathogenesis, we crossed GGC repeat transgenic flies with the flies expressing either APP/BACE1 or Microtubule-associated protein Tau (MAPT), and found that the expression of expanded GGC repeats could enhance the toxicity induced by Abeta and Tau. These results together suggest that the expanded GGC repeats in the human-specific *NOTCH2NLC* gene could contribute to AD pathogenesis.

## PgmNr 3270: Genome-wide 5hmC dynamics during human organoid development and its dysregulation in Alzheimer's disease.

### Authors:

J. Kuehner <sup>1,3</sup>; E. Bruggeman <sup>1,3</sup>; J. Chen <sup>1,3</sup>; F. Wang <sup>1</sup>; C. Xu <sup>2</sup>; Z. Wen <sup>2</sup>; J. Yang <sup>1</sup>; B. Yao <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics; 2) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA; 3) These authors contributed equally to this work

---

Epigenetic modifications on cytosine residues, such as 5-hydroxymethylcytosine (5hmC), undergo dynamic changes in the developing brain, and their aberrant regulation is known to be associated with numerous neurological diseases. Studies have identified abnormal 5hmC patterns that differ in postmortem brains of Alzheimer's Disease (AD) patients versus healthy brains. However, the dynamic regulation of 5hmC during early development and how that contributes to AD pathology remains unexplored. AD is a neurodegenerative disease characterized by extensive memory loss and cognitive impairments. Evidence suggests that key changes occur during neural development long before the onset of AD pathology. Despite considerable efforts, the molecular mechanism underlying AD pathogenesis, particularly during early brain development, remains elusive. Brain organoids derived from healthy and AD patients provide a unique opportunity to study brain development at the cellular and molecular levels. We have cultured forebrain organoids derived from healthy humans and *PSEN1* mutated, early-onset AD patients to model the human AD brain. Our AD organoids show accumulation of phosphorylated Tau and Ab plaques after 60 days in culture, and replicate 5hmC alterations as previously reported in AD postmortem brains. Using 5hmC-capture coupled with high-throughput sequencing, we analyzed the 5hmC landscape across several organoid developmental timepoints in healthy controls. As a result, differentially hydroxymethylated regions (DhMRs) in organoids cultured for 56 days (D56) versus embryoid bodies showed depletion at regions associated with pluripotent processes (cell adhesion and chemical stimulus detection). Additionally, D56 DhMRs were shown to be enriched at regions pertaining to nervous system development, neurogenesis and differentiation. Comparing AD organoids to their matched control organoids, we found substantial 5hmC alterations at known AD-risk genes: *APP*, *CNTNAP2* and *FRMD4A*. Significant differentially expressed genes identified by RNA-sequence analysis and our identified intragenic DhMRs from these organoid samples were shown to be enriched for neuronal development and differentiation pathways. These findings suggest that early epigenetic dysregulation of the 5hmC landscape during neuronal development may predispose AD pathogenesis, which could lead to the understanding of molecular mechanisms underlying this devastating disease, allowing for early intervention treatment options.



## PgmNr 3271: Non-coding regulatory landscape of Alzheimer's disease variants using GWAS of 63,926 individuals.

### Authors:

Y.Y. Leung; P.P. Kuksa; C.-Y. Lee; Y.-F. Chou; A. Amlie-Wolf; G.D. Schellenberg; L.-S. Wang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Genome-wide association (GWAS) studies identified various genetic variants that are associated with the risk of late-onset Alzheimer's disease (LOAD). Most of these genetic signals affect regulatory elements outside coding genes such as transcriptional enhancers and noncoding RNAs, and due to linkage disequilibrium (LD), top reported variants may not be causal but are just correlated with truly causal variants, which remain to be identified.

We used our INFERNO (INFERring the molecular mechanisms of NOncoding genetic variants) tool (<http://inferno.lisanwanglab.org>), which integrates GWAS data with hundreds of functional genomics datasets to identify causal noncoding variants underlying GWAS association signals and their affected regulatory elements, tissue contexts, and target genes. INFERNO uses the COLOC method to identify colocalized GWAS and target gene eQTL signals overlapping enhancers in matching tissue classes and characterizes co-regulatory networks of targeted long noncoding RNAs.

We identified 1,358 significant variants ( $p < 5 \times 10^{-8}$ ) in the largest LOAD GWAS meta-analysis study to date (21,982 cases and 41,944 controls). We pruned significant variants into independent signals (281 LD blocks) by European LD and expanded these into 2,441 candidate causal variants. While only 3.5% of all candidate variants were in coding exons, 5.9% overlapped FANTOM5 enhancers, 66.2% overlapped Roadmap enhancers, and 58.4% overlapped transcription factor binding sites (TFBS). 3.4% of all the candidate causal variants overlapped with FANTOM5, Roadmap enhancers, and were disrupting TFBS. A subset of variants affected non-coding RNAs: 4.1% overlapped 12 classes of small RNA across 185 tissues/cell types using DASHR2 annotation (<http://dashr2.lisanwanglab.org>). 25% of eQTL signals colocalized with GWAS signals, spanning all 44 GTEx tissues, 67% of which were supported by enhancer overlaps in the matching tissue categories.

The target genes of the eQTL and colocalized genes were strongly enriched (corrected  $p < 0.05$ ) in Gene Ontology enrichment analyses, confirming amyloid-beta clearance, tau protein binding, immune response as top candidates. Interestingly, the analyses also revealed lipoprotein particle receptor binding and regulation of metabolic process as potential mechanisms for LOAD. These analyses support the utility of INFERNO for inferring the molecular mechanisms underlying noncoding GWAS signals in common complex diseases.



## PgmNr 3272: Hypermethylation of mitochondrial tRNA p9 sites in the cerebellum of Alzheimer's disease and progressive supranuclear palsy patients.

### Authors:

T. Silzer; G. Pathak; N. Phillips

[View Session](#) [Add to Schedule](#)

**Affiliation:** Graduate School of Biomedical Science, UNT Health Science Center, 3500 Camp Bowie Blvd, Fort Worth, Texas.

---

**Background.** Mitochondrial dysfunction has been implicated in a many age-related disease pathologies, and post-transcriptional sequence modifications in mitochondrial RNA (mtRNA) have been correlated with mitochondrial function. One of the most significant modifications is methylation of mitochondrial transfer RNAs (mt-tRNA) at their 9<sup>th</sup> position ("p9 sites") which alters the efficiency of translation and protein synthesis. The involvement of mt-tRNA modifications on risk for age-related neurodegenerative disease has not been previously explored.

**Purpose.** The purpose of this study was to determine if altered post-transcriptional modification rates to mt-tRNA p9 sites are associated with risk for neurodegenerative disease.

**Methods.** RNA-seq and SNP data was analyzed from cerebellar tissue of 266 Caucasian subjects including elderly controls, Alzheimer's disease (AD) patients, progressive supranuclear palsy (PSP) patients, and pathological aging. Data were obtained through the Synapse data repository and were collected by the Mayo RNAseq Study (Dr. Nilüfer Ertekin-Taner, Mayo Clinic, Jacksonville, FL; with multi-PI U01 AG046139 by Golde, Ertekin-Taner, Younkin, Price). Methylation was inferred from heteroplasmy in mtRNA measured using VarScan. PCA was conducted using EIGENSOFT. Gene-based GWAS was executed in FUMA. Gene expression significantly associated with p9 methylation was determined using Fisher's exact test, and gene set overrepresentation tests were conducted using PANTHER and MAGMA.

**Results.** Average rate of heteroplasmy at each of the 11 p9 sites was compared between control and case groups. Gene-based GWAS identified *SLC28A3* as being significantly associated with p9 methylation. Expression of 5300 genes was also found to be significantly associated with p9 methylation. GO overrepresentation testing using SNP and gene expression datasets displayed significant enrichment for several cellular processes, with overlap for mitochondrial-related compartments/processes.

**Concluding Remarks.** Post-transcriptional modification of mtRNA occurs under normal, non-pathologic states and alteration p9 site methylation of mt-tRNAs is associated with AD and PSP. Here we show similarities in mitochondrial profiles of AD and PSP, suggesting analogous disease pathologies in cerebellar tissue. Targeting mt-tRNA methylation may prove to be a viable therapeutic approach for deficits in mitochondrial function.

## PgmNr 3273: Plasmatic microRNA signature in Friedreich's ataxia: Preliminary findings from a Brazilian cohort.

### Authors:

T.M. Peluzzo <sup>1</sup>; M. Martin <sup>1</sup>; A.R.M. Martinez <sup>4</sup>; A.S. Vieira <sup>3</sup>; I. Lopes-Cendes <sup>1</sup>; M.C. França Jr. <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics and Genomic Medicine, School of Medicine, University of Campinas – UNICAMP, Campinas, SP, Brazil, Rua Tessália Vieira de Camargo, 126, Cidade Universitária “Zeferino Vaz”, Zip Code: 13083-887, Campinas/SP - Brazil; 2) Department of Neurology, School of Medicine, University of Campinas – UNICAMP, Campinas, SP, Brazil, Rua Tessália Vieira de Camargo, 126, Cidade Universitária “Zeferino Vaz”, Zip Code: 13083-887, Campinas/SP -Brazil.; 3) Department of Structural and Functional Biology, Institute of Biology, University of Campinas – UNICAMP, Campinas, SP, Brazil, Av Bertrand Russel S/N, Cidade Universitária “Zeferino Vaz”, Zip Code: 13083-865, Campinas/SP - Brazil.; 4) Neuroimaging Laboratory, School of Medicine, University of Campinas – UNICAMP, Campinas, SP, Brazil, Rua Tessália Vieira de Camargo, 126, Cidade Universitária “Zeferino Vaz”, Zip Code: 13083-887, Campinas/SP - Brazil

---

Friedreich’s Ataxia (FRDA) is the most common hereditary ataxia worldwide with estimated prevalence between 2 to 0.13:100,000 in Caucasians. It’s caused by a homozygous GAA expansion in the first intron of the *FXN* gene located at chromosome 9q21.11. As a result, there is a dramatic reduction in the expression of the mitochondrial protein frataxin. It leads to the cardinal manifestations of FRDA: early onset progressive ataxia and sensory deficits. Dysphagia, dysarthria, subtle cognitive dysfunction, cardiomyopathy and diabetes are also often found. MicroRNAs are well known for their ability to regulate gene expression and lately, have been explored as potential biomarkers for several neurodegenerative diseases. One way to obtain miRNAs that can be used for this purpose is extracting them from peripheral blood plasma, not only because of its known stability, but also for its accuracy in real-time phenotypic alterations. Up to now, there are very few studies looking at plasma microRNAs in patients with FRDA. In this study, we recruited a Brazilian cohort of FRDA patients and extracted miRNAs from plasma samples, then submitted all 48 small-RNAs samples (24 patients and 24 healthy controls) to an RNA-seq (Illumina platform). Our preliminary analysis using DESeq2 (shrinkage estimation for dispersions and fold changes to improve stability and interpretability of estimates; Benjamini-Hochberg procedure for adjusted  $p < 0.05$ ) showed that at least 3 miRNAs were downregulated (hsa-miR-26a-5p, hsa-miR-30b-5p, hsa-miR-423-3p) and 2 were upregulated (hsa-miR-15a-5p, hsa-miR-20a-5p) when comparing patients with healthy controls. Two of them (hsa-mir-15a-5p and hsa-mir-26a-5p) already known to be targeting *BDNF* gene, that its reduced protein levels are associated with ataxic conditions. Further technical validation utilizing qPCR is the next step of this study. Financial support: CNPq, CAPES and Fapesp fomentation.

## PgmNr 3274: Methylation of alpha synuclein intorns in a cohort of Sudanese patients with Parkinson disease.

### Authors:

Y. Bakhit <sup>1</sup>; I. Schmitt <sup>2</sup>; M. Ibrahim <sup>3</sup>; A. Ahmed <sup>4</sup>; O. Seidi <sup>5</sup>; U. Wüllner <sup>6</sup>; Sudan Neuroscience Projects (SNPs). University of Khartoum

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Basic Medical Sciences, University of Khartoum, Bonn, NRV, Germany; 2) Klinik für Neurologie, UKB; 3) Department of Biochemistry, Faculty of Medicine. Alrazi University; 4) Department of Physiology, Faculty of Medicine. University of Khartoum; 5) Department of Internal Medicine. Faculty of Medicine. University of Khartoum; 6) Klinik für Neurologie, UKB und DZNE Bonn

---

Alpha-synuclein (*SNCA*) is a critical risk gene for Parkinson's disease (PD) with mutations reported to cause parkinsonian syndrome in both sporadic and familial forms of it. We found that Methylation of *SNCA* intron 1 was reduced in DNA from Sudanese PD patients, suggesting an epigenetic role in *SNCA* expression in PD. The novelty of this study in the region suggests further screening of PD patients from Sudan as well as all over Africa.

## PgmNr 3275: ASXL3 links extracellular matrix deposition and congenital cardiac hypertrophy.

### Authors:

N. Ghali <sup>1</sup>; A. Srivastava <sup>1</sup>; B. McGarth <sup>2</sup>; Y. Tsan <sup>1</sup>; R. KC <sup>1</sup>; C. Keegan <sup>1,3</sup>; A. Helms <sup>4</sup>; S. Bielas <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, USA.; 2) Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI, USA.; 3) Department of Pediatrics, University of Michigan Medical School, Ann Arbor, MI, USA.; 4) Internal Medicine of Cardiology, University of Michigan Medical School, Ann Arbor, MI, USA.

---

Recent findings emerging from human genome sequencing efforts have highlighted genes encoding chromatin regulators, particularly histone modifiers, as the genetic basis of developmental syndromes and cancers. Covalent histone modifications play a key role in developmental transcriptional plasticity required for cell fate specification. Converging lines of investigation implicate mono-ubiquitination of histone H2A (H2Aub1) as a pathogenic mechanism of developmental disorders. H2Aub1 is an evolutionarily conserved post-translational modification associated with polycomb transcriptional repression. ASXL3 is a component of the polycomb repressive deubiquitinase (PR-DUB) complex, which deubiquitinates H2Aub1. *De novo* dominant truncating variants in *ASXL3* (*Additional-sex combs like 3*) have been identified as the genetic basis of Bainbridge Ropers Syndrome (BRS), with truncating variants result in elevated H2Aub1 and altered transcriptional regulation. To understand epigenomic alterations and early development defects, we used CRISPR/Cas9-mediated genome editing to generate clinically relevant truncating variants in a novel mouse model of BRS. Constitutive loss of *Asx13* results in highly penetrant developmental heart defects and perinatal lethality. *Asx13*<sup>-/-</sup> mice display congenital cardiac hypertrophy and septal defects. To understand the role of ASXL3-dependent epigenomic changes, we have generated human induced pluripotent stem cell (iPSC) lines from BRS and healthy individuals that are differentiated to a cardiomyocyte fate. Transcriptomic and epigenomic analysis implicate a role for ASXL3 in both human and mouse cardiac development, the mechanisms of which is species specific and not a direct phenocopy. These findings underscore the importance of ASXL3 in polycomb transcriptional repression during heart development.

## PgmNr 3276: Investigating the role of aberrant methylation in developmental and epileptic encephalopathies.

### Authors:

M. Hebbar <sup>1</sup>; J.W. MacDonald <sup>2</sup>; T.K. Bammler <sup>2</sup>; A.L. Schneider <sup>3</sup>; L. Sadleir <sup>4</sup>; I.E. Scheffer <sup>3</sup>; H.C. Mefford <sup>1</sup>; University of Washington Center for Mendelian Genomics

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; 2) University of Washington, Department of Environmental & Occupational Health Sciences, Interdisciplinary Center for Exposures, Diseases, Genomics & Environment, Seattle, USA; 3) University of Melbourne, Austin Health and Royal Children's Hospital, Melbourne, VIC, Australia; 4) University of Otago, Wellington, New Zealand

---

**Background:** The developmental and epileptic encephalopathies (DEEs) are a group of severe disorders characterized by developmental delay and intractable seizures. Advances in genomic testing technology have accelerated gene discovery in DEE; in the clinical setting, a genetic diagnosis can now be made in up to 40% of patients. However, more than 50% of patients with DEEs remain without a genetic diagnosis despite state-of-the-art genetic testing. Methylation represents an epigenetic modification of DNA that does not change the underlying DNA sequence but can impact gene expression. We hypothesize that a subset of DEEs, where the cause remains unknown using conventional sequence-based analysis, harbors rare DNA methylation abnormalities that affect gene expression and cause DEE. We predict that pathogenic aberrant methylation will be due to an underlying DNA defect.

**Methods:** We are using Illumina Infinium MethylationEPIC Kit to evaluate methylation of 850,000 CpG sites in 400 probands to identify differentially methylated regions (DMRs) compared to controls. Bioconductor tools are used to identify rare regions with extreme methylation using a sliding window approach. We will validate DMRs and perform segregation analysis to determine inheritance, prioritizing *de novo* DMRs. For *de novo* DMRs, we will perform regional or genome sequence analysis, repeat analysis and copy number studies to identify underlying DNA defects, prioritizing DMRs near known DEE genes.

**Results:** We have phenotyped 400 individuals with DEEs for whom a molecular diagnosis has not been found after sequencing at least 104 DEE genes. We have completed methylation arrays for 96 samples and developed a pipeline to identify outlier individuals with putative DMRs. Analysis and downstream testing are ongoing.

**Conclusion:** Based on studies in a similarly sized cohort of individuals with neurodevelopmental disorders, we predict that 5-10% of unsolved DEEs will be due to rare methylation abnormalities. Though there are rare examples of epilepsy-related disorders due to methylation defects, to our knowledge, this will be the first genome-wide screen for methylation abnormalities in a cohort of individuals with DEE. This study will enhance our understanding of the molecular and genetic mechanisms of epilepsy, improve diagnostics and provide new avenues for therapies for people and families living with DEEs.

## PgmNr 3277: Loss of Dnmt3a dependent methylation in inhibitory neurons impairs neural function through a mechanism that impacts Rett syndrome.

### Authors:

L.A. Lavery<sup>1,3</sup>; K. Ure<sup>1,3,2</sup>; Y. Wan<sup>1,3</sup>; C. Luo<sup>4</sup>; A. Trostle<sup>3</sup>; W. Wang<sup>1,3</sup>; J. Lopez<sup>3</sup>; J. Lucero<sup>5</sup>; R. Castanon<sup>4</sup>; J.R. Nery<sup>4</sup>; M. Durham<sup>3,7,8</sup>; Z. Liu<sup>1,3,9</sup>; M.A. Goodell<sup>1</sup>; J.R. Ecker<sup>4,6</sup>; M.M. Behrens<sup>5,10</sup>; H.Y. Zoghbi<sup>1,3,11</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas.; 2) Animal Behavior Core; University of Ottawa; Ottawa, ON K1N 6n5, Canada; 3) Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital; Houston, TX, 77030, USA; 4) Genomic Analysis Laboratory, The Salk Institute for Biological Studies; La Jolla, CA, 92037, USA; 5) Computational Neurobiology Laboratory, The Salk Institute for Biological Studies; La Jolla, CA, 92037, USA; 6) Howard Hughes Medical Institute, The Salk Institute for Biological Studies; La Jolla, CA, 92037, USA; 7) Program in Developmental Biology; Baylor College of Medicine; Houston, TX, 77030, USA; 8) Medical Scientist Training Program; Baylor College of Medicine; Houston, TX, 77030, USA; 9) Graduate Program in Quantitative and Computational Biosciences, Baylor College of Medicine; Houston, TX, 77030, USA; 10) Department of Psychiatry; University of California San Diego; La Jolla, CA 92093, USA; 11) Howard Hughes Medical Institute; Baylor College of Medicine; Houston, TX, 77030, USA

---

Methylated cytosine is a major effector of epigenetic gene regulation. In the brain, the DNA methyltransferase, Dnmt3a, "writes" a unique methylation pattern that introduces non-canonical non-CpG methylation (mCH) in neurons during a critical phase of maturation. Despite its apparent importance to brain function our understanding of the mechanism of postnatal gene regulation via Dnmt3a methylation is in its infancy. One clear methylation "reader" involved is methyl CpG binding protein 2 (MeCP2). To gain a foot-hold into this novel epigenetic pathway, we set out to determine if MeCP2 is the brain's only reader for Dnmt3a dependent methylation. We compared two lines of mice, lacking either Dnmt3a (the "writer") or MeCP2 (the "reader") in GABAergic inhibitory neurons, at the behavioral, physiological and molecular levels. We found that loss of either the writer or the reader causes overlapping as well as distinct sets of features from the behavioral to the molecular level, with the loss of Dnmt3a causing global loss of mCH and a small subset of mCG sites to drive more severe neurological dysfunction and more widespread transcriptional alterations. These data indicate that MeCP2 is responsible for reading only part of the Dnmt3a dependent methylation pattern in the brain and lays the foundation to elucidate novel readers involved in neuron maturation that may impact neuropsychiatric phenotypes.

## PgmNr 3278: Role of chromatin remodeling protein CHD2 in neurodevelopmental disorders.

### Authors:

K.J. Lamar <sup>1</sup>; A.M. Tidball <sup>2</sup>; K.Y. Zeng <sup>1</sup>; J.M. Parent <sup>2</sup>; G.L. Carvill <sup>1</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Neurology, Northwestern University, Chicago, Illinois.; 2) Neurology, University of Michigan, Ann Arbor, Illinois

---

Pathogenic variants in *CHD2* are associated with autism spectrum disorder, intellectual disability, and epilepsy in humans. Chromodomain helicase DNA binding protein 2 (CHD2) is a chromatin remodeler associated with sites of active transcription. To better understand the role of CHD2 in human neurodevelopmental disorders we utilized CRISPR/Cas9 to create heterozygous disruptions of *CHD2* in human induced pluripotent stem cells (iPSCs). To examine the effect of *CHD2* disruption in neuronal development, we converted these iPSC lines into neuronal progenitor cells (NPCs), cortical excitatory neurons (CENs), and cerebral organoids (COs). mRNA sequencing revealed that *CHD2*<sup>+/-</sup> NPCs exhibited an upregulation of genes involved in neuronal differentiation and a downregulation of genes involved in proliferation. *CHD2* disruption also lead to misregulation of other epilepsy genes. Immunofluorescence, proliferation assays and single-cell RNA-sequencing on COs are in process to examine cell type specific changes that occur upon *CHD2* deficiency in human neurons.

## PgmNr 3279: Altered chromatin accessibility in KAT6A syndrome.

### Authors:

L.M. Sathe <sup>1</sup>; V.A. Arboleda <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Pathology and Lab Medicine, David Geffen School of Medicine at UCLA; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA

---

The organization of the chromatin landscape plays a crucial role in many cellular processes through accessibility of transcription factors and regulation of gene expression. Histone acetyltransferases are chromatin remodelers that acetylate specific lysine (K) residues and may change cells' expression profiles when disrupted. Mutations lysine acetyltransferase 6A (*KAT6A*, also known as *MOZ* and *MYST3*) causes a rare Mendelian disorder with variable features including intellectual disability, speech delays, microcephaly and gastrointestinal disorders. While *KAT6A* is known to affect chromatin structure by acetylating H3K9 and other lysine residues, the mechanism by which *KAT6A* causes disease is unknown. Here, we use the Assay for Transposase-Accessible Chromatin using Sequencing (ATAC-seq) to identify regions of closed and open chromatin in a pilot study using *KAT6A* syndrome patient-derived fibroblasts from 4 patients compared to 2 control samples. Raw paired-end fastq files were mapped using the BWA MEM algorithm, peaks were called using MACS2. Downstream analysis was performed with DASTk, a toolkit for ATAC-seq analysis. We found that *KAT6A* patients have on average 52.3% more open regions compared to average control samples, however, there are 30.3% fewer bases included within those regions, suggesting that there are more peaks, but that these peaks are, on average, smaller. We also identified transcription factor (TF) binding sites that are more available in *KAT6A* patients (*KLF4* and *MAZ*), and those that are less available (*SP1*) in *KAT6A* syndrome patients compared to controls. Further integration with RNA-seq data suggests that these factors may identify gene-regulatory mechanisms downstream of *KAT6A* histone modulation and will integrate our identification to identify differential RNA expression from matched patient-derived fibroblasts. Understanding the gene regulatory aspects can identify underlying mechanisms controlled by *KAT6A* and will broaden our understanding of how mutations to chromatin modifiers directly alter nuclear organization.



## PgmNr 3280: DNA methylation in two pairs of discordant monozygotic twins for ADHD by Illumina MethylationEPIC BeadChip microarray.

### Authors:

T.V.M.M. Costa <sup>1,6</sup>; J.G. Damasceno <sup>1</sup>; F.A.R. Madia <sup>2</sup>; F.A. Marchi <sup>3</sup>; E.A. Zanardo <sup>1</sup>; G.M. Novo-Filho <sup>1</sup>; S.N. Chehimi <sup>2</sup>; A.M. Nascimento <sup>2</sup>; M. Rocha <sup>1</sup>; M.M. Montenegro <sup>2</sup>; A.T. Dias <sup>1</sup>; Y.G. Oliveira <sup>1</sup>; C.A. Kim <sup>4</sup>; V. Schuch <sup>5</sup>; C.B. Mello <sup>5</sup>; M. Muszkat <sup>5</sup>; L.D. Kulikowski <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina da USP, Universidade de São Paulo, São Paulo, Brasil; 2) Laboratório de Citogenômica, Departamento de Pediatria, Faculdade de Medicina da USP, Universidade de São Paulo, São Paulo, Brasil; 3) Laboratório NeoGene, AC Camargo Cancer Center, São Paulo, Brasil; 4) Unidade de Genética, Departamento de Pediatria, Instituto da Criança, Hospital das Clínicas da Faculdade de Medicina HC/FMUSP, Universidade de São Paulo, São Paulo, Brasil; 5) Núcleo de Atendimento Neuropsicológico Infantil Interdisciplinar - NANI, Departamento de Psicobiologia da Escola Paulista de Medicina, UNIFESP, São Paulo, Brasil; 6) Instituto Federal de Educação, Ciência e Tecnologia do Maranhão - IFMA, Campus Barreirinhas, Barreirinhas, Brasil

---

Attention deficit hyperactivity disorder (ADHD) is characterized by inadequate levels of hyperactivity/impulsivity and/or inattention and is one of the most common neurobehavioral disorder diagnosed in childhood. ADHD affects around 5% of school children worldwide and, in most cases (50-80%), the disorder persists during adolescence and adulthood. Due to its high heritability (estimated at 76%), it is known that the etiology of ADHD is strongly influenced by genetic factors. However, the genetic architecture of ADHD is not clear. Behavioral genetic studies have demonstrated that genetic influences play a role in the etiology of ADHD. However, it is very important to study whether epigenetic factors, such as DNA methylation, are related to the etiology of ADHD. Thus, performing the DNA methylation profiles of monozygotic twins are an important tool for studying the role of epigenetic factors in ADHD. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) was used as a reference for the diagnosis. DNA was extracted from blood lymphocytes from two pairs of monozygotic twins. Each pair of twins was composed of two boys: one with ADHD and one with neurotypical development. The first pair with 12 years old and the second pair with 09 years old. Then, DNA methylation assays were performed using Illumina MethylationEPIC BeadChip. Arrays data were treated and analyzed using specific packages in the R environment. Some Bioconductor packages were used to obtain a list of statistically significant probes, considering  $p < 0.05$ . Thus, we identified differentially methylated probes in the comparison between the two boys with ADHD and the two boys without ADHD. Probes were annotated according to the data provided by Illumina using the genome hg19 reference. Our results suggest that DNA methylation profiles may present significant differences between monozygotic twins discordant for ADHD. Therefore, we emphasize the importance of investigating DNA methylation in more cases of ADHD discordant twins in order to verify if the differences found in this study are, in fact, relevant for the development of ADHD.

## PgmNr 3281: DNA methylation patterns associated with cyanogenic cassava exposure in Sub-Saharan Africa.

### Authors:

K. Kocher<sup>1,2</sup>; M. Bramble<sup>1</sup>; D. Okitundu-Luwa<sup>3</sup>; M. Ngoyi<sup>4</sup>; D. Tshala-Katumbay<sup>3,4,5</sup>; E. Vilain<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genetic Medicine, Children's National Medical Center, Washington DC; 2) Institute for Biomedical Sciences, The George Washington University, Washington DC; 3) Department of Neurology, University of Kinshasa, Kinshasa, DRC; 4) National Institute of Biomedical Research, Kinshasa, DRC; 5) Department of Neurology, School of Medicine, Oregon Health and Science University, Portland, OR

---

Cassava is an innately cyanogenic, fibrous tuber that serves as the predominant food source for impoverished regions throughout Sub-Saharan Africa. If sufficiently processed, cassava can be safely consumed; however, failure to do so causes sub-lethal exposure to cyanide through the diet and epidemics of konzo. Konzo is a neurological disease that is characterized by the sudden onset of irreversible spastic paraparesis or paralysis. It has been classified by the World Health Organization as having the highest case-fatality ratio of any disease associated with exposure of chemicals through the food supply. Children and women of child-bearing age appear to be more vulnerable, and familial clusters of disease occur for reasons that have yet to be elucidated. Even with the same homogenous diet of poorly processed cassava, only up to 10% of the population is typically affected by konzo. As such, we hypothesize that the pathophysiology of konzo is mediated by changes to the DNA methylome in response to sub-lethal exposure to cyanogenic cassava.

This study investigates the global DNA methylation differences in individuals exposed to cyanogenic cassava that develop konzo compared to age-matched healthy controls from the same outbreak zones. We showed that there were 118 differentially methylated loci between those affected by konzo and healthy controls. Notably, 80% of the differentially methylated loci were hypomethylated compared to controls. Genes represented by these loci are enriched for biological processes, including neuronal cell death, oxidative stress, cellular detoxification, and epigenetic regulation, which are directly relevant to konzo presentation and cyanide metabolism. This study provides the first analysis of the durable DNA methylation differences attributed to sub-lethal cyanide exposure through diet and reveals disruption of biological processes that may underpin the etiology of konzo.

## PgmNr 3282: Transgenerational disruptions in DNA methylation and gene expression by ancestral folate supplementation.

### Authors:

A. Madrid; L. Borth; N. Hariharan; L. Papale; R. Alisch; B. Iskandar

[View Session](#) [Add to Schedule](#)

**Affiliation:** Neurological Surgery, University of Wisconsin - Madison, Madison, Wisconsin.

---

The central nervous system (CNS) has limited ability to heal and regenerate after injury. Folate is crucial in enhancing axonal regeneration in CNS systems after sharp injury, potentially by influencing the abundance of heritable environmentally sensitive molecular substrates, such as DNA methylation. We have shown that folate-induced axonal regeneration and DNA methylation in the spinal cord following sharp injury is dose-dependent and persists through 4 generations of untreated progeny, depicting a transgenerational epigenetic inheritance. These findings led us to hypothesize that DNA methylation may influence inheritance of the axonal regeneration phenotype. In order to test this hypothesis, we collected spinal cord tissue that was distal to a sharp injury from adult F3 male rats derived from an F0 lineage exposed to daily injections of folic acid or a vehicle control beginning three days before breeding and continuing until the pups were weaned. Genomic DNA from these tissues was profiled by whole genome bisulfite sequencing to examine an average of 22.7 million CpG dinucleotides. This approach found 1,636 folate-related differentially methylated regions (DMRs), and ~86% ( $N = 1,411$ ) were hypomethylated. Pathway analysis of the DMR-associated genes revealed gene enrichments related to axon and neuronal processes, such as *axonogenesis* and *neuron-projection guidance*. RNA sequencing was conducted on RNA from these same tissues to find differentially methylated candidates linked to altered gene regulation. Three genes particularly relevant to the axonal regeneration process that exhibited both folate-related differential methylation and expression were *Rxrb*, *Syn2*, and *Dscam*. These transcripts have well defined roles in regeneration, including cell-fate commitment, neuronal apoptosis, and dendritic extension, respectively. Together, these data support that folate supplementation leads to transgenerational inheritance of DNA methylation alterations and subsequent gene expression changes resulting in enhanced axonal regeneration of injured spinal cord axons. These findings are of particular interest as manipulations of modifiable substrates (*i.e.*, DNA methylation levels) are likely to provide insights into injury mechanisms and novel treatment targets for individuals suffering from spinal cord (and brain) trauma or disease.

## PgmNr 3283: Epigenome-wide meta-analysis of cerebral white matter hyperintensities on MRI and integrated genetic analysis.

### Authors:

Y. Yang <sup>1,2</sup>; R. Xia <sup>2</sup>; H. Adams <sup>3,4</sup>; J. Bis <sup>5</sup>; S. Li <sup>6</sup>; M. Jhun <sup>7</sup>; M. Luniano <sup>8</sup>; A. Teumer <sup>9</sup>; C. van Duijn <sup>3</sup>; S. Kardina <sup>7</sup>; I. Deary <sup>8,10</sup>; H. Grabe <sup>11</sup>; L. Launer <sup>12</sup>; W. Longstreth <sup>13,14</sup>; M. Ikram <sup>3,4,15</sup>; S. Seshadri <sup>16,17</sup>; M. Fornage <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA; 2) Brown Foundation Institute of Molecular Medicine, The University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 4) Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 5) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; 6) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 7) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA; 8) Centre for Cognitive Ageing and Cognitive Epidemiology, Psychology, University of Edinburgh, Edinburgh, UK; 9) Institute for Community Medicine, University Medicine Greifswald, Greifswald 17475, Germany; 10) Department of Psychology, University of Edinburgh, Edinburgh, UK; 11) Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Germany; 12) Intramural Research Program, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA; 13) Department of Epidemiology, University of Washington, Seattle, WA, USA; 14) Department of Neurology, University of Washington, Seattle, WA, USA; 15) Department of Neurology, Erasmus MC University Medical Center, Rotterdam, The Netherlands; 16) Department of Neurology, Boston University School of Medicine, Boston, MA, USA; 17) Glenn Biggs Institute of Alzheimer and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, TX, USA

---

Cerebral white matter hyperintensities (WMH) are an indicator of cerebral small vessel disease, a major risk factor for vascular dementia and stroke. DNA methylation (DNAm) may contribute to the molecular underpinnings of WMH, which are highly heritable. We aimed to identify individual CpGs and differentially methylated regions (DMRs) associated with WMH, to investigate their genetic control and to identify potential regulatory variation at known WMH genetic loci using DNAm. We included 6,019 middle-aged to elderly individuals from 11 population-based cohorts, who were free of dementia and stroke and were of African (AA) or European (EA) descent. In each cohort, association between WMH volume and each CpG was tested within ancestry using a linear mixed model, adjusted for age, sex, total intracranial volume, white blood cell count, technical covariates, smoking, body mass index and blood pressures. Meta-analyses were performed within ancestry and in the total sample. We identified DMRs using region-based p-values accounting for spatial correlations among CpGs. Finally, we performed colocalization analyses of methylation quantitative trait loci (mQTLs) and WMH-associated genetic loci using a Bayesian method. No individual CpG reached epigenome-wide significance, but suggestive novel associations were

identified with cg17577122 (*CLDN5*,  $P=2.39E-7$ ), cg24202936 (*LOC441601*,  $P=3.78E-7$ ), cg03116124 (*TRIM67*,  $P=6.55E-7$ ), and cg04245766 (*BMP4*,  $P=3.78E-7$ ). Gene set enrichment analyses implicated pathways involved in regulation of cell development and differentiation, especially of endothelial cells. We identified 11 DMRs (Sidak-adjusted  $P<0.05$ ) including two mapping to blood pressure-related genes (*HIVEP3*, *TCEA2*). We identified 14 DMRs in EA, including one mapping to a stroke gene (*SH3PXD2A*), and 17 DMRs in AA, including 3 mapping to 3 BP genes (*MCF2L*, *PLEC*, and *TM9SF2*). Genes mapping to DMRs were enriched in biological processes related to metabolism and transport of lipoprotein. Finally, the colocalization analysis provided evidence that 4 GWAS loci (2p16, 7q32, 14q32, and 6q25) colocalized with mQTLs.

In conclusion, we identified novel epigenetic loci associated with WMH, which may provide new clues about pathogenesis. We also show evidence for the colocalization of WMH genetic associations and differential DNAm.

## PgmNr 3284: CpG methylation signature define human temporal lobe epilepsy and predict drug-resistant.

### Authors:

B. Xiao; W.B. Xiao; H.Y. Long; L.L. Long; L. Feng

[View Session](#) [Add to Schedule](#)

**Affiliation:** department of neurology, Xiangya Hospital Central South University, Changsha, China

### Purpose—

Based on genome-wide DNA methylation profiling, machine learning methods were applied to screen and verify CpG sites as biomarkers of temporal lobe epilepsy (TLE) diagnosis and drug resistance prediction, providing a new idea for precision treatment of epilepsy.

### Methods—

We initially conducted 450k and 850k methylation chips to detect the DNA methylation of TLE and healthy control in peripheral blood; 450k data were screened for differentially methylated sites, Least Absolute Shrinkage and Selection Operator scored the importance of features, and Support Vector Machine modeled different number of feature combinations, and the optimal model was selected for cross-validation in 850k. The TLE group can be divided into drug-responsive and drug-resistant group, and the same method was used for drug-response related CpGs screening and validation. Furthermore, a Nomogram comprising methylation score and clinicopathological data was generated to establish TLE drug resistance prediction model.

### Key Findings—

As a result, eight CpGs were identified (cg25838818, cg27564766, cg07782795, cg09383187, cg09293614, cg09270525, cg09197288, cg08664849), yielded a sensitivity of 71%, a specificity of 73%, and an accuracy of 77%, with an area under the curve (AUC) of 0.81 to detect TLE. Six CpGs were identified (cg15999964, cg08768218, cg11954680, cg17706086, cg21761639, cg26119877), yielded a sensitivity of 77%, a specificity of 71%, and an accuracy of 73%, with an AUC of 0.79 to distinguish drug-responsive from drug-resistant patients. TLE drug resistance prediction model established by Nomogram included hippocampal sclerosis pathology, epilepsy course, methylation score.

### Significance—

This study introduced a methodological framework to screen and validate biomarker, and demonstrates the ability to use machine learning as a potential clinical tool for epilepsy diagnosis and drug-resistant prediction after more comprehensive validation.

## **PgmNr 3285: Expression of hnRNP A2/B1 suppresses rCGG repeat-mediated neuronal toxicity associated with fragile X-associated tremor/ataxia syndrome in a RAN translation-independent manner.**

### **Authors:**

Y. Kang <sup>1</sup>; L. Huang <sup>1</sup>; K. Zhang <sup>1</sup>; R. Hukema <sup>2</sup>; D. Nelson <sup>3</sup>; P. Jin <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Human genetics, Emory University, atlanta, Georgia.; 2) Department of Clinical Genetics, Erasmus MC, 3015 Rotterdam, the Netherlands; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

---

FXTAS (Fragile X-associated tremor/ataxia syndrome) is a neurodegenerative disorder in aged premutation carriers with 55–200 CGG repeats in the 5'UTR of fragile X mental retardation gene (*FMR1*). Two mechanisms have been proposed for FXTAS: RNA toxicity and non-AUG-initiated (RAN) translation toxicity. Expanded CGG repeat RNA is reported to sequester proteins such as Pur $\alpha$  and hnRNP A2/B1 (heterogeneous nuclear ribonucleoprotein A2/B1), thereby blocking them from performing from their proper function. Recently, polyglycine-containing protein (FMRpolyG) generated by RAN translation of the expanded CGG repeats was shown to be toxic to cells. Here, we have generated a transgenic mouse line that expresses hnRNP A2/B1 in Purkinje cells, and we have crossed this line with the FXTAS mouse model that we generated previously. We found that the expression of hnRNP A2/B1 can ameliorate the Purkinje cell loss and locomotor deficits induced by the expression of the expanded CGG repeat. Gene expression analyses further confirmed the molecular rescue by hnRNP A2/B1 in the FXTAS mouse model. Surprisingly, pathological analyses revealed that the distribution and number of FMRpolyG inclusions were unchanged in the rescued mice. Furthermore, using a FXTAS cell culture model, we observed a similar suppression of rCGG repeat toxicity by hnRNP A2/B1 without altering the distribution and number of FMRpolyG inclusions. These results together suggest that hnRNP A2/B1 has an important neuroprotective function by alleviating RNA toxicity, and FMRpolyG is not essential for the rCGG repeat toxicity associated with FXTAS.

## PgmNr 3286: Cell type, genetic content and presence of the 3'UTR influence TREM2 promoter activity.

### Authors:

L. Bekris<sup>1</sup>; M. Khrestian<sup>1</sup>; G. D'Aleo<sup>1</sup>; S. Formica<sup>1</sup>; Y. Shao<sup>2</sup>; J. Leverenz<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genomics Medicine Institute, Cleveland Clinic Lerner Research Institute, Cleveland, Ohio.; 2) Cleveland Clinic Lou Ruvo Center for Brain Health, Cleveland Clinic, Cleveland, Ohio

---

Genetic variation in triggering receptor expressed on myeloid cells 2 (TREM2) is associated Alzheimer's disease (AD) and plays a role in neuroinflammation in AD mouse models. Little is known about how the expression of TREM2 is differentially regulated. In the human brain, TREM2 mRNA is differentially expressed in multiple brain regions and human cell types. We hypothesized that trans-acting factors present in certain cell types modulate TREM2 regulatory element activity. Therefore, TREM2 expression or promoter activity will differ according to tissue or cell type. The objectives were to demonstrate that: 1) Cerebrospinal fluid (CSF) and plasma sTREM2 levels are elevated in AD. 2) Cell line, hippocampus and cerebellum TREM2 protein expression varies by tissue type. 3) TREM2 promoter region genetic variants and 3' UTR influence to TREM2 promoter reporter expression in certain human cell lines. Human DNA, plasma and CSF samples were obtained from the Cleveland Clinic Center for Brain Health Aging and Neurodegeneration Biobank. Post-mortem brain samples were obtained from the University of Washington Alzheimer's Disease Research Center. CSF and plasma soluble TREM2 (sTREM2) were evaluated using a custom capture sandwich immunoassay. Brain TREM2 was analyzed by Western blot. PCR amplification of genomic DNA was performed using TREM2 primers located outside of the promoter or 3'UTR region. TREM2 PCR products were cloned into a Luciferase reporter constructs. A transient transfection using these reporter constructs was performed in human cell lines. CSF sTREM2, but not plasma, levels were marginally higher in carriers of the AD GWAS rs7748513 risk allele suggesting that TREM2 genetic variation might influence gene expression depending on tissue type. Post-mortem brain and multiple human cell lines also showed differential expression by tissue or cell type. TREM2 promoter + 3'UTR reporter levels showed differential expression depending on genetic content of the promoter in human neuronal SHSY5Y cell lines. TREM2 promoter reporter inhibition in the presence of the TREM2 3'UTR was cell type specific. This inhibition was lost in the presence of genetic variants located in the TREM2 promoter region suggesting that SHSY5Y cells contain inhibitory factors that target this region. This finding is important because it will allow us to identify factors that regulate TREM2 expression and therefore potentially modulate neuroinflammation in AD.



## PgmNr 3287: *HLA-DRB1\*1501* gene expression is modulated by rs4959030 (A) in multiple sclerosis patients.

### Authors:

C. Khunsriraksakul; T. Salameh; R. Kim; B. Zhang; H. Yang; K. Sheldon; L. Schneper; B. Zhang; G. Thomas; J. Broach; F. Yue

[View Session](#) [Add to Schedule](#)

**Affiliation:** Pennsylvania State University College of Medicine, Hershey, PA

---

Multiple sclerosis (MS) is the most common chronic immune-mediated disorder of the central nervous system (CNS) in young adults, characterized by demyelinating lesions that occur across time and space. Multiple large-scale genome-wide association studies (GWAS) have identified over 200 risk variants related to MS; however, all identified risk variants together could only explain ~28% of the sibling recurrence risk. Only *HLA* locus (e.g. *HLA-DRB1\*15:01* allele, *HLA-A\*02* allele) has been repeatedly verified to be associated with the pathogenesis of MS. The role of *HLA* genes in autoimmune diseases has been largely studied in terms of MHC antigenic specificity; yet, much less is known about how these genes are transcriptionally regulated. We hypothesized that the aberrant epigenetic regulation in MS patients may contribute to the pathogenesis of this disease. Here, we recruited 246 MS patients and utilized both whole-exome sequencing (WES) and ImmunoArray genotyping to maximally interrogate the intragenic and intergenic regions of the genome in a cost-effective manner. Both technologies combined identified over half a million SNVs, in which 70.1% and 20.8% of those SNVs located within the intragenic and intergenic regions, respectively. Our pathway enrichment analysis confirmed the complexity of MS pathogenetic mechanisms. Additionally, epigenetic profiling further provided a deeper understanding of potential transcriptional dysregulatory mechanisms in MS. Specifically, we identified rs4959030, which is located within a super-enhancer region nearby *HLA-DRB1* gene, as a potential risk SNV that can modulate the expression of *HLA-DRB1\*1501* through the disruption of transcription factor PU.1 binding motif. We further confirmed the computational prediction with quantitative real-time PCR in whole blood of 30 MS patients. We found that the expression of *HLA-DRB1* was, indeed, decreased by 30% ( $P = 0.0002$ ) when patients have homozygous high-risk allele (A) in comparison to that of homozygous low-risk allele (G). This finding highlighted the importance of understanding epigenetic regulation in MS and may open up new possibilities in personalized health care for MS patients.

## **PgmNr 3288: Is Down syndrome an example of multiple mechanisms of gene dose Dysregulation at Critical Region (DSCR) on chromosome 21? A neurogenomics approach.**

### **Authors:**

A. Sanchez-Gomez; F. Garcia-Vallejo; A. Rodriguez; J.M. Satizabal-Soto

[View Session](#) [Add to Schedule](#)

**Affiliation:** Physiological Sciences, Universidad del Valle, Cali, Valle, Colombia

---

The classical etiological interpretation on Down syndrome (DS), is an aneuploidy caused by a complete, or occasionally partial, triplication of chromosome 21 resulting in a complex and variable phenotype which compromise the structure and function of brain. However no many approaches look for a different explanation about the complexity of its neurological phenotype. The objective of the present study was to analyze the dysregulation of dose in genes located at Down Syndrome Critical Region (DSCR) along of several structures of brain. The searching of DNA microarrays experiments with Down syndrome brain samples in the GEO Data Set of NCB, produced one which fit on our experimental design; it contains data of global gene expression in 11 structures of brain cortex, hippocampus and cerebellum from patients with Down syndrome and their euploid controls. We calculated DS/Control ratio of 40 DSCR genes from the intensity log2 values obtained of the selected DNA microarray experiment. Differential transcriptional profiles, lead us to postulate that three main different mechanisms of DSCR gene dose dysregulation in brain are contributing to partially explain the Down neurophenotype. These mechanisms include: unbalance of the gene dose (triplication and/or single expression); under and over expression, and gene amplification. Our results support the hypothesis of complex spatial and temporal mechanisms regulate the gene dose imbalance in the brain of Down syndrome.

## PgmNr 3289: Screening Alzheimer disease serum for long noncoding RNA biomarkers.

**Authors:**

E. Dreskin; M. Hammond; R. Drevland; Y. Jouvenot

[View Session](#) [Add to Schedule](#)

**Affiliation:** Bio-Rad Laboratories, Hercules, CALIFORNIA.

---

Long noncoding RNAs have become the fastest growing class of biomarkers in genomics research. They have been demonstrated to play roles in normal development mechanisms as well as pathological processes, such as cancer, autoimmune diseases and neurodegenerative diseases. In order to find new and relevant biomarkers for Alzheimer's Disease, we designed an approach to quantify changes in expression of long noncoding RNA targets from patient serum. We purified cell free RNA from a portion of each serum sample, and found that the majority of samples did have low amounts of RNA, although a few were too degraded for further analysis. Using a one-step pre-amplification reverse transcription qPCR strategy, we screened for differential expression of long non coding RNAs proposed to be associated with inflammation and epigenetics. Alzheimer's Disease is proposed to be a disease of inflammation, and disease states are often the result of epigenetic dysregulation. Of the 145 unique long noncoding RNAs screened, we identified four with a greater than two fold expression change over healthy controls that were deemed significant ( $p < 0.05$ ) by ANOVA, and had normal sample distributions by the Shapiro-Wilks normality test. All four of these lncRNAs (AC093323.3, BDNF-AS, H19 and HAR1B) are linked to epigenetic processes.

## PgmNr 3290: Chromatin interaction map elucidates regulatory relationships in human interneuron-like progenitors that are associated with atypical Rett syndrome.

### Authors:

R.Y. Birnbaum<sup>1</sup>; R. Eshel<sup>1</sup>; E. Farhi<sup>1</sup>; E. D'haene<sup>2</sup>; S. Vergult<sup>2</sup>; T. Kaplan<sup>3</sup>; I.C. Harabula<sup>4</sup>; A. Kukalev<sup>4</sup>; A. Pombo<sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Ben-Gurion University of the Negev, Department of Life Sciences, Beer-Sheva, Israel.; 2) Center for Medical Genetics, Ghent University, Belgium; 3) The Hebrew University in Jerusalem, School of Computer Science and Engineering, Jerusalem, Israel; 4) 4Berlin Institute for Medical Systems Biology. Max-Delbrück-Centrum für Molekulare Medizin, Berlin, Germany

---

Gene regulatory elements such as enhancers dynamically regulate gene expression in a tissue-specific manner. However, the transcriptional regulatory elements during human inhibitory interneuron differentiation and their role in neurodevelopmental disorders are unknown. Here, we generated gene regulatory element maps of human inhibitory-like interneurons derived from embryonic stem cells (H9-ESC), permitting large-scale annotation of previously uncharacterized interactions of regulatory elements that associated with neurodevelopmental disorders. Using histone modification marks of active enhancers (H3K27ac, ATAC-seq), we identify neuronal progenitor enhancers that are associated with cell-specific gene expression (RNA-seq), especially of neuronal transcription factors. Using genome architecture mapping (GAM), we determined the chromatin interactions of regulatory elements and their potential targeted genes. Furthermore, chromatin interactions in *FOXP1* locus unraveled promoter-enhancer looping that likely associated with atypical Rett syndrome. Using *in vivo* enhancer assay, we identified 8 enhancers in the *FOXP1* locus with activity patterns that resembled *FOXP1* expression. Using CRISPRCas9 genome editing, two *FOXP1* enhancers were deleted (but not the *FOXP1* gene body) and showed significant reduction in *FOXP1* expression and alteration of cell proliferation in human cell line. Furthermore, a microdeletion proximal to *FOXP1* encompassing these neuronal *FOXP1* enhancers was found in patient with atypical Rett syndrome, supporting the role of *FOXP1* enhancers in this syndrome. Our study provides a framework for understanding the impact of non-coding regulatory elements during inhibitory interneuron differentiation, and highlights novel mechanisms underlying neurodevelopmental disorders.

## PgmNr 3291: Transcriptomic analyses in a diverse ancestry cohort implicate convergent immune related pathways in Alzheimer disease.

### Authors:

A.J. Griswold <sup>1,2</sup>; F. Rajabli <sup>1</sup>; D. Van Booven <sup>1</sup>; K. Hamilton <sup>1</sup>; V. Rodriguez <sup>1</sup>; L.D. Adams <sup>1</sup>; T.D. Starks <sup>1</sup>; M.L. Cuccaro <sup>1,2</sup>; J.M. Vance <sup>1,2</sup>; W.S. Bush <sup>3,4</sup>; G.S. Byrd <sup>5</sup>; B.E. Feliciano-Astacio <sup>6</sup>; A.R. Betancourt <sup>6</sup>; H. Acosta <sup>7</sup>; A. China <sup>8</sup>; J.L. Haines <sup>3,4</sup>; G.W. Beecham <sup>1,2</sup>; M.A. Pericak-Vance <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Population & Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH; 4) Cleveland Institute for Computational Biology, Cleveland, OH; 5) Maya Angelou Center for Health Equity (MACHE) / Wake Forest School of Medicine, Winston Salem, NC; 6) Universidad Central del Caribe, Caguas, Puerto Rico; 7) Clinica de la Memoria, San Juan, Puerto Rico; 8) Universidad Central del Caribe, Bayamon, Puerto Rico

**Background.** Our group and others have recently expanded genomic studies of Alzheimer disease to individuals of diverse ancestries, however, identification of downstream effects on gene regulation is lacking, particularly in under-represented groups such as African Americans (AA) and Caribbean Hispanics from Puerto Rico (PR). Understanding these transcriptomic effects is critical in discovery of both ancestry-specific and convergent functional mechanisms in AD. In this study, we analyzed the protein coding transcriptome from peripheral blood collected from 76 PR (39 AD, 37 controls), 234 African American (AA) (115 AD, 119 controls), and 241 non-Hispanic Whites (NHW) (121 AD, 120 controls).

**Methods.** Total RNA was extracted from peripheral whole blood, sequenced to 40 million paired-end 100bp reads, and analyzed with a standard bioinformatics pipeline. Linear regression models were used to test differentially expressed genes between cases and controls within each ethnicity adjusting for covariates including age, ascertainment centers, sex, and sequencing batch as required. Gene Set Enrichment Analysis of Gene Ontology and KEGG gene sets were used to identify underlying biological pathways.

**Results.** Using the criteria of adjusted p-value  $\leq 0.05$ , a total of 761 (518 up-regulated, 243 down-regulated), 419 genes (291 up-regulated, 127 down-regulated) and 490 genes (353 up-regulated, 137 down-regulated) were differentially-expressed between AD and controls in the PR, AA, and NHW datasets, respectively. Interestingly, only 25 genes (18 up-regulated, 7 down-regulated) were differentially expressed in both PR and AA, 16 genes (15 up-regulated, 1 down-regulated) were commonly differential in AA and NHW, and no genes were differentially expressed in both PR and NHW. Despite the generally unique set of differentially expressed genes, pathway analysis revealed that adaptive and innate immune response pathways were enriched in all three groups.

**Conclusions.** These results support the hypothesis that there is genetic heterogeneity for AD across

ancestral groups, but suggest convergent underlying processes. Further evaluation of genetic modulation of specific transcriptomic changes using ancestry-aware methods will uncover underlying mechanisms of the genomic regulation of AD pathophysiology.

## **PgmNr 3292: The expanded GAA triplet-repeat in Friedreich ataxia establishes a mutation-specific differentially methylated region (FRDA-DMR) that is predictive of *FXN* gene silencing and reactivation.**

### **Authors:**

S.I. Bidichandani; L. Rodden; K. Gilliam; Y. Chutake

[View Session](#) [Add to Schedule](#)

**Affiliation:** Dept of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma.

---

Transcriptional silencing caused by the expanded GAA triplet repeat mutation in intron 1 of the *FXN* gene is the primary molecular defect in Friedreich ataxia (FRDA). We found that patients who are homozygous for the expanded GAA triplet-repeat establish a region of DNA hypermethylation in its vicinity. This region remains largely unmethylated in non-FRDA controls. The FRDA-specific differentially methylated region (FRDA-DMR) is mutation-dependent as hypermethylation was completely reversed in isogenic cells with the expanded repeat removed. Hypermethylation of the FRDA-DMR was observed in disease-relevant tissues (dorsal root ganglia, cerebellum and heart) of the humanized mouse model of FRDA, and in iPSC-derived FRDA neurons. Analysis of PBMCs from 50 FRDA patients revealed a strong correlation between DNA methylation and *FXN* gene silencing. Small molecule HDAC inhibitors are currently being developed for *FXN* gene reactivation in FRDA. PBMCs from the same 50 patients were tested for response to *ex vivo* HDAC inhibitor treatment, which revealed that lower levels of FRDA-DMR methylation correlated with superior gene reactivation in FRDA. Indeed, the same CpG sites that correlated with *FXN* transcriptional silencing in FRDA were also the best predictors of *FXN* gene reactivation. Repeat-mediated hypermethylation of the FRDA-DMR is a novel epigenetic silencing signal in FRDA.

## PgmNr 3293: Cell type and brain region-specific differential chromatin accessibility analysis in Alzheimer's disease.

### Authors:

J. Bendl<sup>1,2</sup>; M.E. Hauberg<sup>4,5</sup>; J.F. Fullard<sup>1,2</sup>; K. Girdhar<sup>1,2</sup>; G.E. Hoffman<sup>1,2</sup>; R. Misir<sup>1,2</sup>; S.M. Reach<sup>1,2</sup>; V. Haroutunian<sup>3</sup>; P. Roussos<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Genetics and Genomic Sciences, Icahn Institute of Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA; 2) Department of Psychiatry and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, USA; 3) Mental Illness Research, Education, and Clinical Center (VISN 2 South), James J. Peters VA Medical Center, Bronx, NY, USA; 4) Department of Biomedicine, Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark; 5) The Lundbeck Foundation Initiative of Integrative Psychiatric Research (iPSYCH), Denmark

---

**Background:** Regions of open chromatin house regulatory elements required to mediate cell-type and tissue-specific gene expression. Studies of human brain have shown that dysregulation of these regulatory mechanisms is associated with Alzheimer's Disease (AD). The majority of previous studies, however, has been conducted using homogenate tissue derived from a single brain region and have included small numbers of AD cases and controls. Here, we present the largest cell type and brain region-specific study of differential chromatin accessibility in AD.

**Methods:** Using frozen postmortem tissue from 209 cases with AD and controls, we performed ATAC-seq to profile chromatin accessibility in 2 distinct populations of cells (neuronal and non-neuronal), isolated by FACS from two different brain regions (parahippocampal gyrus and superior temporal cortex). We characterized epigenetics changes associated with multiple AD phenotype ratings, i.e., clinical dementia rating (CDR), density of neurofibrillary tangles (Braak&Braak score), the average count of senile plaques as well as a linear version of the CERAD criteria. The availability of RNA-seq and whole-genome data for this cohort allowed us to measure the overlap between differentially regulated transcriptome and epigenome signatures and related pathways and also run fine mapping.

**Results:** We observed widespread differences in chromatin accessibility associated with AD phenotype, mostly in neuronal samples. While substantial changes between AD-cases and controls can be seen in both brains regions of interest, parahippocampal gyrus is, in general, more affected than superior temporal cortex. The differentially regulated regions of open chromatin were enriched in neurobiological and disease-related pathways. Finally, by integrating this dataset with whole-genome sequencing data, we identified thousands of new chromatin accessibility quantitative trait loci (caQTLs).

**Conclusions:** This dataset provides a unique insight into molecular mechanisms underlying brain region and cell type-specific vulnerability to AD at different stages of the disease progression. Supported by R01AG050986 and R01AG057440.



# PgmNr 3294: The relationship between subunits of a protein complex and gene co-expression networks in the human brain.

## Authors:

M. Xu <sup>1</sup>; C. Jiao <sup>2</sup>; Y. Chen <sup>1</sup>; S. Liu <sup>1</sup>; J. Tian <sup>1</sup>; C. Chen <sup>1,3</sup>; C. Liu <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Central South University, Changsha, Hunan, China; 2) SUNY Upstate Medical University, Syracuse, NY, USA; 3) Xiangya Hospital, Central South University, Changsha, China

---

**Background:** Protein complexes work synergistically as a basic functional unit within cells. Interactions among the subunits that constitute protein complexes reveal specific regulatory mechanisms critical for cellular function. Protein complexes are currently identified by expensive and time-consuming methods (i.e., two-hybrid screening, mass spectrometry, and glutathione S-transferase pull-down assay). We posit that gene co-expression, which implies coordinate regulation, can be a viable alternative for the study of gene interactions. Recent studies have shown that the subunits of protein complexes express in similar patterns, indicating that gene co-expression networks may reflect gene interaction at the protein level. However, the overall concordance between the subunits of a protein complex and gene co-expression network has not been assessed. The systematic evaluation of the co-expression of protein subunits may help to identify specific genes that work in concert.

**Methods:** We constructed a gene co-expression network by weighted gene co-expression network analysis (WGCNA) using the RNA-seq data from the Genotype-Tissue Expression project (GTEx). Protein complex data originating from CORUM was mapped into gene co-expression networks. We used four other independent datasets to replicate our findings and defined protein complexes co-expressed in the four datasets as stably co-expressed.

**Results:** Of the 2,390 protein complexes in the human brain, 36% were found to contain at least two co-expressed subunits when compared with random modules ( $P < 0.05$ ). In the frontal cortex, 43% of the complexes contained at least two co-expressed subunits. Of those, 243 were stably co-expressed, relating to basic cellular functions such as ribosomal function and PI3K-Akt signaling. When comparing stably co-expressed protein complexes with all other proteins, we found that the protein complex subunits were over-presented in the other proteins. (one-tail t-test  $P = 1.07 \times 10^{-69}$ ). The subunits of 210 stably co-expressed protein complexes, such as the psychiatrically-associated SNARE complex, were differentially expressed in autism spectrum disorder ( $FDR = 4.2 \times 10^{-3}$ ) and schizophrenia ( $FDR = 1.8 \times 10^{-3}$ ).

**Conclusion:** Our study revealed a consistency between protein complexes and gene co-expression networks. The finding that co-expressed protein subunits are differentially expressed in psychiatric disorders suggests a new mechanism for the unraveling of their pathology.

## PgmNr 3295: Gene network analysis reveals unique and shared modules associated with alcohol dependence within the PFC and NAc.

### Authors:

E. Vornholt <sup>1,2</sup>; G.O. McMichael <sup>1</sup>; J. Drake <sup>2</sup>; Z.N. Taylor <sup>1</sup>; V.I. Vladimirov <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA.; 2) Integrative Life Sciences, Virginia Commonwealth University, Richmond, VA; 3) Department of Psychiatry, Virginia Commonwealth University, Richmond, VA; 4) Lieber Institute for Brain Development, John Hopkins University, Baltimore, MD

---

Excessive alcohol consumption has become a growing public health concern worldwide due to the potential development of alcohol dependence (AD). Prolonged alcohol abuse leads to dysregulation of the mesocorticolimbic dopaminergic pathway (MCL) effectively disrupting executive functioning and the allostatic conditioning of reward response. We utilized a human postmortem brain case-control study design to identify differentially expressed genes (DEG) and co-expressed gene networks associated with AD within two different MCL associated structures, the prefrontal cortex (PFC) and nucleus accumbens (NAc). We used weighted gene co-expression network analysis (WGCNA) to identify gene modules significantly correlated with AD in PFC and NAc and further test whether networks are preserved between the two regions. Finally, a gene ontology analysis was used to reveal the biological pathways of significant AD associated gene networks. Gene expression data from 17 AD cases and 18 matched controls ( $n = 35$ ) were obtained via the Affymetrix GeneChip Human Genome U133A 2.0 (HG-U133A 2.0) array and analyzed using a bidirectional stepwise regression. We identified 1841 and 70 DEG (*FDR adjusted*  $p \leq 0.05$ ) in the NAc and PFC, respectively. Using WGCNA on the nominally significant DEG ( $p \leq 0.05$ ) from each brain region we identified 6 and 3 modules (categorically labeled using conventional colors) which were significantly associated with AD (*Bonferroni corrected*  $p \leq 0.05$ ) in the NAc and PFC, respectively. The network preservation analysis identified some of the significant modules to be highly preserved between NAc and PFC, i.e. the magenta module in NAc ( $Z_{summary\_magenta} = 13.0$ ), while others showed little to no network preservation, i.e. the purple ( $Z_{summary\_purple} = 0.55$ ) and dark orange ( $Z_{summary\_dark\ orange} = 0.83$ ) modules. Gene ontology analyses via DAVID Bioinformatics Resource 6.8 revealed the magenta module is primarily associated with immune response and more specifically metallothioneins (*FDR adjusted*  $p \leq 0.05$ ) whereas the purple and dark orange modules contain genes associated with neuronal cell signaling pathways (*FDR adjusted*  $p \leq 0.05$ ) and cell protrusion (*FDR adjusted*  $p \leq 0.05$ ). The upregulation of immune response mechanisms has been consistently reported to be the result of alcohol neurotoxic effects. More interestingly, these results show that the effects of chronic alcohol use on genes important for neuronal signaling and synaptic growth may be brain region specific.

## PgmNr 3296: Cell-type specific investigation of the methylome and hydroxymethylome in the prefrontal cortex of alcohol dependent patients.

### Authors:

S.L. Clark <sup>1</sup>; R.F. Chan <sup>2</sup>; L.Y. Xie <sup>2</sup>; M. Zhao <sup>2</sup>; K.A. Aberg <sup>2</sup>; E.J.C.G. van den Oord <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Psychology, Michigan State University, East Lansing, MI; 2) Center for Biomarker and Precision Medicine, Virginia Commonwealth University, Richmond, VA

Alcohol dependence (AD) is characterized by compulsive and uncontrolled consumption of alcohol despite its negative effects. Mounting evidence suggests a role for DNA methylation in AD. The brain methylome includes both methylation at CG dinucleotides (mCG) and hydroxymethylation (hmCG). We present a methylome-wide association study (MWAS) of both mCG and hmCG, across cell types in human prefrontal cortex (PFC) samples from 50 AD cases and controls. With a sequencing-based approach, we used optimized MBD-seq and HmSeal-seq protocols to study the near complete mCG and hmCG methylomes. FACS was used to sort NeuN+ (neurons) and NeuN- (mainly glia) nuclei from PFC to generate reference methylomes that, in combination with a deconvolution approach, enabled us to perform cell-type specific MWAS in neurons and glia separately. RNA-seq was performed on the same samples to examine the functional impact of mCG/hmCG on gene expression.

For mCG, in bulk brain tissue, the top findings were located in *GOSR2* ( $p=3.9 \times 10^{-07}$ ) and in a replicated gene associated with alcohol consumption, *AUTS2* ( $p=4.6 \times 10^{-07}$ ). Cell-type specific findings revealed further associations with *AOX1* ( $p=1.6 \times 10^{-07}$ ) in glia and *SMAD4* in neurons ( $p=2.4 \times 10^{-08}$ ). Top findings were enriched for bivalent transcription start sites ( $p=7.3 \times 10^{-04}$ ) and depression-related CG methylation in human PFC ( $p=4.3 \times 10^{-03}$ ).

For hmCG, the top finding in bulk brain tissue was located in *SPAST* ( $p=1.5 \times 10^{-07}$ ). In glia, the top result was located in *ACSS3* ( $p=4.9 \times 10^{-08}$ ) that encode acetyl-CoA synthetase, an enzyme that processes the ethanol metabolite acetate. The top genic finding in neurons was located in the autophagy-related *ATG7* ( $p=5.0 \times 10^{-08}$ ). As autophagy is induced by ethanol exposure, differences in hydroxymethylation may indicate upregulation of autophagy. Top findings were enriched for active transcription start sites ( $p=5.0 \times 10^{-05}$ ) and DNase hypersensitive sites ( $p=7.0 \times 10^{-07}$ ).

The work presented here provides a comprehensive consideration of the methylomic and hydroxymethylomic landscape of AD in human prefrontal cortex. These analyses provide potential mechanisms for confirmed genetic loci previously associated with AD and revealed several potentially novel loci. Preliminary gene expression analyses indicate that several associated genes identified here are differentially expressed in AD cases vs controls.

## PgmNr 3297: Dose and time dependency of the *AHRR* cg05575921 demethylation response in nascent smoking.

### Authors:

K. Dawes <sup>1,2</sup>; E. Papworth <sup>1</sup>; A. Becker <sup>1</sup>; W. Philibert <sup>1</sup>; B. Hundley <sup>1</sup>; N. Hutchens <sup>1</sup>; F. Gibbons <sup>4</sup>; M. Gerrard <sup>5</sup>; R. Philibert <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry, University of Iowa, Iowa City, Iowa.; 2) Department of Molecular Medicine, University of Iowa, Iowa City, Iowa.; 3) Behavioral Diagnostics, Coralville, Iowa.; 4) Department of Psychological Sciences, University of Connecticut, Storrs, Connecticut.; 5) Department of Psychology, Iowa State University, Ames, Iowa.

---

In prior work, we and others have shown that methylation of cg05575921 of whole blood DNA accurately predicts smoking status of adults and is sensitive to nascent smoking. Unfortunately, the dose and time dependency of the demethylation response in nascent is not known. In order to determine this relationship, we interviewed 300 high school sophomores biannually (5 time points) and determined yearly methylation, cotinine and cannabinoids levels annually (3 time points). At intake, 29 subjects reported using tobacco and 39 subjects reported using cannabis at least once in their life with 19 and 11 subjects having a positive cotinine and cannabinoid levels, respectively. At exit, 15 subjects reported using tobacco and 89 subjects reported using cannabis at least once in their life with 52 and 37 subjects having positive cotinine and cannabinoid levels at exit, respectively. The rates of unreliable self-report of tobacco or other nicotine products in those positive for cotinine was 74% at intake and 77% at exit. Those 149 subjects had undetectable cotinine levels at all three time points had cg05575921 methylation of 88.67% at entry and 88.84% at exit, while those 12 subjects who had a positive cotinine levels at all time points had an entry methylation levels of 81.84% and exit levels of 71.31%. In those with reliable levels of self-report the amount of demethylation at cg05575921 was correlated with time and intensity of smoking. We conclude that cg05575921 methylation is a sensitive indicator of nascent smoking and suggest that screening of blood or saliva for changes in cg05575921 could identify smokers in the most malleable phases of smoking.

## PgmNr 3298: Methylation sensitive digital PCR tests for rapidly assessing heavy alcohol consumption and smoking using DNA from blood or saliva.

### Authors:

R. Philibert <sup>1,2</sup>; M. Dogan <sup>1,2</sup>; J. Mills <sup>2</sup>; J. Long <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept Psychiatry, Univ Iowa, Iowa City, Iowa.; 2) Behavioral Diagnostics, Coralville, IA

---

**Background:** Smoking and Excessive Drinking are the 1<sup>st</sup> and 3<sup>rd</sup> leading causes of mortality in the world. Unfortunately, there are no existing clinically implementable tests for quantifying the extent of either lifestyle factor. In 2012 and 2014, respectfully, we published the first genome wide study of smoking and heavy alcohol consumption (HAC). Herein, we detail the translation of those genome wide findings to easy to perform quantitative digital PCR (dPCR) tests for both smoking and drinking using DNA from blood or saliva from 233 smokers, 143 HAC and 200 controls. **Results:** Using DNA from whole blood, we show that dPCR assessment of methylation at cg05575921 has a Receiver Operating Characteristic (ROC) area under the curve (AUC) of 0.995 for predicting smoking with the amount of demethylation at cg05575921 being proportional to daily cigarette consumption. Similarly, a set of 4 dPCR markers specific for HAC has an AUC of 0.95 using DNA from whole blood. By using a recently described dPCR marker capable of correcting for the cellular heterogeneity in saliva, similar AUC values and dose response curves were seen with respect to cigarette and alcohol consumption using DNA prepared from saliva. **Conclusions:** We conclude that clinically implementable dPCR tools using DNA from blood or saliva can sensitively detect the presence and intensity of both smoking and HAC. Since the methylation signatures for both alcohol and cigarette consumption revert with abstinence, we believe that these easy to perform, rapid digital PCR tools will be useful to clinicians and researchers in assessing and monitoring those enrolled in substance use disorder treatment, employee wellness programs and insurance underwriting.

## PgmNr 3299: Epigenome-wide analysis uncovers a blood-based DNA methylation biomarker of lifetime cannabis use.

### Authors:

C.A. Markunas<sup>1</sup>; D.B. Hancock<sup>1</sup>; Z. Xu<sup>2</sup>; B.C. Quach<sup>1</sup>; D.P. Sandler<sup>2</sup>; E.O. Johnson<sup>1,3</sup>; J.A. Taylor<sup>2,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Omics Discovery and Epidemiology, RTI International, Research Triangle Park, NC, USA; 2) Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA; 3) Fellow Program, RTI International, Research Triangle Park, NC, USA; 4) Epigenetic and Stem Cell Biology Laboratory, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA

---

**Background:** Cannabis, the most commonly used illicit drug in the US, is associated with adverse and beneficial effects. Biomarkers that accurately classify cannabis use would facilitate research on potential health consequences. DNA methylation (DNAm) is an excellent candidate biomarker, yet no blood-based epigenome-wide association studies (EWAS) exist.

**Methods:** We conducted an EWAS using blood-based DNAm data (Illumina 450K) from a case-cohort study within Sister Study, a prospective cohort of women at risk of developing breast cancer (Discovery N=1,730; Replication N=853). We ran robust linear regression models of DNAm by lifetime cannabis (ever versus never) use, adjusting for age, breast cancer, tobacco and alcohol use, technical factors, and blood cell-type proportions. A multi-CpG classifier of lifetime cannabis use was developed using penalized regression of top EWAS CpGs in the discovery sample and validated in the replication sample. We also examined published *cis*-methylation quantitative trait loci (*cis*-meQTLs) for reported lifetime cannabis use-associated SNPs in relation to our EWAS findings.

**Results:** We identified and replicated an association with lifetime cannabis use at cg15973234 (CEMIP): combined  $P=3.3 \times 10^{-8}$ . We found no overlap between published cannabis-associated SNPs ( $P < 0.05$ ) and blood-based *cis*-meQTLs of cg15973234. A 3-CpG classifier of lifetime cannabis use had an area under the receiver operating characteristic curve of 0.79 (95% confidence interval [0.76, 0.82];  $P < 2.2 \times 10^{-16}$ ) in the replication sample.

**Conclusions:** Our findings provide evidence that blood-based DNAm is associated with lifetime cannabis use. DNAm levels at cg15973234 were not associated with reported genetic risk factors for lifetime cannabis use, suggesting that the observed DNAm difference was driven by cannabis exposure.

## PgmNr 3300: Genetic, epigenetic, and pharmacologic influences on gene expression regulation in iPSC-derived neural progenitor cells.

### Authors:

F. Gordovez <sup>1,2</sup>; S.D. Detera-Wadleigh <sup>1</sup>; D. Hernandez <sup>3</sup>; N. Akula <sup>1</sup>; L. Kassem <sup>1</sup>; W. Corona <sup>1</sup>; A. Smith <sup>4</sup>; J. Nevado <sup>2</sup>; F.J. McMahon <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Human Genetics Branch, National Institute of Mental Health, Bethesda, Maryland.; 2) College of Medicine, University of the Philippines Manila, Manila, Philippines; 3) Laboratory of Neurogenetics, National Institute of Neurological Diseases and Stroke, Bethesda, Maryland; 4) National Human Genome Research Institute, Bethesda, Maryland

---

Neurodevelopmental programs depend upon dynamic changes in gene expression, but the underlying regulatory mechanisms are poorly understood. We are developing an experimental platform based on iPSC-derived neural cells designed to investigate the relative roles of genetic variation, epigenetics, and pharmacologic treatment on neurodevelopment in vitro. Here we sought to pilot the platform in neural progenitor cells (NPC), with and without treatment with valproic acid (VPA), a commonly used CNS drug and neural teratogen whose impact on DNA methylation remains controversial.

Fibroblasts from 6 unrelated donors were reprogrammed to iPSC, differentiated into NPCs, and treated 4 wks later with 1mM VPA (or medium only) for 5d. DNA and RNA were extracted on day 5. DNA bisulfite-converted DNA was analyzed with the Illumina MethylationEPIC array and Minfi pipeline. Methylation was expressed as percent intensity (B). RNAseq was done using NovoSeq 6000 with 40M, 150 bp, paired-end reads per sample. Mapped reads were variance stabilized and analyzed for differential expression through DESeq2. For each sample, methylation and expression levels were compared in treated and untreated conditions. EnrichR was used for functional enrichment analysis of differentially methylated (DM) or differentially expressed (DE) genes.

The methylation array reliably detected 830,000 CpG sites mapped to 34,440 genes. As expected, there was significant correlation between B values and gene expression ( $r^2=0.06$ ) that increased near the transcription start site ( $r^2 = 0.20$ ) and was not changed by VPA treatment ( $r^2=0.22$ ).

After VPA treatment, 2108 sites in 860 genes were DM and 502 genes were DE at FDR 5%, with only chance overlap at the gene level. Consistent with these results, only 1% of variance in B-values could be attributed to VPA treatment. Functional enrichment analysis of DM genes revealed ectoderm differentiation and axon guidance, while DE genes were enriched for genes previously associated with VPA treatment. There was no overlap.

These results confirm the impact of local CpG methylation on gene expression and support known changes in gene expression caused by VPA exposure. Contrary to some previous findings, we detected no evidence that VPA acts as a widespread DNA demethylating agent in NPCs. iPSC-derived neural cells provide a simple and efficient experimental setup for the investigation of genetic and

epigenetic influences on gene expression regulation during neurodevelopment.



## PgmNr 3301: GWAS of epigenetic age acceleration in AUD reveals a role for *APOL2*.

### Authors:

J. Jung <sup>1</sup>; A. Luo <sup>1</sup>; M. Longley <sup>1</sup>; D. Rosoff <sup>1</sup>; K. Charlet <sup>1,2</sup>; C. Muench <sup>1</sup>; J. Lee <sup>1</sup>; Z.A. Kaminsky <sup>3,4</sup>; F.W. Lohoff <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Section on Clinical Genomics and Experimental Therapeutics, National Institute on Alcohol Abuse and Alcoholism, Bethesda, Maryland.; 2) Department of Psychiatry and Psychotherapy, Charite – Universitaetsmedizin Berlin, Berlin, Germany; 3) Department of Psychiatry and Behavior Sciences, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

---

Studies have shown that individuals with alcohol use disorder (AUD) have a significantly shorter lifespan compared to healthy individuals. Given the potential role of AUD in the aging process, it is important to understand the underlying biological mechanism. We employed Levine's epigenetic clock method to estimate DNA methylation age in 532 participants consisting of 331 AUD and 201 healthy controls recruited in the National Institute on Alcohol Abuse and Alcoholism (NIAAA) at the National Institutes of Health (NIH). We further evaluated effects of heavy, chronic alcohol consumption on the epigenetic age acceleration (EAA) using clinical biomarkers, including LFTs and clinical measures. To further characterize potential underlying genetic variation that contribute to age acceleration in AUD, we performed GWAS on EAA in our sample, including pathway analyses. Relevant top findings were follow-up with in silico eQTL analyses for biological function using the BRAINEAC database. Individuals with AUD showed a 2.22 years of age acceleration compared to controls after adjusting for gender and blood cell composition ( $p=1.85 \times 10^{-5}$ ). This effect became diluted but remained significant after additional adjustment for race, BMI, and smoking status (1.38 years,  $p=0.02$ ). Secondary analyses of AUD endophenotypes showed elevated heavy drinking days, GGT, and ALT were associated with even more severe EAA (all  $p<0.05$ ). The genome-wide meta-analysis of EAA in AUD revealed a genome-wide significant SNP in *APOL2* (rs916264,  $p=5.43 \times 10^{-8}$ ) that was also functional, with the minor allele A associated with increased EAA also affected higher mRNA expression on hippocampus brain tissue ( $p\text{-value} = 0.0015$ ). Taken together, our data show EAA in AUD, and suggest that more severe phenotypes of AUD show even greater EAA. Potential underlying mechanisms might include genetic variation in *APOL2*, a apolipoprotein of the high density lipoprotein family involved in cholesterol homeostasis, that affects accelerated biological aging in AUD.

## PgmNr 3302: Quantifying regulatory effects of schizophrenia-associated variation directly from patient DNAs.

### Authors:

G.D. Johnson <sup>1,2</sup>; A. Barrera <sup>1</sup>; Y. Kim <sup>1,2</sup>; W.H. Majoros <sup>2</sup>; P. Giusti-Rodríguez <sup>3</sup>; M.W. Halvorsen <sup>3</sup>; A.S. Allen <sup>2</sup>; P.F. Sullivan <sup>3,4,5</sup>; G.E. Crawford <sup>2,6</sup>; T.E. Reddy <sup>1,2,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Genomic and Computational Biology, Duke University, Durham, NC; 2) Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC; 3) Department of Genetics, Center for Psychiatric Genomics, University of North Carolina, Chapel Hill, NC; 4) Department of Psychiatry, University of North Carolina, Chapel Hill, NC; 5) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 6) Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC; 7) Computational Biology and Bioinformatics Program, Duke University, Durham, NC

---

Schizophrenia is a chronic, debilitating, and highly heritable mental illness that affects approximately 1% of the population. The relative absence of protein-coding variants linked with schizophrenia suggests that risk alleles associated with the disease may alter gene expression through perturbing gene regulation. To identify noncoding regulatory variants associated with schizophrenia, we used a population-scale high-throughput reporter assay to quantify the effect of tens of thousands of genetic variants obtained from patient genomes. Specifically, those assays measured the effects of regulatory variants across 104 schizophrenia-associated loci captured from 96 control and 96 schizophrenia patients of Swedish ancestry. Together, the captured regions spanned ~8 Mb of genomic DNA. The assays were performed in the SK-N-SH neuroblastoma cell model. In total, 87,437 variants were evaluated in the schizophrenia patients, including 31,772 rare variants (MAF<0.01). We then tested putative regulatory elements and variants in vivo in the fetal mouse brain, and in vitro via dCas9 epigenome editing in cell culture models. Using this GWAS screening methodology we can quantify the regulatory impact of thousands of disease-associated alleles and predict variants likely to influence gene expression in complex diseases such as schizophrenia.

## PgmNr 3303: Genetic variants affecting chromatin accessibility during human neuronal differentiation.

### Authors:

D. Liang<sup>1</sup>; A. Elwell<sup>1</sup>; O. Krupa<sup>1</sup>; K. Cheek<sup>1</sup>; K. Countney<sup>1</sup>; M. Love<sup>5</sup>; L. de la Torre-Ubieta<sup>4</sup>; D. Geschwind<sup>2,3,4</sup>; J. Stein<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of genetics and neuroscience center, University of North Carolina at Chapel Hill, Chapel Hill, NC.; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 3) Department of Psychiatry, Center for Neurobehavioral Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 4) Neurogenetics Program, Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 5) Department of genetics and Department of biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

---

Although genome-wide association studies (GWAS) have identified hundreds of common genetic variants associated with neuropsychiatric disorders, the molecular mechanisms underlying these risk variants are still poorly understood. GWAS risk loci are often in noncoding regions and likely play important roles in gene regulation. In order to determine how genetic variation impacts a measure of gene regulation (chromatin accessibility) during a critical time period of cortical neuronal differentiation, we performed a chromatin accessibility quantitative trait locus (caQTL) in cell-type specific population of human neural progenitor cells and their differentiated, labelled, and sorted neuronal progeny. We cultured neural progenitor cells (n=91 donors) derived from genotyped developing human cortical tissue donors (gestation weeks 14-21), and then differentiated and sorted virally labeled neurons (AAV2-hSyn1-EGFP) after 8 weeks of differentiation (n=66 donors). We then generated chromatin accessibility profiles for both neural progenitor cells and sorted neurons using ATAC-seq. We found global chromatin accessibility profiles are strongly different based on cell-type. We identified 136,714 peaks, of which 27,682 (FDR<0.05 and |log2FC|>1) peaks were differentially accessible between neurons and progenitors. As expected, we observed neurogenesis ontologies enriched in progenitor accessible peaks and also observed a strong enrichment in previously identified *in vivo* peaks. Common genetic variants within differentially accessible peaks in progenitors are significantly enriched in heritability explained for several human neuropsychiatric disorders and cognitive traits, most notably neuroticism, schizophrenia and educational attainment. QTL studies were then conducted on accessible peaks in progenitors and neurons separately. We found 22,892 caQTLs impacting 4,615 peaks in progenitors and 12,623 caQTLs impacting 2,715 in neurons. These caQTLs were highly cell-type specific with 11% of progenitor caQTLs (20% neuron caQTLs) having a significant effect in both cell types. Some caQTLs were useful in suggesting cell-type specific mechanisms of GWAS variants. For example, we identified a neuron-specific caQTL (rs9930307) within an intron of *GRIN2A* is also associated with the risk of schizophrenia. Using these caQTLs, we suggest molecular mechanisms of non-coding risk variants affecting gene regulations in particular cell types during human cortical neuronal differentiation.

## PgmNr 3304: DNA methylation profile of narcolepsy-affected brain and overlapping methylation profile between narcolepsy and multiple sclerosis.

### Authors:

M. Shimada <sup>1</sup>; T. Miyagawa <sup>1</sup>; K. Tokunaga <sup>2</sup>; M. Honda <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Genome Medical Science Project (Toyama), National Center for Global Health and Medicine (NCGM), Tokyo, Japan

---

Narcolepsy is a sleep disorder characterized by excessive daytime sleepiness and cataplexy. The loss of orexin- (hypocretin-) producing cells in the lateral hypothalamus (LH) causes the disease. Several genetic factors including *HLA-DQB1\*06:02* have been identified to date; however, the reason for orexin-producing cell loss and the pathogenesis are yet to be uncovered. Epigenetics has been suggested to play a certain role in the pathophysiology of complex diseases. Here we focused on DNA methylation and performed an epigenome-wide association study (EWAS) for narcolepsy and replication analyses with DNA samples extracted from two brain regions; LH (Case N=4, Control N=4) and temporal cortex (TC) (Case N=7, Control N=7). We first estimated proportions of surrogate measures of cell types present in brain using genome-wide methylation data. We found that there was no difference of cell populations between cases and controls both in LH and in TC, while there were significant differences between LH and TC. We then investigated differentially methylated regions (DMRs) of narcolepsy both in LH and in TC respectively. Seventy-seven DMRs were identified in the LH analysis, with the top association of a DMR in the *myelin basic protein (MBP)* region. In contrast, only five DMRs were detected in the TC analysis, suggesting that DNA methylation might be associated with narcolepsy especially in LH that is responsible area for the disease. Further, we developed two new approaches making the best use of the EWAS data; i) Investigation of methylation profiles shared between narcolepsy and other disorders and ii) An integrative analysis of the EWAS and the genome-wide association study (GWAS). The results of the two approaches, which included significant overlap of methylation profiles associated with narcolepsy and multiple sclerosis, indicated that the two diseases might partly share their pathogenesis. These major findings were replicated with independent LH-derived control samples (N=4). Although further replication studies are necessary, the DNA methylation in LH, that is responsible for narcolepsy, might have a certain role in the pathophysiology of the disease.

## PgmNr 3305: Recent stress and DNA methylation on emergent suicidal ideation: An epigenome-wide association study.

### Authors:

K.Z. Wang <sup>1,2</sup>; A. Bani-Fatemi <sup>1</sup>; V. De Luca <sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Centre for Addiction and Mental Health, Toronto, ON, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, ON, Canada

---

Suicide is a global public health problem, accounting for over 788,000 lives lost annually. In particular, patients with schizophrenia have a tenfold increase in suicide risk, which contributes greatly to their reduced life expectancy. The presence of suicidal thoughts, or suicidal ideation (SI), provides a robust predictor of future suicide. Thus, schizophrenia patients with newly emergent SI are especially at risk. In order to predict and prevent suicide, it is imperative to identify reliable biomarkers. Studies have shown that suicide is often triggered by recent stressful events. DNA methylation, an epigenetic signaling mechanism, plays a role in mediating gene expression and is sensitive to environmental factors such as stress. The present 3-month longitudinal study examines the impact of recent stress on emergent SI, and investigates the epigenome-wide methylation changes in relation to emergent SI in patients with schizophrenia. In a sample of 4 subjects with emergent SI and 4 subjects without SI, we found that increased recent stress was associated with emergent SI ( $p = 0.047$ ), specifically in the domains of family and interpersonal relationships ( $p = 0.027$ ) and personal health ( $p = 0.00002$ ). Furthermore, we quantified the changes in DNA methylation at 850,000 CpG sites using the Illumina Infinium MethylationEPIC BeadChip, and determined that the CpG site cg06371916 in chromosome 10 was the most differentially methylated site, characterized by hypomethylation of white blood cells of subjects with emergent SI compared to those without SI. Combined, these findings suggest that increased recent stressors is predictive of emergent SI, and that there are variations in DNA methylation at certain CpG sites associated with emergent SI. Further work can be done to investigate the mediating effect of methylation on stress causing emergent SI, in order to better establish molecular biomarkers for suicide prevention.

## **PgmNr 3306: A dynamic transcriptional program of transposable elements mediate cell fate specification and lineage commitment during corticogenesis.**

### **Authors:**

G. Guffanti Masetti <sup>1</sup>; F. Macciardi <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Psychiatry, McLean Hospital/Harvard Medical School, Belmont, Massachusetts.; 2) Department of Psychiatry, University of California Irvine, Irvine, California

---

Human brain development is an extensive process that begins during early embryonic state with the differentiation of distinct cell types from neural stem cells (NSC) in a strictly regulated temporal order. The molecular mechanisms that give rise to NSC fate specification are still mostly unexplored, but noncoding regulatory elements are consistently implicated. We developed and validated a straightforward approach to identify unambiguous genome-wide expression profiles of discrete transposable elements (TEs) using a *de novo* assembly strategy. We set out to investigate the role of TE in the embryonic development of the human cortex at single-cell resolution in an in vitro model of hESC differentiating into cortical neurons over 77 days using RNA sequencing. We found that TEs 1) can be profiled in in vitro models of hESC; 2) play a role in the cell fate decision along the differentiation of human cortical development; 3) regulate target gene expression during neurodevelopment.

## PgmNr 3307: Dysregulation of miR-1271-5p in schizophrenia.

### Authors:

M. Geaghan<sup>1,2</sup>; D. Kiltschewskij<sup>1,2</sup>; J. Atkins<sup>1,2</sup>; M. Cairns<sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) School of Biomedical Sciences and Pharmacy, University of Newcastle, Australia, Callaghan, NSW, Australia; 2) Centre for Brain and Mental Health Research, Hunter Medical Research Institute, New Lambton, NSW, Australia

---

### Background:

While the cause of schizophrenia (SZ) is still unknown, multiple genes have been associated with the condition, at both the genetic and gene expression levels. Particular interest has been given to small non-coding RNAs called microRNA (miRNA). Many miRNA have been found differentially expressed in both peripheral and brain tissues, and several miRNA have been consistently linked to SZ. This study aimed to identify putative miRNA-mRNA interactions in peripheral blood mononuclear cell (PBMC) tissue, and investigate the roles of SZ-associated miRNAs in regulating disease-relevant genes.

### Methods:

We obtained PBMC tissue from individuals with and without SZ. Total RNA was extracted and both miRNA and mRNA sequencing was performed. One downregulated miRNA, miR-1271-5p, was further investigated through overexpression and knockdown in an in vitro neuronal cell culture model using synthetic RNA mimics and inhibitors. Both mRNA sequencing and ribosome profiling were performed to analyse differential transcription and translation.

### Results:

Several miRNAs and genes were differentially expressed in PBMCs in SZ, with an enrichment of upregulated immune-related genes. Several miRNA were negatively correlated with their predicted target genes, including several immune-relevant genes. One miRNA of interest, miR-1271-5p, was downregulated in both males and females, replicating previous findings. In vitro perturbation of miR-1271-5p identified roles in regulating a diverse range of gene networks, including cell membrane and cytoskeletal maintenance, neurodevelopment and neuronal function, as well as several SZ-associated genes including *GRM7*, *SLC1A1*, *HOMER1*, and *RELN*.

### Discussion:

These results add to growing evidence that disruption of miRNA regulation is associated with SZ. Peripherally, miRNA dysregulation may be partially responsible for an increased expression of immune system-relevant genes. We replicated previous findings of miR-1271-5p downregulation in PBMCs in SZ. Interestingly, this miRNA is also known to be a brain-enriched homolog of miR-96-5p, miR-182-5p, and miR-183-5p, all of which are contained within a single miRNA cluster and have known roles in regulating neuronal function. Altered expression of miR-1271-5p in vitro affected gene networks relevant to neuronal function and SZ. Thus, miR-1271-5p may be important to the pathophysiology of SZ through disruption of its target genes both peripherally and in the brain.

## PgmNr 3308: Explore the value of circRNA-miRNA multimolecular regulatory network in the diagnosis of early onset schizophrenia.

### Authors:

S. Liu <sup>1,3</sup>; F. Zhang <sup>2</sup>; Z. Li <sup>1</sup>; Y. Xu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) First Hospital/First Clinical Medical College of Shanxi Medical University, Taiyuan, China.; 2) Wuxi Mental Health Center of Nanjing Medical University, Wuxi, China.; 3) Broad Institute of MIT and Harvard, Cambridge, MA.

---

Schizophrenia (SZ) is a severe chronic mental health disorder with 1% lifetime prevalence and unclear pathological mechanism. At present, the diagnosis and prognosis evaluation of SZ still mainly depends on symptomatic observations, which is likely to cause misdiagnosis and missed diagnosis. Early-onset schizophrenia (EOS) was defined as SZ with onset before 18 years of age and experienced even more severe psychiatric symptoms and worse prognosis. Therefore, exploring objective and readily available biomarkers for EOS has become an urgent matter to be solved.

Here we selected 10 EOS patients and 10 matched controls to identify the differentially expressed (DE) circRNAs and miRNAs in plasma by Illumina HiSeq high throughput sequencing. Based on the interaction between circRNA and miRNA, bioinformatics methods were also used to integrate multilevel indicators of gene regulatory expression and construct circRNA-miRNA centered gene regulatory network for EOS. Further, verification of the specificity and sensitivity of these multilevel indicators will be done in a larger sample.

10 peripheral plasma samples from patients with EOS and 10 matched normal controls were selected to quantify the expression of miRNA and circRNA. 44 miRNAs and 50 circRNAs were showed significant DE level in case ( $P \leq 0.05$  and  $FC \geq 2$ ). Among them, 7 miRNAs and 1 circRNA showed a down-regulation trend in the case group, and 37 miRNAs and 49 circRNAs were up-regulated in the patient group. By using silicon analysis, target mRNAs and DE circRNA-miRNA interaction network was further predicted. Enrichment of the DE circRNAs and miRNAs in KEGG and GO (biological processes and molecular function) pathway were also analyzed. We discovered a significant enrichment of these genes associated with nervous system development ( $P = 4.706 \times 10^{-7}$ ,  $FDR = 0.001$ ), even with neuron projection, Glutamatergic synapse, MAPK signaling pathway and etc.

Our study clearly suggests that multilevel-based molecular network in plasma could be served as a more reliable biomarker for EOS. Although further verification should be done, our exploration is still helpful to complement current diagnostic strategies, even shed light on the pathology of SZ.



## PgmNr 3309: Characterizing differential *cis*-regulatory domains in neuronal epigenome across schizophrenia cases and controls.

### Authors:

K. Girdhar; G. Hoffman; J. Bendl; Y. Jiang; L. Brown; O. Devillers; Y. Wang; H. Shah; E. Zharovsky; R. Jacobov; J. Wiseman; E. Flatow; R. Park; J. Johnson; B. Kassim; P. Sklar; P. Roussos; S. Akbarian

[View Session](#) [Add to Schedule](#)

**Affiliation:** Mount Sinai, New York, New York.

---

Genome wide association studies (GWAS) of schizophrenia (SCZ) have revealed that disease-associated variants are mapped to gene regulatory sequences. Moreover, pathway analyses of GWAS-SCZ risk loci have shown the strongest association with histone modifications. Yet, we understand very little about the role of regulatory elements underneath these disease specific signals. For example, how do regulatory elements (REs) coordinate to modulate genes' expressions in schizophrenic patients' brains vs. healthy brains? To answer this question, we profiled histone modifications: histone H3-trimethyl-lysine 4 (H3K4me3) and histone H3-acetyl-lysine 27 (H3K27ac) in fluorescence-activated cell sorted (FACS) nuclei from the prefrontal cortex (PFC) brain regions of 149 SCZ postmortem brains compared to the 145 healthy brains. We estimated the correlation of chromatin peaks across all the individuals' brains to probe the coordination of regulatory elements. We found the neuronal genome was partitioned into ~9,000 neuronal *cis*-regulatory domains (nCRDs) that encompassed distinct patterns of histone marks. A total of 75 nCRDs were identified in SCZ brains that showed differential neuronal dysregulation than the healthy brains. Pathway analyses of differential nCRDs showed association with biological processes such as calcium ions signaling. We then characterized the association of genes with nCRDs, wherein ~50% of the genes were associated with nCRDs within the window of  $\pm 1$  Mb. Annotation of nCRDs, showed clusters of active REs.

Our research work provides unique insight into the role of regulatory elements in modulating the gene expression that might be causal for neuronal dysfunction in schizophrenia. Our PsychENCODE sponsored research provides a resource of genome-wide map of neuronal *cis*-regulatory domains (nCRDs) and schizophrenia specific nCRDs. The disease specific nCRDs can be further explored using chromosome conformation capture technique to study psychosis in animal models to advance our knowledge in schizophrenia. In addition to this, we provide a computational pipeline that translates linear functional genomics assays such as ChIP-seq into three dimensional higher chromatin structures. Supported by NIMH U01MH103392

## **PgmNr 3310: Generating a resource for integrating transcription factor binding profiles with functional genomic data in multiple human brain regions.**

### **Authors:**

J.M. Loupe <sup>1</sup>; W.E. Bunney <sup>2</sup>; P. Cartegena <sup>2</sup>; S.J. Watson <sup>3</sup>; G. Cooper <sup>1</sup>; R.M. Myers <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) Department of Psychiatry and Human Behavior, University of California, Irvine, CA; 3) The Molecular & Behavioral Sciences Institute, University of Michigan, Ann Arbor, MI

---

Gene regulation is crucial to the physiology of human cells and disease risk, but our knowledge of transcriptional regulatory elements in the brain and their contribution to psychiatric disease is limited. As part of the HudsonAlpha Brain TF Project, we are developing a deep resource of regulatory information including thousands of maps of transcription factor (TF) binding using ChIP-seq in 9 brain regions from 4 individuals and neuronal progenitor cells from 4 individuals. In addition to ChIP-seq, we are gathering other valuable data pertaining to gene expression, including RNA-seq, microRNA-seq, Whole Genome Sequencing (WGS), and DNA methylation profiling. This work has yielded rich maps of regulatory elements to facilitate analyses of the molecular effects of genetic variation in the brain and discovery of pathways and regulatory mechanisms relevant to psychiatric disease. Beyond our own analyses, these data will be a resource for other studies.

Here, we show recent progress in optimizing a ChIP-seq pipeline that requires a minimal input of brain tissue to collect data of multiple histone-markers and TFs in post-mortem human brain tissues. We have thus far implemented this pipeline to produce high quality TF binding data across nine brain regions from two individuals using small amounts of tissue, thus maximizing the potential for future experiments on precious samples. In addition, we have generated distinct TF binding maps for neuronal and non-neuronal cell populations isolated from cortex tissue by sorting nuclei based on NeuN signal with flow cytometry. The TF signals produced from sorted nuclei are consistent with expected regulation of genes specific to each cell-type and can distinguish neuronal from glial populations by assessing gene ontology of cis-regulatory regions. Furthermore, this approach distinguishes exclusive epigenetic signatures that are unique to neuronal and non-neuronal cell populations. This technique, when implemented in the resource-building component for TF maps across tissues, will provide an invaluable resource for gene regulation in the human brain.

## PgmNr 3311: Investigating the role of the epigenetic reader *Brd4* in hearing and development of the inner ear.

### Authors:

A. Kannan-Sundhari<sup>1,2,4</sup>; C. Abad<sup>3</sup>; M. Maloof<sup>1,5</sup>; N. Ayad<sup>5</sup>; J. Young<sup>3,4</sup>; S. Blanton<sup>2,3,4</sup>; X.Z. Liu<sup>2,4</sup>; K. Walz<sup>3,4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) University of Miami, Miami, FLORIDA.; 2) Department of Otolaryngology, University of Miami, Miami, FL, USA; 3) John P. Hussman Institute of Human Genomics, University of Miami, FL, USA; 4) The Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, FL, USA; 5) Department of Psychiatry and Behavioral Sciences, University of Miami, Miami, FL, USA

---

Hearing loss (HL) is the most common sensorineural disorder in humans, with profound congenital HL occurring in approximately 1/500 births. In mammals, sensorineural HL (SNHL) arises from irreversible damage to the sensory hair cells (HC) in the cochlea. Investigating transcriptional regulation in inner ear would help understand the mechanisms involved in sensory HC development, function and maintenance. Recently, genes involved in DNA methylation, histone modification and chromatin remodeling have been associated with hereditary HL. Syndromic forms of deafness, such as autosomal dominant cerebellar ataxia, deafness, and narcolepsy (ADCA/DN) and hereditary sensory neuropathy type IE (HSN1E) have been shown to be caused by mutations in *DNMT1*, important in methylation. In this study, we investigated the role of *Brd4*, a chromatin reader protein that recognizes and binds acetylated histones which are relevant in transcription regulation. To generate the conditional knock-out model we crossed *Brd4*<sup>tm1a(EUCOMM)Wtsi</sup> heterozygous mice with B6.Cg-Tg (ACTFLPe)9205Dym/J to obtain *Brd4*<sup>fl/fl</sup> mice. These were then crossed to B6.Cg-Tg (Atoh1-cre)1Bfri/J mice to obtain conditional expression of Cre in the inner ear to eventually generate *Brd4*<sup>fl/fl</sup>; *Atoh1-cre*<sup>+/-</sup>, *Brd4*<sup>fl/fl</sup>; *Atoh1-cre*<sup>-/-</sup> and *Brd4*<sup>wt/wt</sup> littermates. Auditory brainstem response (ABRs) were recorded and compared to *Brd4*<sup>fl/fl</sup>; *Atoh1-cre*<sup>-/-</sup> and *Brd4*<sup>wt/wt</sup> mice using a Smart EP Universal Smart Box (Intelligent Hearing Systems, Miami, FL) at P21. ABR stimuli were induced as 0.1 ms broadband clicks and 0.1 ms pure-tone pips at frequencies 8, 16 and 24 kHz through 20-100dB intensities. To test the mechanical activity of outer hair cells (OHCs) in the cochlea, distortion product acoustic emission (DPOAEs) were recorded in the mutant mice and compared to controls. The ABR thresholds for click response in *Brd4*<sup>wt/wt</sup> mice were ~ 65dB SPL, ~67 dB SPL in *Brd4*<sup>fl/fl</sup>; *Atoh1-cre*<sup>-/-</sup>, and ~100 dB SPL in *Brd4*<sup>fl/fl</sup>; *Atoh1-cre*<sup>-/-</sup>. As with the pure-tone frequencies, mutant mice had significantly elevated ABR thresholds at P21 (~100dB SPL) compared with control mice. The DPOAE recordings at P21 produced by OHCs in the mutants had reduced amplitudes in 8-16 KHz compared to both control groups but had similar profiles at 24 KHz. Thus, the mutant mice were profoundly deaf, indicating a relevant role for *Brd4* in the postnatal development of HCs. This study has been supported by R01 DC005575 and R01 DC012115 from NIH/NIDCD to Xue Zhong Liu.

## PgmNr 3312: DNA hypomethylation in high myopia: Differentially methylated genes in overrepresented molecular pathways could contribute to high myopia in Polish children.

### Authors:

M. Gajecka<sup>1,2</sup>; S. Vishweswaraiah<sup>3</sup>; J.A. Karolak<sup>2,1</sup>; M. Mrugacz<sup>4</sup>; U. Ratnamala<sup>5</sup>; N.K. Mishra<sup>6</sup>; C. Guda<sup>6</sup>; S.S. Chettiar<sup>7</sup>; K.R. Johar<sup>7</sup>; R. Uppala<sup>3</sup>; J. Swierkowska<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland; 2) Department of Genetics and Pharmaceutical Microbiology, Poznan University of Medical Sciences, Poznan, Poland; 3) Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA; 4) Department of Ophthalmology and Eye Rehabilitation, Medical University of Bialystok, Bialystok, Poland; 5) Department of Pharmacology, Creighton University, Omaha, NE, USA; 6) Department of Genetics, Cell Biology & Anatomy College of Medicine, University of Nebraska Medical Center Omaha, NE, USA; 7) Department of Zoology, School of Sciences, Gujarat University, Ahmedabad, India

**Introduction:** High myopia (HM) is an eye disorder characterized by refractive error (RE) greater than -6.0 diopters (D) and axial length (AL) more than 26.0 mm. The etiology of HM is complex and involves genetic and environmental factors. Recently, we have completed the global analysis of DNA hypermethylation in Polish HM children (PMID:30858441). Here, we focus on the aspect of hypomethylation to complement the methylation studies in HM.

**Materials and Methods:** The DNA hypomethylation was assessed based on the results of genome-wide DNA methylation analysis performed on DNA of 18 Polish HM children (aged 4-12 years, RE -6.0 – -15.0 D, AL 26.22 mm – 27.85 mm) and 18 age-matched Polish control individuals. In total, 358 hypomethylated genes, containing 333 CpG dinucleotides with more than 12% difference of methylation level between HM children and controls and FDR p-value < 0.00001, were chosen for overrepresentation analyses of molecular pathways in ConsensusPathDB-human server. Only pathways with p-value ≤ 0.01 and sharing at least five genes with our uploaded gene set were considered.

**Results:** The most hypomethylated CpGs with more than 25% difference in the methylation level between children with HM and controls was located in long intergenic non-coding RNA 1088 (*LINC01088*), erythrocyte membrane protein band 4.1 like 4A (*EPB41L4A* [MIM 612141]), ADAM metallopeptidase domain 20 (*ADAM20* [MIM 603712]), grainyhead like transcription factor 1 (*GRHL1* [MIM 609786]), and WW domain containing oxidoreductase (*WWOX* [MIM 605131]). Enrichment analysis of hypomethylated genes set revealed overrepresentation of genes with even more than 10 shared genes in several pathways, including steroid hormone biosynthesis, phase II - conjugation of compounds, metabolism biotransformation phase I and II, signaling by receptor tyrosine kinases, biological oxidations, intracellular signaling by second messengers, focal adhesion-PI3K-Akt-mTOR-signaling pathway, axon guidance, and miR-targeted genes in epithelium. Signaling by receptor tyrosine kinases, focal adhesion-PI3K-Akt-mTOR-signaling, and axon guidance pathways were also identified in our previous hypermethylation studies.

Conclusions: The observed differences in methylation level of genes of the overrepresented pathways indicate the involvement of methylation aspects in HM etiology.

## PgmNr 3313: Early and progressive hearing loss in a mouse model of Kabuki syndrome.

### Authors:

T.R. Luperchio<sup>1</sup>; K. Schrode<sup>2</sup>; V. Deleon<sup>3</sup>; J. Fahrner<sup>1,4</sup>; A. Lauer<sup>2</sup>; H.T. Bjornsson<sup>1,4,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 2) Department of Otolaryngology-Head and Neck Surgery and Center for Hearing and Balance, Johns Hopkins School of Medicine, Baltimore, Maryland, USA; 3) Department of Anthropology, University of Florida, Gainesville, Florida, USA; 4) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 5) Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik 101, Iceland; 6) Landspítali University Hospital, Reykjavik, 101, Iceland

**Introduction:** Kabuki syndrome type 1 (KS1) is a Mendelian disorder of the epigenetic machinery caused by mutations in *KMT2D*, a gene encoding a histone-modifying enzyme. *KMT2D* is responsible for histone H3K4 methylation especially at enhancers and promoters of genes. We have characterized a mouse model of KS1 (*Kmt2d*<sup>+/ $\beta$ Geo</sup> mice) that demonstrates many KS1 features including growth retardation, craniofacial defects, and immune dysregulation. *Kmt2d*<sup>+/ $\beta$ Geo</sup> mice also demonstrate a disruption of adult neurogenesis in the dentate gyrus and hippocampal memory defects, phenotypes rescued postnatally with agents targeting the epigenetic machinery.

**Objective:** Many of our KS1 patients develop progressive hearing loss requiring intervention with aids or cochlear implants during early childhood. Although the etiology of this hearing phenotype is poorly understood, current clinical dogma states it is caused by repeated ear infections in childhood or anatomic abnormalities. However, in our clinic we have observed individuals with hearing loss in early life without anatomic abnormalities or any history of ear infection. We hypothesize that hearing loss in KS1 is sensorineural due to the loss of *KMT2D* and not recurrent ear infections.

**Methods:** To explore the basis of this hearing loss we phenotyped the *Kmt2d*<sup>+/ $\beta$ Geo</sup> mice by imaging and auditory brainstem responses. We further focused on hair cells by directly testing their function through distortion product otoacoustic emissions (DPOAE) and histology of the sensory tissue.

**Results:** We observe no anatomic abnormalities but we do observe auditory response deficits at onset of hearing in our *Kmt2d*<sup>+/ $\beta$ Geo</sup> mice compared to wild type litter mates. This phenotype becomes progressively worse over time. Interestingly, we observe abnormalities of hair cell function by DPOAE and found outer hair cell death at later ages but not at early postnatal time points.

**Conclusion:** Since we find impaired hearing in early life of *Kmt2d*<sup>+/ $\beta$ Geo</sup> mice, we conclude that loss of *KMT2D* results in sensorineural hearing loss independent of recurrent ear infections common in Kabuki syndrome. These data suggest that conditional, transcriptional, and epigenetic studies prior to hair cell demise could help us understand how the deficiency of a chromatin modifier leads to hearing loss through disruption of epigenetic and transcriptional programs. Furthermore, this offers another quantitative postnatal outcome measure for therapeutic development for Kabuki syndrome.

## PgmNr 3314: DNA methylation of microRNA-coding genes in non-syndromic high myopia.

### Authors:

N. Saiyed <sup>1,3</sup>; K. Johar <sup>2,3</sup>; S.K. Nath <sup>4</sup>; S. Chettiar <sup>3</sup>; U. Ratnamala <sup>5</sup>; S. Vishweswaraiah <sup>6</sup>; U. Radhakrishna <sup>6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Green Cross Pathology & Molecular Laboratory, India.; 2) Raghudeep Eye Hospital, India; 3) Department of Zoology, Gujarat University, India; 4) Department of Pharmacology, Creighton University, Omaha NE; 5) Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, OK, United States; 6) Department of obstetrics and gynecology, Oakland University William Beaumont school of medicine, Royal Oak, Mi, USA.

---

Myopia, commonly referred to as nearsightedness, is one of the most common causes of visual disability throughout the world. It affects more people worldwide than any other chronic visual impairment condition. Ocular tissue gene expression variation, including the dysregulation of MicroRNA (miRNA) genes, may lead to the development of myopia. Blood-based miRNA profiling is being used as a diagnostic tool for multiple diseases. Until, there are no studies thus far to address epigenetic methylation of miRNA genes that may affect the progression of myopia. To elucidate the epigenetic mechanism(s) underlying the pathophysiology of high-myopia, we investigated DNA methylation levels of miRNA encoding genes using the Illumina MethylationEPIC BeadChip array in 50 axial HM subjects and an equal normal of age, gender, sex and ethnic matched controls (aged 4-12 years). The degree of myopia was variable among subjects, ranging from -6 to -15D. Several “R” packages were used to identify the significant CpG targets of miRNA. We have performed bioinformatic analysis by mining the MicroRNA Target Prediction and Functional Study Database (miRDB) to predict the possible genes regulated by the identified methylated miRNAs. We have identified seven significantly differentially methylated miRNAs: miR-6737, miR-636, miR-548H4, miR-548G, miR-423, miR-7704 and miR-2277. All identified miRNAs were found to be methylated at their transcription start site or at the 5’UTR region. One of the differentially methylated miRNAs, miR-548G, targets the paired box 6 (PAX6) gene with high target score. This gene has been identified as a candidate for developing HM. This is the first study of the methylation difference of global miRNA genes in HM cases. The study presents a novel approach to understand the molecular mechanism of HM. Altered methylation of miRNAs detectable in blood may potentially be a sensitive biomarker to detect early risk of HM. Understanding the complex mechanism leads to the potential of using miRNAs as prognostic tools and therapeutic agents for HM.

## PgmNr 3315: Characterization of *GJB2* *cis*-regulatory elements in the *DFNB1* locus.

### Authors:

A. Le Nabec <sup>1</sup>; A. Quillevere <sup>1</sup>; C. Le Marechal <sup>1,2</sup>; C. Ferec <sup>1,2</sup>; S. Moisan <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Univ Brest, Inserm, EFS, UMR 1078, GGB, F-29200 Brest, France; 2) Laboratoire de Génétique Moléculaire et d'Histocompatibilité, CHRU de Brest, Bretagne, Brest, France

---

Three-dimensional chromatin organization plays a key role on gene expression. Gene regulation depends on *cis*-regulatory elements which can interact with gene promoter by chromatin loop. Alteration of chromatin architecture and/or *cis*-acting elements can lead to *cis*-ruption disorder. Numerous unelucidated nonsyndromic hearing loss and deafness 1 (*DFNB1*) cases carrying out only one heterozygous pathogenic mutation on Gap Junction Beta 2 (*GJB2*) gene, led to strongly suggest the presence of distant *cis*-regulation.

To analyze chromatin conformation of a large *DFNB1* locus, we performed the 5C technology and showed several chromatin contacts with *GJB2* promoter. Then, to identify potential regulatory elements of the *GJB2* gene, we analysed these chromatin interacting regions with the *GJB2* promoter by reporter test activity and identifying *cis*-regulatory elements which have enhancer action and silencer effect on *GJB2* expression. Moreover, to explore *DFNB1* locus architecture we analysed CTCF (CCCTC-binding factor) binding along *DFNB1* locus by ChIP-qPCR and defined an active domain in which *cis*-acting elements were closer brought to the *GJB2* promoter through chromatin looping. This first study of *DFNB1* three-dimensional chromatin organization allows identification of *GJB2* *cis*-acting elements. Finally, better understand of the molecular mechanisms of *DFNB1* deafness could be important to improve genetic diagnosis of hearing loss, to adapt patients care and to develop small therapeutic molecules.



## PgmNr 3316: A circulating microRNA profile in a laser-induced mouse model of neovascular age-related macular degeneration (AMD).

### Authors:

C. Kiel <sup>1</sup>; P. Berber <sup>1</sup>; M. Karlstetter <sup>2</sup>; A. Aslanidis <sup>2</sup>; T. Langmann <sup>2</sup>; F. Grassmann <sup>1,3</sup>; B.H.F. Weber <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Human Genetics, University of Regensburg, Regensburg, Germany; 2) Laboratory for Experimental Immunology of the Eye, Department of Ophthalmology, University of Cologne, Cologne, Germany; 3) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

---

**Purpose:** Age-related macular degeneration (AMD) is the major cause of severe visual impairment in people over the age of 55 in Western societies. Given that AMD is a complex disease, investigations into its pathomechanism are challenging. Circulating microRNAs (cmiRNAs) have been shown to represent valuable biomarkers for several age-related diseases and can be readily measured in blood, since they are common cargo of exosomes or microvesicles. Dysregulation of cmiRNA expression at defined stages of AMD may point towards molecular pathways involved in its pathobiology and may therefore give rise to novel treatment options. In this study, we aimed to investigate the role of dysregulated cmiRNAs in a laser-induced mouse model of neovascular AMD.

**Methods:** Laser coagulation was used to induce six lesions per eye in six C57BL/6J mice. Blood samples were taken the day before (day 0) and on three defined time-points after laser treatment (days 3, 7 and 14). CmiRNA profiles were determined by Next Generation Sequencing. Results were validated independently by quantitative real-time RT-PCR (qRT-PCR) with samples from an additional mouse cohort with six laser-treated and six control animals. Moreover, we used qRT-PCR to investigate the local expression of miRNAs in the retina and retinal pigment epithelium (RPE)/choroid complex at the same time points, in 12 lasered and 12 control mice, respectively.

**Results:** The cmiRNA levels were analyzed at the four defined time-points in a two-step approach, including a discovery and a validation phase. Accordingly, we identified 10 cmiRNAs to be significantly dysregulated after laser-treatment compared to baseline at least at one time-point ( $p < 0.05$ ). Expression of these 10 candidate cmiRNAs were also analyzed in the retina and the RPE/choroid tissues. In each tissue, seven miRNAs were also found to be dysregulated in laser-treated mice compared to control mice, including five miRNAs being dysregulated in both tissues.

**Conclusion:** We detected 10 cmiRNAs to be differentially expressed in a laser-induced mouse model of neovascular AMD. Nine of these also displayed dysregulation in the retina or RPE/choroid complex. Our data highlight the potential role of cmiRNAs in the pathology of neovascular AMD, and point to the use of cmiRNAs as biomarkers in this disease. This is underpinned by the fact that two cmiRNAs, miR-20a and miR-140, were previously identified in a cmiRNA biomarker study in patients with neovascular AMD.

## PgmNr 3317: Biochemical quantification of pseudouridine in spliceosomal RNAs in relation to snoRNAs expression levels: A panacea to developmental defects?

### Authors:

A. Bello <sup>1</sup>; A. Marshall <sup>2</sup>; J.E. O'Brien <sup>2</sup>; N. Kibiyeva <sup>2</sup>; M. Filla <sup>2</sup>; D.C. Bittel <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biological science, Kansas City University of medicine and bioscience, Kansas City, Missouri.; 2) Children's Mercy Hospital, Kansas City Missouri

---

Messenger RNA splicing is the process of removing introns from pre-messenger RNA and ligating together exons to produce a mature messenger RNA (mRNA) that represents the template for protein translation. Alternative pre-mRNA splicing adds diversity to the proteome and greatly increases its diversity. scaRNAs catalyze 2'-O-methylation or pseudouridylation of specific nucleotides in spliceosomal RNAs. We showed that a reduction in the expression level of scaRNA is associated with alternative mRNA splicing and congenital heart defects. Furthermore, our studies and those of others have demonstrated that scaRNA levels are developmentally dynamic and must be precisely regulated in order to achieve normal embryological development. Messenger RNA splicing is catalyzed by the spliceosome, a complex macromolecular machine that consists of a dynamic set of hundreds of proteins and five small RNAs. Although we assume that methylation or pseudouridylation were the intermediary steps in adjusting alternative splicing we have not yet examined the levels of methylation or pseudouridylation in direct response to changing scaRNA levels. Therefore, we developed a process to quantify pseudouridylation based on modifications of the technique developed by Carlile et al (2015). Here we present an assessment of U/ψ89 in spliceosomal RNA after experimental manipulation of scaRNA1, which directs the pseudouridylation of U89. Many genetic diseases in humans are a consequence of splicing defects often due to mutation in splicing machinery, so developing tools to target the spliceosome to correct splicing represents a novel approach to therapy. However, a complete understanding of the mechanisms that influence spliceosome function must be clearly understood in order to successfully target the spliceosome. ScaRNA expression levels that attenuate biochemical modification of spliceosomal RNAs that consequently fine-tune alternative splicing represents an unappreciated functionally important mechanism that might be amenable to therapy.

## PgmNr 3318: Examining *SRCAP* mutations in Floating-Harbor syndrome: An induced pluripotent stem cell disease modeling approach.

### Authors:

R.L. Hood <sup>1,2</sup>; K.M. Boycott <sup>2,3,4</sup>; M.J.M Nowaczyk <sup>5</sup>; W.L. Stanford <sup>1,6,7</sup>; D.E. Bulman <sup>1,2,4,7</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Biochemistry, CHEO Research Institute, University of Ottawa, Ottawa, Ontario, Canada; 2) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa; 3) Department of Genetics, Children's Hospital of Eastern Ontario; 4) Department of Pediatrics, University of Ottawa; 5) Department of Molecular Medicine and Pathology, McMaster University, Hamilton; 6) The Sprott Centre for Stem Cell Research, Regenerative Medicine Program, Ottawa Hospital Research Institute; 7) Department of Cellular and Molecular Medicine, University of Ottawa

---

Floating-Harbor Syndrome (FHS; OMIM# 136140) is characterized by short stature, delayed osseous maturation, cognitive deficit, and distinct facial dysmorphology. We previously identified heterozygous truncating mutations in the 3' end of *SRCAP* as the genetic cause underlying FHS. We also demonstrated that FHS-causing mutations result in unique and highly reproducible methylation alterations in the peripheral blood-derived DNA of FHS patients. *SRCAP* has been described as having coactivator roles in CREB and CBP-mediated and nuclear (steroid) hormone receptor signaling pathways. It has also been shown to co-precipitate as part of a large chromatin remodeling complex which catalyzes ATP-dependent displacement of the histone variant H2A by H2A.Z. Based on the described roles of *SRCAP*, the clustered nature of our mutations, and our finding that FHS mutations result in genomic methylation alterations, we expect that FHS-causing mutations result in widespread gene dysregulation. To examine the functional consequences of these alterations in cell lineages implicated in the FHS phenotype, we employed an induced pluripotent stem cell (iPSC) disease modeling approach. Patient and gender-matched fibroblast cell lines were reprogrammed into iPSCs using an episomal vector system. Clones were characterized by pluripotency cell surface marker expression, differentiation capability (*in vitro* and *in vivo*), and karyotyping assays. iPSCs were differentiated into neural crest cells (NCCs) using a dual SMAD inhibition approach. Expression of NCC markers p75, HNK-1, and SOX10 demonstrated normal NCC specification for all lines. For downstream analysis, pure populations of NCCs were generated by FACs sorting (p75+/PI-). RNA and transposed-DNA were isolated from the initial fibroblast cell lines, reprogrammed iPSC lines, and iPSC-derived FACs sorted NCCs. From these, a combination RNAseq/ATAC-seq strategy is being used to examine alterations in downstream expression and chromatin architecture. By comparing data between these two technologies, and across three cell lineages, we expect to identify key targets and downstream pathways altered in FHS, which will allow us to better understand the FHS disease mechanism and the dysmorphology seen in FHS in a precise tissue specific manner.

## PgmNr 3319: Chromatin structure changes are more closely associated with adipogenesis than osteogenesis.

**Authors:**

R. Hao; J. Guo; S. Dong; T. Yang; Y. Guo

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, Shaanxi, China

---

Chromatin conformation change has been proved to have close connection with gene regulation mechanism changing in cells, especially during cell differentiation procedures. Mesenchymal stem cells (MSCs) are multipotent and can be differentiated into adipocytes and osteoblasts. While the transcriptomic and epigenomic changes in response to adipogenesis and osteogenesis have been characterized, what happened to the 3D chromatin conformation stabilization is hardly known. Here we induced human MSCs to adipogenic and osteogenic differentiation in vitro. We performed 2k resolution Hi-C experiments in 3 cells and detected cell specific “loop” structures. After integration with RNA-seq, H3K4me1, H3K27ac ChIP-seq and ATAC-seq data, we demonstrate that loop structures are associated with active gene expression in general. However, loops in adipocytes show evidence of harboring adipogenesis related genes with active enhancers, whereas osteogenesis associated genes are more likely to locate outside of loops in osteoblasts. We also marked distal and proximal cell specific enhancers in 3 cells, and by mapping open chromatin landscapes, we proposed different gene regulation patterns between adipogenesis and osteogenesis. At the same time, we also identified associated transcription factors that might play a crucial role in lineage determination. Overall, our study provides a comprehensive view of chromatin structure changes and its associated regulation landscape differences during adipogenesis and osteogenesis.

# PgmNr 3320: Impact of structural variations on human genome topology at the population scale.

## Authors:

D. Plewczynski <sup>1,3</sup>; M. Sadowski <sup>1</sup>; A. Kraft <sup>1,3</sup>; P. Szalaj <sup>1</sup>; M. Wlasnowolski <sup>1,3</sup>; Z. Tang <sup>2</sup>; Y. Ruan <sup>2</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Centre of New Technologies, University of Warsaw, Warsaw, mazovia, mazovia, Poland; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; 3) Faculty of Mathematics and Information Science, Warsaw University of Technology, Warsaw, Poland

---

## Background

Recent studies have demonstrated the importance of chromatin spatial organization in human health by comparing genomic interaction patterns between normal and abnormal cells. However, no systematic studies were published so far on the variability of 3D chromatin structure in human population of healthy individuals. We combine various types of data from a pool of human lymphoblastoid cell lines and computational modeling to address this lack.

## Results

We extend our previous study on the human 3D genome architecture[1] to account for genome topological variation within population and to gain a deeper insight into mechanisms of gene regulation by chromatin topology

We earlier demonstrated that using genomic interaction data a 3D model of an averaged genome structure which recovers architectural features of the genome can be built[2]. We extended our modeling approach (3D-GNOME) to include information on SVs in recovery of 3D chromatin structures. Our algorithm models individual chromatin loops, meaning that the remodeling effect of a genetic variant disrupting a single pair of interacting genomic segments will be represented in the model.

High-quality ChIA-PET data (combined with data from the 1000 Genomes Catalogue of Human Genetic Variation and from the GWAS Catalogue) enables us to perform genome-wide population-scale analysis at the finest resolution. We show different effects genetic variations have on genome regulation when altering chromatin interaction networks mediated by different protein factors. Our results show a close link between variation in chromatin interaction networks mediated by RNA polymerase II and differential gene transcription[3].

## Conclusions

Altogether, our studies show the critical impact of genetic variants on the higher-order organization of chromatin folding and provide insight into the mechanisms regulating gene transcription at the population scale.

[1] Tang, Z., et al. "CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription." *Cell*, December 2015.

[2] Szalaj, P., et al. "An integrated 3-Dimensional genome modeling engine for data-driven simulation

of spatial genome organization." *Genome Res.*, October 2016.

[3] Sadowski, M., et al. "Spatial chromatin architecture alteration by structural variations in human genomes at population scale." *bioRxiv* 266981, September 2018.

## **PgmNr 3321: Targeted bisulfite sequencing: An efficient and streamlined approach to maximize throughput and minimize cost.**

### **Authors:**

I. Andreou <sup>1</sup>; B. Dugan <sup>2</sup>; M.N. Ranjbar <sup>2</sup>; S. Strauss <sup>1</sup>; Q. Peng <sup>2</sup>; S. Huebner <sup>1</sup>; K. Heitz <sup>1</sup>; C. Meerschiff <sup>1</sup>; C. Wach <sup>1</sup>; E. Lader <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) QIAGEN GmbH, Hilden, Germany; 2) QIAGEN Sciences Inc., Frederick, MD, USA

---

Epigenetic changes have a significant impact on human health and disease susceptibility since they play a crucial role in the regulation of important cellular processes. Over the past few years significant progress has occurred in this rapidly advancing field. One of the most stable form of epigenetic regulation is DNA methylation.

Bisulfite sequencing (BS) is considered to be the best method for detecting DNA methylation with single nucleotide resolution. However, the high cost of whole genome BS and the low throughput of amplicon BS are challenges that must be overcome.

Here we present a streamlined workflow which combines optimal bisulfite treatment and targeted bisulfite sequencing. This method utilizes single primer enrichment (SPE) and unique molecular identifier (UMI) barcoding technology. This enables high-efficient, flexible and scalable targeted analysis of DNA methylation from gDNA, FFPE and ccf-DNA samples.

## **PgmNr 3322: Filling in the blanks: Simultaneous imputation of epigenetic profiles using deep neural networks to investigate optimal experimental design in large sequencing studies.**

### **Authors:**

S. Batra <sup>1</sup>; J. Spence <sup>2</sup>; J. Fischer <sup>3</sup>; Y. Song <sup>1,3,4</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Computer Science, University of California at Berkeley, Berkeley, CA.; 2) Department of Computational Biology, University of California at Berkeley, Berkeley, CA; 3) Department of Statistics, University of California at Berkeley, Berkeley, CA; 4) Chan Zuckerberg Biohub, San Francisco, CA

---

Biological data can often be represented via multi-dimensional arrays. For example, expression levels may be measured for a set of genes across both tissues and individuals, or a number of different assays may be performed across the genome in many cell types. These data are often plagued by structural missingness; entire rows or columns may be missing if a given assay has not been performed or if a given tissue was not assayable. If these missing entries could be accurately imputed, it would be possible to infer the results of unperformed experiments, such as measurements in difficult-to-assay tissues, like the brain, by using analogous measurements in easier-to-assay tissues, such as whole blood. Unfortunately, imputing these missing entries is difficult, and most previous approaches assume a low-dimensional linear latent structure. Biological data are often highly non-linear, however, and with massive datasets it should be possible to relax these assumptions. Correspondingly, we have developed a method based on deep convolutional neural networks which captures the nonlinear relationships among missing entries and observed values. Further, to leverage the exchangeability inherent in many biological datasets, we train our model using a novel permutation-equivariant scheme. We then apply our method to a number of different real datasets and demonstrate that our method compares favorably to existing methods and baselines. Our results also allow us to elucidate the effect of different missingness patterns on imputation performance, thus informing optimal experimental design in the presence of technical or resource constraints.



## PgmNr 3323: *In vivo* response to respiratory viruses assessed by whole genome methylation sequencing in infants.

### Authors:

W.A. Cheung; D. Banerjee; R. Biswell; F. Hassan; R. Selvarangan; T. Pastinen; E. Grundberg

[View Session](#) [Add to Schedule](#)

**Affiliation:** Children's Mercy, Kansas City, MO

---

The methylation patterns of approximately 28M CpG dinucleotides in our genome provide dense information on gene regulatory circuitries. Whole genome bisulphite sequencing (WGBS) quantitatively captures the methylation state of individual CpGs. Canonically, methylation levels are low (0-3%) at active promoters and intermediate at active enhancers (5-40%). We studied nasal epithelial cells from 100 infants with suspected respiratory tract infections, either being virus-positive (N=9 different pathogens) or virus-negative as defined by clinical testing. We pooled DNA of 10 patients from each viral group and negative controls, and WGBS (30-40x coverage) from each pool was used to characterize virus-specific epigenome response. We used the MethylSeekR Hidden Markov model algorithm to identify active promoters (UnMethylated Regions, UMRs) and active enhancers (Low Methylated Regions, LMRs) in each pool. On average, 80K regulatory regions were observed per cell pool, however, Influenza B (Flu-B) exposed children showed wide activation (hypomethylation), with >100K regulatory regions present and nearly 15K were unique to this virus. On the other end of the epigenomic response, Influenza A (Flu-A) showed only 2K regions (2.4%) uniquely activated. However, the limited Flu-A response indicated relatively specific enrichment of IL4 signaling pathways (GREAT overrepresentation analysis of GO Biological processes, 7-fold enrichment,  $P < 0.001$ ). Nearly 4K regulatory elements were consistently inactivated (hypermethylation) among patients with documented infection vs. negative patients. These regions were enriched among genes in monocyte activation pathways, suggesting that suppression of these gene networks is important for viral colonization in infant nasal epithelium. Furthermore, we observed strong enrichment of asthma associated genetic markers (SNPs from genome-wide association studies) among these commonly inactivated regulatory elements, potentially linking dysregulated viral response to genetic susceptibility for asthma. Altogether, our genome-wide monitoring of infant viral response provides first catalogue of viral species and infection associated host DNA regulatory elements in early childhood. Future assessments of epigenetic variation in these regions in individual patients may reveal epigenetic evidence for viral triggers for childhood disease.

## PgmNr 3324: Social disadvantage, gene expression and depressive symptoms in a sample of African American from the Minority Health - GRID.

### Authors:

A. Gaye <sup>1</sup>; G. Goodney <sup>1</sup>; M. Abbas <sup>1</sup>; R. Quarells <sup>2</sup>; G. Gibbons <sup>1</sup>; S. Davis <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) National Institutes of Health, Bethesda, MD, USA; 2) Morehouse School of Medicine, Atlanta, GA, USA

### Background

In addition to individual risk factors and behaviors, health is also affected by a range of socio-economic (SES) conditions often measured as income or level of education. More evidence is needed to advance our understanding of how adverse SES conditions are internalized as psychosocial stressors that are manifested as perturbations in biological systems such as patterns of gene expression within the blood circulation.

### Methods

Adverse social conditions were measured as a composite score, social disadvantage, generated from perceived neighborhood environment, level of education, per-capita family income and wealth. Random Forests was used to investigate the relationship between social disadvantage and gene expression, in the whole blood RNA-seq profiles of 180 African Americans and the mediatory role of stress and depression scales in that relationship.

### Results

A set of 191 mRNA could classify social disadvantage with an area under the curve (AUC) of 0.98. Subsets of those 191 mRNA were related to depression scale, perceived stress scale and financial stress scale. The relationships between social disadvantage and the three psychosocial factors were also established through the questionnaire data. A total of 44 of the 191 mRNAs were reported as differentially expressed between depressive subjects and controls in a study of clinical depression (major depressive disorder, MDD).

### Conclusions

The mRNAs identified are involved in mainly 3 areas related to the brain: brain development/neuroplasticity, neurobehavioral response to chronic stress and neurobehavioral and neurodegenerative diseases. The results suggest that subjects above the threshold of depressive symptoms, measured in the data, are likely exhibit a gene expression signature that corresponds to an increased risk of a clinical diagnosis of depression.

## PgmNr 3325: TDG-mediated epigenetic regulation in retinoic acid signaling during spermatogonial differentiation.

**Authors:**

T. Lee; A. Luk; J. Liao; W. Chan

[View Session](#) [Add to Schedule](#)

**Affiliation:** School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin N.T., China

---

Male fertility is sustained by male germline stem cells (GSCs) known as spermatogonial stem cells (SSCs) that either self-renew or differentiate to progenitors to initiate spermatogenesis. A key question in SSC biology is whether and how epigenetic programs contribute to the cellular identity of SSC and specific gene expression programs for committing self-renewal or differentiation. Active DNA demethylation through 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by TDG is essential for mouse embryonic and stem cell development, but its role remains elusive in SSC biology. Therefore, we examined the functional role of TDG in male germ cell development using our recent established germline-specific TDG knockout animal. We found that TDG knockout males expressed a progressive loss of fertility associated with increased PLZF+ undifferentiated spermatogonia and reduced elongated spermatids after 6 months of age. We further traced spermatogonial differentiation using the differentiated spermatogonial cell marker KIT at P8, the initial stage of spermatogonial differentiation and observed that KIT-positive cells were significantly reduced in the knockout testes. To examine the molecular phenotypes of the TDG knockout germ cell, we performed RNA-Seq analysis and revealed altered gene expression of a subset of retinoic acid target genes, including *Stra8* and *Cyp26a1*. Furthermore, we found that the demethylation of *Stra8* during the KIT transition was impaired after depletion of TDG in Oct4-GFP+ undifferentiated spermatogonia. These results indicate that TDG-mediated active demethylation regulates the differentiation of undifferentiated spermatogonia through the coupling of the retinoic acid signaling pathway.

## **PgmNr 3326: Development of programmable CRISPR/Cas9-based histone kinases to link histone phosphorylation to human gene regulation.**

### **Authors:**

J. Li <sup>1</sup>; B. Mahata <sup>1</sup>; I.B. Hilton <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Department of Bioengineering, Rice University, Houston, US; 2) Department of Biosciences, Rice University, Houston, US

---

### **Purpose:**

We have built new synthetic biology tools that enable us to synthetically engineer histone phosphorylation at specific locations in human chromatin. Ultimately this will permit precision control of the epigenome and gene regulation and the amelioration of disease phenotypes in human cells.

### **Method:**

Taking advantage of the programmable CRISPR/Cas9 platform, we fused variants of histone H3 kinases to deactivated Cas9 (dCas9) and delivered these fusion proteins to specific genomic regulatory elements in human cells. We used ChIP-qPCR to measure changes in histone phosphorylation and qPCR/RNA-seq to measure changes to the human transcriptome.

### **Results:**

dCas9-based histone kinases activate human genes when targeted to endogenous promoters. This activation is dependent upon intact kinase activity, as the catalytically inactivated fusion proteins fail to activate gene expression. Furthermore, ChIP-qPCR results show that the phosphorylation levels at specific histone residues are significantly increased coincident with gene activation.

### **Conclusion and Discussion:**

Although numerous important studies have demonstrated a correlation between histone phosphorylation and gene activation, direct evidence of a causal relationship has been lacking. Using the CRISPR/dCas9 system, we have provided the first direct evidence that histone phosphorylation activates the expression of human genes. This activation specific to gRNA-targeted loci and requires intact kinase activity. In addition to gene regulation, histone phosphorylation is associated with many critical biological processes, including DNA replication, DNA repair, and the onset of human diseases. Therefore, our new CRISPR/dCas9-based histone kinases expand the ability mechanistically understand and control human genes, and also enable new ways to explore fundamental biological pathways and model human diseases.

## PgmNr 3327: Delving into epigenetics and gene regulation on the ENCODE portal.

### Authors:

Y. Luo; I. Gabdank; J. Hilton; J. Zamanian; P. Adenekan; U.K. Baymurado; K. Graham; J.P. Jou; J. Lee; K.Z. Lin; C. Litton; Z. Myers; E. O'Neill; P. Sud; F. Tanaka; J.S. Strattan; B.C. Hitz; J.M. Cherry

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Genetics, School of Medicine, Stanford University, Stanford, CA 94305

---

The Encyclopedia of DNA Elements (ENCODE) project is an international collaboration of research groups funded by the NHGRI. The main goal of the project is to identify the functional elements of the human genome. The ENCODE portal (<https://www.encodeproject.org/>), established and maintained by the ENCODE Data Coordination Center (DCC), provides unrestricted access to the data generated by the members of the ENCODE consortium and other related projects such as Model organism ENCODE (modENCODE), Model organism Encyclopedia of Regulatory Networks (modERN), Genomics of Gene Regulation (GGR), and Roadmap Epigenomics Project (Roadmap). The DCC develops and updates the data model and uniform processing pipelines that allow transparency and reproducibility of the experimentally collected and computationally generated data. The categories of metadata and their implementation in a structured, machine-readable data model ensure that the users of the ENCODE portal can search and identify related experiments, understand how assays were performed, and track the provenance of the data. Currently the ENCODE portal hosts data from more than 15 thousand datasets, representing experimental results of more than 40 diverse assay types, performed in multiple cell lines and tissues under various conditions. The portal provides free access to over 640 terabytes of raw experimental, processed, and analysis data files with the corresponding metadata.

Richness, diversity, and structure of the data on the ENCODE portal make it an invaluable resource for the scientific community. The presence of curated and standardized experimental results from ChIP-seq, DNase-seq, ATAC-seq, HiC and ChIA-PET assays, that were uniformly processed, makes the portal into a one-stop shop for a chromatin investigator. Here we show how the ENCODE portal can be used as a multifaceted lens for chromatin exploration.

## **PgmNr 3328: Pediatric age acceleration: Potential biomarker for individual differences on the effects of maternal mental health on child development.**

### **Authors:**

M.G. McGill <sup>1</sup>; L.M. McEwen <sup>2</sup>; M. Kobor <sup>2</sup>; N. Karnani <sup>3</sup>; M. Kee <sup>3</sup>; I. Pokhvisneva <sup>1</sup>; M. Oldach <sup>1</sup>; E. Garg <sup>1</sup>; M.J. Meaney <sup>1,3</sup>; K.J. O'Donnell <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) McGill University; 2) University of British Columbia; 3) National University of Singapore

---

Maternal anxiety is associated with adverse behavioural and socioemotional outcomes that vary markedly at the level of the individual child. There is a need for the development of biomarkers of exposure that reflect variation of this impact at the level of the individual child. Epigenetic clocks, which derive estimates of biological age on the basis of DNA methylation, have emerged as clinically relevant biomarkers that predict adverse health outcomes in adults. Error within more well-established epigenetic clocks ( $\geq 3.6$  years) limits predictive power for understanding child development. A new epigenetic clock specifically designed for use in children, the pediatric clock (error = 0.35 years), has recently been developed and shown to associate with measures of child neurodevelopment. In this project, we used longitudinal DNA methylation data to examine how prenatal and postnatal maternal anxiety scores associated with epigenetic aging in children. All pediatric clock analyses were benchmarked against the conventional epigenetic clock. We hypothesized that increased maternal anxiety would show a corresponding increase in pediatric epigenetic aging and that pediatric clock estimates would provide a more robust association than using conventional epigenetic clock estimates. Generalized linear models indicated an association between increased prenatal maternal anxiety and accelerated epigenetic aging in both the pediatric clock and the conventional epigenetic clock. These analyses highlight the potential for the pediatric clock as a biomarker for variation in effects of maternal prenatal anxiety at the level of the individual child.

## PgmNr 3329: EM-seq enables accurate and precise methylome analysis of challenging DNA samples.

### Authors:

L. Williams; V.K.C. Ponnaluri; B. Sexton; L. Saleh; K. Marks; M. Samaranayake; Z. Sun; M. Campbell; B. Langhorst; T.C. Evans; R. Vaisvila; E. Dimalanta; T.B. Davis

[View Session](#) [Add to Schedule](#)

**Affiliation:** New England Biolabs, Ipswich, Massachusetts.

---

DNA isolated from blood draws (cell-free DNA (cfDNA)) or from archival material like formalin fixed paraffin embedded (FFPE) tissues have advanced the field of cancer genetics. DNA methylation (5-methylcytosines (5mC) and 5-hydroxymethylcytosines (5hmC)) is a key epigenetic factor that plays an important role in cellular processes and it's misregulation results in diseased states like cancer. Advances in the field of sample preparation from biological matrices and genomics have enabled cancer biomarker identification based on methylation profiling.

Bisulfite sequencing is the standard method to detect methylation and has been employed for both targeted and whole genome methylation analysis. However, the chemical based bisulfite conversion of cytosines to uracils also results in DNA damage which subsequently results in shorter DNA insert sizes as well as introducing bias into the data. Hence, analysis of DNA methylation from cfDNA and FFPE DNA is challenging as the DNA is typically of low quality and quantity. Robust biomarker detection relies primarily on the ability to profile methylation accurately. To overcome the drawbacks of bisulfite sequencing, we developed an enzyme based methylation detection technology, called NEBNext® Enzymatic Methyl-Seq (EM-Seq™), that minimizes damage to DNA enabling longer insert sizes, lower duplication rates and minimal GC bias resulting in more accurate quantification of methylation in the sample DNA.

Using EM-Seq, we profiled clinically relevant cfDNA and FFPE DNA from multiple tissue types. Results for these challenging DNA types showed that the EM-Seq libraries had longer inserts, lower duplication rates, higher percentages of mapped reads and less GC bias compared to WGBS libraries. These libraries also identified a higher number of CpGs and the estimated global methylation levels were in good agreement with the absolute levels quantified using LC/MS. In conclusion, EM-Seq libraries have superior sequencing metrics resulting in robust methylation profiling for these types of challenging DNA samples.

## PgmNr 3330: OSCA: A tool for omic-data-based complex trait analysis.

### Authors:

F. Zhang<sup>1</sup>; W. Chen<sup>1</sup>; Z. Zhu<sup>1</sup>; Q. Zhang<sup>1</sup>; M.F. Nabais<sup>1,2</sup>; T. Qi<sup>1</sup>; I.J. Deary<sup>3</sup>; N.R. Wray<sup>1,4</sup>; P.M. Visscher<sup>1,4</sup>; A.F. McRae<sup>1</sup>; J. Yang<sup>1,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, 4072, Australia; 2) University of Exeter Medical School, Devon, EX2 5DW, UK; 3) Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK; 4) Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, 4072, Australia; 5) Institute for Advanced Research, Wenzhou Medical University, Wenzhou, 325027, Zhejiang, China

---

The rapid increase of omic data has greatly facilitated the investigation of associations between omic profiles such as DNA methylation (DNAm) and complex traits in large cohorts. Here, we propose a mixed-linear-model-based method called MOMENT that tests for association between a DNAm probe and trait with all other distal probes fitted in multiple random-effect components to account for unobserved confounders. We demonstrate by simulations that MOMENT shows a lower false positive rate and more robustness than existing methods. MOMENT has been implemented in a versatile software package called OSCA together with a number of other implementations for omic-data-based analyses.



## PgmNr 3331: DNA hydroxymethylation levels may explain environment-specific epigenetic changes: An example of smoking-discordant monozygotic twin pairs by more than 10 pack years.

### Authors:

S.J. Lee <sup>1</sup>; E. Kim <sup>2</sup>; H.Y. Shin <sup>2</sup>; J. Kim <sup>2</sup>; S. Li <sup>3</sup>; J. Hopper <sup>3</sup>; J. Lee <sup>1</sup>; J. Sung <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Institute of Health and Environment, Seoul National University, Seoul, Republic of Korea; 2) Department of Epidemiology, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea; 3) Centre for Epidemiology & Biostatistics, Melbourne School of Population & Global Health, The University of Melbourne, Victoria, Australia

Unlike genetic influences on DNA methylation (Gen-DNAM), the DNAm specific to environmental exposure (Env-DNAM) is less reproducible, mainly due to the composite and cumulative nature of environmental influences. The authors first hypothesized that the presence of ubiquitous genetic influences further confounds the Env-DNAM of the methylome-wide association studies (MWAS). To test our hypothesis, authors dissected 18,496 smoking-associated epiloci from a large population study into Env-DNAM and Gen-DNAM. 1) Env-DNAM: an MWAS involving 33 smoking-discordant monozygotic twin pairs (MZs) with more than a 10 pack-year difference; and, 2) Gen-DNAM: methylation quantitative trait loci (MeQTLs) for 534 participants. Presumably, pure Env-DNAM, if exists, would evade MeQTL controls but show strong signals in the discordant MZ comparison.

About 21% (3,865 out of 18,496) of the candidate smoking-associated epialleles were under strong MeQTLs, and 842 loci were replicated for the smoking-discordant MZ analysis (Bonferroni corrected  $p < 0.05$ ). A reverse correlation existed between the extent of Gen-DNAM ( $= -\log(p)$  of MeQTL) and Env-DNAM ( $= -\log(p)$  of MZ control) for the same CpG probes. Notably, CpG loci on the *AHRR* (hypo-), *MYO1G* (hyper-) and *F2RL3* (hypomethylated) were the top hits from the MZ analysis with none or weak genetic influences.

We further hypothesized that the mechanisms of the Env-DNAM being initiated by the over-expression of genes, either through active demethylation or through increases in gene copies. We did not find any evidence that the copy number of the CpG sites differs, which negated the possibility that the association might have resulted from gene copy changes. We measured 5-hydroxymethylcytosine (5hmC) levels as a marker of active demethylation, which is indistinguishable from plain methylation (5-methylcytosine, 5mC) in conventional analysis. For hypomethylated loci in *AHRR* (smokers) showed significant decreases in 5hmC, whereas hypermethylated loci in *MYO1G* showed random patterns in 5hmC.

In conclusion, our findings suggest that even with a mild correlation with exposure, genetics may paradoxically comprise a large part of environment-DNAM associations further complicating the complex Env-DNAM; and, for the specific Env-DNAM changes, the active demethylation process of

over-expressed genes might provide one mechanical explanation for the hypomethylated changes.

## PgmNr 3332: A database of experimentally tested regulatory relationships relevant to brain development and disorder.

### Authors:

E. Chu <sup>1,2,3</sup>; A. Morin <sup>1,2,3</sup>; T. Chang <sup>2</sup>; T. Nguyen <sup>2</sup>; A. Sharma <sup>2</sup>; C. Liu <sup>2</sup>; Y. Tsai <sup>2</sup>; P. Pavlidis <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Psychiatry, UBC, Vancouver, BC, Canada; 2) Michael Smith Laboratories, UBC, Vancouver, BC, Canada; 3) Graduate Program in Bioinformatics, UBC, Vancouver, BC, Canada

---

Disruption in transcriptional regulation during early brain development is implicated in neurodevelopmental disorders. For example, mutations in several transcriptional regulators including TBR1, MYT1L, and CHD8 have been causally linked to autism. Genome-wide regulatory network models reconstructed using high-throughput functional genomics data have immense potential in identifying the transcriptional mechanisms associated with the etiology of neurodevelopmental disorders. Due to the lack of a reliable gold standard dataset of direct regulatory relationships, it is currently difficult to evaluate and effectively utilize these predicted network models. We have undertaken the first large-scale attempt to assemble a database of experimentally tested direct regulator-target relationships with relevance to the developing brain by mining the published literature. For each one-to-one relationship, we establish confidence by finding and integrating multiple lines of experimental evidence obtained from transcription factor (TF) perturbation, TF binding, and reporter gene assays. To date, we have identified more than 1000 direct regulatory relationships involving approximately 200 transcription factors. Curated factors include key regulators of brain development such as PAX6, SOX2, and POU5F1. We estimate that we have captured approximately 25% of all experimentally tested relationships relevant to the development of the central nervous system (CNS). Moreover, most regulatory relationships in our database are supported by at least two independent lines of experimental evidence. More than 100 of the relationships include evidence of specific activity in the embryonic mouse CNS. The results of our curation have highlighted that the research literature is dominated by investigations into a small number of transcription factors. Finally, to demonstrate a use case for this resource, we show that a set of putative targets regulated by Pax6 in the embryonic mouse forebrain, detected by intersecting publicly available ChIP-seq and TF knockout transcriptomic datasets, is significantly enriched for the target genes recorded in our database. In summary, we provide a set of experimentally tested, high confidence regulatory relationships to support the inference of large-scale regulatory networks using high-throughput approaches as well as more reliable investigations into the transcriptional regulatory mechanisms involved in neurodevelopmental disorders.

## PgmNr 3333: Functions and dependency of enterocyte-specific enhancers at the *CFTR* locus.

### Authors:

S. Yin; M. NandyMazumdar; J.L. Kerschner; S.H. Leir; A. Harris

[View Session](#) [Add to Schedule](#)

**Affiliation:** Case Western Reserve University, Cleveland, OH.

---

The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene is a large gene with multiple tissue-specific *cis*-regulatory elements (CREs) located in introns and intergenic regions. Among these are enhancers that recruit transcription factors (TF) governing the cell-type specific expression of *CFTR*. We showed previously that the boundaries of the *CFTR* topologically associating domain (TAD) are conserved among diverse cell types, and are dependent on CCCTC-binding factor (CTCF) and the cohesin complex. Within the TAD, enhancers coordinate the loading of TFs onto the gene promoter by cell-specific looping mechanisms. Here we focus on the interactions and dependency of multiple CREs in one cell type: enterocyte-like intestinal epithelial cells. We removed CREs individually and in pairs and measured the effects on *CFTR* expression by RT-qPCR and locus architecture by circular chromatin conformation capture followed by deep sequencing (4C-seq). Enhancers looping is known to be dependent on CTCF recruitment, though non-CTCF binding enhancers may also impact chromatin structure. We identified a new role for a well-studied non-CTCF binding enhancer in the first intron of *CFTR*, showing it is critical for both chromatin structure and gene expression. Another strong enhancer in intron 11 is essential for enterocyte expression of *CFTR*, but has little impact on locus architecture. Combined deletion of these two enhancers abolished *CFTR* transcription and caused a major reorganization of locus structure. Our data suggest the cooperation of enhancers with different roles in regulating *CFTR* transcription: the first has an active role in modulating chromatin structure but modest enhancer function; the second is a strong enhancer with apparently little independent impact on locus architecture. Current experiments are aimed at elucidating the cooperative mechanism of these two intronic enhancers in intestinal epithelial cells and may have broader application to the function of enhancers genome-wide.

This work is supported by the NIH (R01 HL094585) and CFF (Harris 16G0).

## **PgmNr 3334: Chromatin state annotations across 833 tissue/cell types help annotate thousands of additional GWAS loci.**

### **Authors:**

C. Boix <sup>1,2</sup>; W. Meuleman <sup>3</sup>; Y. Park <sup>1,2</sup>; M. Kellis <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) EECS, Massachusetts Institute of Technology, Cambridge, Massachusetts.; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Altius Institute for Biomedical Sciences, Seattle, WA

---

Large-scale community projects have generated rich epigenomic datasets, with either broad coverage (hundreds of cell types for some assays) or deep coverage (dozens of assays for some cell types). However, owing to cost, time, sample availability, or limited experiment success rates, it has not always been possible to map every assay in every cell type.

Here, we enhance and complete the matrix of ENCODE datasets to enable integrative analyses of gene regulation and disease mechanism, by aggregating, processing, and imputing 30+ epigenomic assays across 800+ human cell types. These include ChIP-seq on histone marks, chromatin regulators, and DNase-seq, painting a rich picture of the non-coding landscape in each cell type.

We combine the resulting imputed maps with experimentally-observed maps to annotate regulatory elements in each cell type using an 18-state chromatin state model with ChromHMM, including enhancer, promoter, repressed, bivalent, and transcribed regions.

We cluster the resulting regulatory annotations across cell lines to infer functional modules of co-regulated elements, and to distinguish constitutive vs. lineage- and cell-specific regions. We also predict driver motifs and co-regulatory motifs in each cluster based on their enrichment in digital genomic footprints (DGF) and surrounding regions. We prioritize cell types for traits and diseases by evaluating GWAS enrichment in enhancers from both full epigenomes and functional modules. We find that novel epigenomes help prioritize an additional 33% GWAS studies over existing annotations and are the strongest associated samples in 64% of all GWAS with significant enrichments.

Lastly, we characterize transitions between cell types and states by their changes in enhancer, promoter, and repressed modules, and we use these transitions to link enhancers to their target genes using coordinated expression changes, and to predict key transcription factors and chromatin regulators involved in cell fate transitions.

## PgmNr 3335: Isolating protein-occupied open chromatin regions.

### Authors:

T. McDonald <sup>1</sup>; R. Hurto <sup>2,3</sup>; P. Freddolino <sup>2,3</sup>; A. Boyle <sup>1,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Human Genetics, University of Michigan, Ann Arbor, MI.; 2) Department of Biological Chemistry, University of Michigan, Ann Arbor, MI; 3) Department of Computational Medicine and Bioinformatics

---

Modern technology and recent advances in genomics have given rise to a variety of high-throughput methods that interrogate regulatory function of the human genome. In particular, the advent of techniques that measure open chromatin have enabled the identification of loci of active regulatory elements. These methods, in addition to mapping open chromatin, have been widely used to characterize footprints of transcription factors (TFs). The ability to infer TF binding motifs from footprints is dependent upon the lack of sequencing coverage over the footprint (strength of binding of TF) within open chromatin regions. This presents an unmet challenge for critical TFs that do not associate as strongly or as frequently with their binding sites. Indeed, it has been well documented that some TFs present very clear footprints (e.g. CTCF) while others lack any discernable difference in chromatin accessibility and footprinting techniques for these perform no better than a basic motif match. Here we present an orthogonal open chromatin assay that isolates TF binding footprints to be made into sequencing libraries. Briefly, we use micrococcal nuclease to extensively digest crosslinked chromatin to mononucleosomal size and extract the low molecular weight fragments. These fragments represent protein-occupied open chromatin regions protected from nuclease digestion. Analysis of these regions will allow us to finely map functional units within open chromatin and better understand how these regions are utilized in eukaryotic gene regulation.

## PgmNr 3336: Ensemble imputation for DNA methylation levels across platforms.

### Authors:

G. Li <sup>1</sup>; L. Raffield <sup>2</sup>; M. Logue <sup>3,4</sup>; M. W Miller <sup>3,4</sup>; H. P. Santos Jr. <sup>7</sup>; T. M. O'Shea <sup>8</sup>; R. C. Fry <sup>9</sup>; Y. Li <sup>2,5,6</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Dept. of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Dept. of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) National Center for PTSD at VA Boston Healthcare System; 4) Dept. of Psychiatry, Boston University School of Medicine; 5) Dept. of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 6) Dept. of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC; 7) School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC; 8) Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 9) Department of Environmental Sciences and Engineering, University of North Carolina at Chapel Hill, Chapel Hill, NC

---

DNA methylation at CpG dinucleotides is one of the most extensively studied epigenetic marks due to its control of gene expression and critical importance in normal development. With technological advancements (e.g., DNA methylation microarrays and bisulfite sequencing), geneticists can profile DNA methylation at increasingly higher resolutions. However, different DNA methylation profiling platforms differ in their resolutions, hindering joint analysis. We have proposed a penalized functional regression (PFR) model to impute missing-by-design methylation levels and demonstrated the power of our PFR approach: it excelled alternative methods when imputing the Illumina HumanMethylation450 (HM450) BeadChip from the HumanMethylation27 (HM27) BeadChip. Here, we extend our method to impute from HM450 to ~850K CpG sites on the HumanMethylationEPIC (HM850) BeadChip, by ensembling with 4 additional methods, K-nearest-neighbors, logistic regression, random forests, and XGBoost, to further increase imputation accuracy. We analyzed data from two population cohorts measured both by HM450 and HM850: the Extremely Low Gestational Age Newborns (ELGAN) study ( $n=127$ ) and the Posttraumatic Stress Disorder (PTSD) study ( $n=144$ ). We assess three datasets: ELGAN and PTSD separately, and a combined dataset with batch effects corrected via the ComBat R function. The cross-validation results show that there is no uniformly best imputation method across the three datasets, which inspires the use of ensemble method. For example, in the PTSD dataset, the ensemble method achieves the lowest predicted MSE (0.0007) and the highest accuracy (99.97%) compared with individual methods. Specifically, the predicted MSE of the ensemble method improved by 28.5% compared with the second-best method (PFR), by 85.7% compared with random forest. Among all 339,033 EPIC-only CpG sites shared between ELGAN and PTSD after quality control, 289,604 (85.4%) exceeding a threshold of 0.0025 for the probe-level predicted MSE and a threshold of 95% for the probe-level predicted accuracy when dichotomizing DNA methylation level at a cutoff of 0.5. In summary, our data highlight that the ensemble method outperforms all the individual methods when imputing from HM450 to HM850, which will boost power for discovery in subsequent epigenome-wide association studies.

## PgmNr 3337: A pipeline to predict prenatal air pollution exposures using DNA methylation data.

### Authors:

Y. Wang; J. Guo; J. Herbstman; F. Perera; S. Wang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Columbia University, New York, NEW YORK.

---

DNA methylation has been associated with air pollution such as prenatal exposure to polycyclic aromatic hydrocarbons (PAH), fine particulate matter (PM<sub>2.5</sub>), nitrogen dioxide (NO<sub>2</sub>), etc. As air pollution leave traces on a person's epigenome, it is of interest to use DNA methylation measures to recover exposure history. Using the 432 newborns from the mother/newborn (M/N) cohort of the Columbia Center for Children's Environmental Health (CCCEH) whose cord blood DNA methylation measures using Illumina arrays and pregnancy PM<sub>2.5</sub> and NO<sub>2</sub> are available, we developed a pipeline that identifies cord blood DNA methylation signatures that predict binary (high/low) prenatal exposure to air pollutants. The pipeline starts by randomly splitting the M/N cohort into training (50%) and testing (50%) sets and then selects 500 top ranked CpGs that differ between the high/low exposure groups by two-sample t-test p-values. Random Forest (RF) is then applied using CpGs from the selected 500 CpGs with a grid search increasing number of CpGs ranked by importance measures. The aim is to select the optimal number of informative CpGs that best predict air pollution levels. The procedure was repeated 200 times and the final air-pollution predictor is the RF model with the highest mean area under the receiver operating curve (AUC). This pipeline generated two good (significant prediction with a 95% confidence interval not covering random chance) air pollution predictors, a predictor of high/low NO<sub>2</sub> exposure during the entire pregnancy with 130 CpGs (testing set mean AUC=0.605, 95% CI=(0.517, 0.687)) and a predictor of high/low 3<sup>rd</sup> trimester NO<sub>2</sub> exposure with 160 CpGs (testing set mean AUC=0.602, 95% CI=(0.506, 0.693)). Using an internal validation cohort of siblings of the M/N children, we demonstrated the predictability of selected DNA methylation signatures for NO<sub>2</sub> during these pregnancy periods (entire pregnancy AUC=0.561 (0.525, 0.593), 3<sup>rd</sup> trimester AUC=0.587 (0.553, 0.622)). These results demonstrate that this newly developed pipeline has potential to be used to successfully recover exposure history of air pollutants.



## PgmNr 3338: Transcribed cis-regulatory elements across primates.

### Authors:

H. Kawaji<sup>1,2,3</sup>; N. Hirose<sup>3</sup>; M. Ema<sup>4</sup>; E. Sasaki<sup>5</sup>; J. Kawai<sup>2</sup>; Y. Murakawa<sup>3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) RIKEN Preventive Medicine and Diagnosis Innovation Program, Yokohama, Japan; 3) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 4) Shiga University of Medical Science, Otsu, Japan; 5) Central Institute for Experimental Animals, Kawasaki, Japan

---

Unveiling regulatory roles of non protein coding regions within the human genome has been a major challenge in assessment of the impact of genome variations on gene expression underlying physiology and disease states. The FANTOM5 consortium identified promoters and transcribed enhancers by monitoring transcription initiation activities in a wide range of samples consisting of primary cells, tissues, and cell lines (Forrest et al 2014; Andersson et al 2014), by using CAGE (Cap Analysis of Gene Expression) that determine capped-5' ends of RNAs in a nucleotide resolution (Kanamori-Katayama et al. 2011). We subsequently found that only a limited fraction of human cis-regulatory elements, in particular enhancers, have corresponding regions in mouse. The lack of such corresponding regions hinders analysis of cis-regulatory regions from the aspect of transcriptional activity and evolution.

To cope with the problem, we mapped cis-regulatory elements in two species of non-human primates (NHPs), crab-eating macaques (*Macaca fascicularis*) and common marmosets (*Callithrix jacchus*) by using CAGE. We collected more than hundreds of tissue samples from the both species, and found a similar magnitude of promoters and enhancers to human. Our data indicated that a larger fraction of corresponding regions to the human promoters and enhancers are transcriptionally active in the non-human primates than mouse. Our data set provides a unique opportunity to study human cis-regulatory elements from both of the functional and evolutionary aspects.

## **PgmNr 3339: Extent, heritability, and functional relevance of single cell expression variability in highly homogeneous populations of human cells.**

### **Authors:**

J. Cai; D. Osorio; X. Yu; Y. Zhong; G. Li; P. Yu; E. Serpedin; J. Huang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Texas A&M University, College Station, TX

---

**INTRODUCTION:** Because of recent technological developments, single-cell assays such as single-cell RNA sequencing (scRNA-seq) have become much more widely available and have achieved unprecedented resolution in revealing cell heterogeneity. The extent of intrinsic cell-to-cell variability in gene expression, or single cell expression variability (scEV), has thus been increasingly appreciated. However, it remains unclear whether this variability is functionally important and, if so, what its implications are for multi-cellular organisms.

**METHODS:** We therefore analyzed multiple scRNA-seq data sets from lymphoblastoid cell lines (LCLs), lung airway epithelial cells (LAECs), and dermal fibroblasts (DFs). For each of the three cell types, we estimated scEV in homogeneous populations of cells; we identified 465, 466, and 291 highly variable genes (HVGs), respectively.

**RESULTS:** These HVGs were enriched with specific functions precisely relevant to the cell types, from which the scRNA-seq data used to identify HVGs were generated---e.g., HVGs identified in lymphoblastoid cells were enriched in cytokine signaling pathways, LAECs collagen formation, and DFs keratinization. HVGs were deeply embedded in gene regulatory networks specific to corresponding cell types. We also found that scEV is a heritable trait, partially determined by cell donors' genetic makeups. Furthermore, across genes, especially immune genes, levels of scEV and between-individual variability in gene expression were positively correlated, suggesting a potential link between the two variabilities measured at different organizational levels.

**CONCLUSION:** Our results support the 'variation is function' hypothesis, which argues that scEV is required for higher-level system function. Quantifying and characterizing scEV in relevant cell types may deepen our understating of normal as well as pathological cellular processes.

## PgmNr 3340: Genome-wide unliganded estrogen receptor 1 (ESR1) binding in human liver.

**Authors:**

J. Collins; D. Ling; D. Wang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Pharmacotherapy and Translational Research, College of Pharmacy, University of Florida, Gainesville, FL, 32610.

---

The estrogen receptor 1 (ESR1) has been implicated in the transcriptional regulation of thousands of genes. Canonically, ESR1 is activated by its ligand estradiol (E2). However, the role of unliganded ESR1 in transcriptional regulation has been gaining attention in the last decade. For example, ChIP-Seq experiments detailing ESR1 DNA-binding sites in breast cancer cells implicated distinct binding profiles for the liganded and unliganded forms of ESR1, and that unliganded ESR1 plays a key role in transcriptional regulation of genes with developmental function (Caizzi et al. 2014; Bojcsuk et al., 2019). Previous results generated in our lab indicate that unliganded ESR1 also plays a role in liver cells, functioning as a major regulator of several cytochrome P450 (CYP) genes (Wang et al., under review). Also, alternative ESR1 isoforms lacking N-terminal and C-terminal motifs known to exhibit altered transactivation and DNA-binding profiles (Flouriott et al., 2000; Hattori et al., 2017) are predominant in liver tissue (Sun & Wang, unpublished). In an effort to improve our understanding of transcriptional regulation by ESR1 in the liver, ESR1 ChIP-Seq was conducted in human hepatocytes and in liver tissue. To differentiate between liganded- and unliganded-ESR1 binding sites, hepatocytes were treated with or without 17beta-estradiol (E2) for 12 hrs. Preliminary analyses indicate approximately 700 ESR1 binding sites shared among all samples and that ESR1 enrichment occurs at proximal promoter regions, the 5'UTR, the introns of genes and intergenic regions. Interestingly, similar to studies in breast-cancer cells, unliganded ESR1 bound to 679 unique sites. Furthermore, in both the liver and E2-treated samples, the Estrogen Response Element (ERE) was identified as an ESR1-enriched motif, but not in the unliganded ESR1 sample. This indicates that in the liver, binding of unliganded ESR1 to DNA may instead occur through interaction with additional transcription factors, as suggested by Bojcsuk et al. (2019) and supported by our previous data (Wang et al., under review). Overall, these results further support a role of unliganded ESR1 in transcriptional regulation in liver. Characterization of ESR1 genomic occupancy and identifying genes associated with differential ESR1 binding will provide insight into the function of ESR1 in liver and its implications in liver drug metabolism and liver diseases.

## PgmNr 3341: Haplotype-specific binding among by transcription factors in human brain tissue.

### Authors:

B. Moyers <sup>1</sup>; J. Loupe <sup>1</sup>; S. Felker <sup>2</sup>; W. Bunney <sup>3</sup>; P. Cartegena <sup>3</sup>; S. Watson <sup>4</sup>; G. Cooper <sup>1</sup>; R. Myers <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) HudsonAlpha Institute for Biotechnology, Huntsville, AL.; 2) Department of Biotechnology, University of Alabama Huntsville, Huntsville AL; 3) Department of Psychiatry and Human Behavior, University of California, Irvine CA; 4) The Molecular & Behavioral Sciences Institute, University of Michigan, Ann Arbor MI

---

Gene expression variation is crucial to both normal biology and disease, and can be mediated by transcription factors (TFs) binding to cis-regulatory sequences. Genetic variation that alters TF binding can thereby lead to altered gene expression levels. However, finding variants that alter TF binding is a challenge. To characterize the binding patterns of TFs in human brains, the HudsonAlpha Brain TF Project is performing ChIP-seq on up to 100 TFs in 9 anatomically-defined brain regions from multiple donors. This project has also generated other datasets, including RNA-seq and phased genome sequences. We have thus far generated 250 distinct maps of TF binding for 15 TFs. We have analyzed these data to search for variants that alter TF binding. In particular, we used the phased genome sequences to create haplotype-specific reference genomes against which ChIP-seq reads were aligned. We then identified loci in which a donor is heterozygous for one or more variants and there exists a haplotype-specific bias in the number of ChIP-seq reads. As our approach integrated phased genomic variants into a reference genome, we greatly reduced reference-haplotype alignment bias and can also detect the effects of multiple nearby variants in cis. Even after stringent FDR corrections, we find hundreds of cases of haplotype-specific binding in each donor, many of which occurred in promoters. Variants altering TF-binding in one tissue often had consistent effects in other tissues, supporting their sequence-dependent effects on TF preference. Similarly, specific haplotypes frequently showed congruent TF bias across donors, further supporting robustness of our analysis. We intersected these events with the location of TF-binding motifs, and discovered that the production or destruction of a motif frequently had an unexpected impact on binding. We also examined deltaSVM and CADD scores to determine how well these capture TF binding effects. We plan to integrate this data with phased variants in RNA data to identify cases where haplotype-specific binding is associated with a change in mRNA production, and generate a more comprehensive view of genetic effects on gene regulation in human brain.

## PgmNr 3342: Sex differences at the molecular level: Lessons from the human transcriptome.

### Authors:

M. Oliva<sup>1</sup>; S. Kim-Hellmuth<sup>2,3</sup>; M. Muñoz-Aguirre<sup>4</sup>; V. Wucher<sup>4</sup>; A.D. Gewirtz<sup>5</sup>; P. Parsana<sup>6</sup>; D. Cotter<sup>7</sup>; B. Balliu<sup>8</sup>; F. Aguet<sup>9</sup>; Y. Zou<sup>10</sup>; R. Mohammadi<sup>11</sup>; S.E. Castel<sup>2,3</sup>; S. Kasela<sup>2,3</sup>; H.K. Im<sup>1</sup>; M. Stephens<sup>10</sup>; R. Guigo<sup>4</sup>; K. Ardlie<sup>9</sup>; T. Lappalainen<sup>2,3</sup>; A. Battle<sup>6</sup>; S.B. Montgomery<sup>7,12</sup>; B.E. Engelhardt<sup>14</sup>; B.E. Stranger<sup>1,13</sup>; GTEx Sex subgroup, GTEx Consortium

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medicine, University of Chicago, Chicago, IL; 2) Department of Systems Biology, Columbia University, New York, NY; 3) New York Genome Center, New York, NY; 4) Centre de Regulacio Genomica, Barcelona, Spain; 5) Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ; 6) Department of Computer Science, Johns Hopkins University, Baltimore, MD; 7) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 8) Department of Biomathematics, University of California-Los Angeles, Los Angeles, CA; 9) Broad Institute, Cambridge, MA; 10) Department of Statistics, University of Chicago, Chicago, IL; 11) The Scripps Translational Science Institute, The Scripps Research Institute, La Jolla, CA; 12) Department of Pathology, Stanford University School of Medicine, Stanford, CA; 13) Institute for Genomics and Systems Biology, Chicago, IL; 14) Computer Science Department, Princeton University, Princeton, NJ

To understand the basis of organismal sex differences, characterization of sex differences at the molecular level is warranted. Here, we explore the role of sex on a variety of transcriptome-related phenotypes in 44 tissues of the Genotype+Tissue Expression (GTEx) project (v8 release). We performed differential gene expression analysis and identified 12,627 (FDR<0.05) sex differentially expressed genes (sexDEGs) in at least one tissue. Tissue-shared sexDEGs are generally more highly expressed in females and enriched for X-linked genes. Across-tissue meta-analysis (*MASH*, FDR<0.05) identifies 20,023 sexDEGs and higher tissue sharing: sexDEGs shared across all 44 tissues increase from dozens to hundreds, ~93% of which are located on autosomes and distribute non-randomly across the genome. To characterize tissue similarities based on sex-biased gene expression, we compared tissue clusters based on sexDEG effect sizes to those based on transcriptome abundances. Tissues cluster similarly by both metrics, but we observe differences in the strongly supported clustering of immune-rich tissues, which are very similar at the transcriptome level but notably different for sexDEGs. We identify 44 male- and 16 female-specific gene networks across tissues with *BicMix*, including breast sex-specific networks enriched for milk secretion and hormone genes. Overall, breast is the most sex-differentiated tissue. Based on *xCell* cell-type estimates, endothelial and epithelial cells are more abundant in male and female breast, respectively. For cis-eQTLs derived by the GTEx Consortium, we performed a genotype-by-sex interaction test and identified 445 sex-biased eQTLs (sb-eQTLs), ~70% in breast (FDR<40%), of which 24% validate with allelic fold change. More than 50% of breast sb-eQTLs are mediated by cell abundances, suggesting these sb-eQTLs are eQTLs specific to cell types(s) that correlate with sex. To explore sex differences in the genetic architecture of disease, we performed colocalization analysis (*coloc*) between eQTLs and summary statistics from 87 GWASs and identified dozens of sex-biased colocalized loci (PP4>0.5). For HKDC1

and CRLF3, birth weight and breast cancer GWAS signals colocalize with female ( $PP4=0.83, 0.66$ ) but not male ( $PP4=0.07, 0$ ) eQTLs in liver and breast, respectively. Collectively, our integrative analyses provide the most comprehensive characterization to date of sex differences in the human transcriptome, with implications for complex traits.

## PgmNr 3343: Do *Alu* elements in intron affect alternative splicing? Identification of minimum sequence of inducing downstream exon skipping within antisense partial *AluSx* element in a minigene model.

### Authors:

M. Nakama <sup>1,2</sup>; H. Otsuka <sup>2</sup>; Y. Ago <sup>2</sup>; H. Sasai <sup>1,2</sup>; E. Abdelkreem <sup>2,3</sup>; Y. Aoyama <sup>2,4</sup>; T. Fukao <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Clinical Genetics Center, Gifu University Hospital, Gifu, Japan; 2) Department of Pediatrics, Graduate School of Medicine, Gifu University, Gifu, Japan; 3) Department of Pediatrics, Faculty of Medicine, Sohag University, Sohag, Egypt; 4) Education and Training Center of Medical Technology, Chubu University, Japan

**BACKGROUND:** *Alu* elements are known as primate-specific repeats and commonly locate in introns. We have been analyzing effects of *Alu* elements on alternative splicing. We recently reported that antisense *AluSx* inserted in intron 9 affected exon 10 skipping in a minigene model consisting of exons 9–11 of the *ACAT1*, especially when splice acceptor sites are suboptimal (Nakama et al., 2018). Here, we further examined to identify the minimum sequence elements in the intronic antisense *AluSx* affecting RNA splicing in the same model.

**METHODS:** We have a minigene splicing vector consisting of exon 9-intron 9-exon 10-intron 10-exon11 of *ACAT1* in a pCAGGS. An *AluSx* originally from intron 5 of *ACAT1* gene was inserted into intron 10 (-226 bp from the beginning of exon 10) with antisense-direction in the above vector. The *AluSx* consists of two similar, but distinct monomers-the left and the right arms. Hence first we made antisense left arm or right arm constructs separately to examine which arms could affect splicing. We then generated further fifteen deletion constructs. Each expression vector was transfected into SV40-transformed fibroblasts from a patient with *ACAT1* deficiency. RNA was harvested from the transfected cells and was reverse-transcribed by strand-specific primer and oligo(dT) primer. Chimeric cDNAs (fusion of human *ACAT1* with rabbit beta-globin) were amplified.

**RESULTS:** Exon 10 skipping was more effectively observed in the antisense right arm than full-length antisense *AluSx*. Analysis of successive deletion mutants confirmed a minimum sequence element of 23 bp within the right arm. Exon 10 skipping was less effectively observed in the antisense left arm only than full-length antisense *AluSx*. The corresponding region to the minimum sequence element of 23 bp within the right arm was 22 bp in the left arm and also responsible for exon 10 skipping. These two minimum sequence elements in *AluSx* had a character of splice acceptor site sequence.

**DISCUSSION:** We identified minimum sequence elements, 23 bp and 22 bp in the right arm and left arm, respectively, for affecting downstream exon skipping. Especially, the identified sequence in the the right arm was highly conserved in other *Alu* subfamilies and was predicted as new splice acceptor site by *in silico* splicing analysis. Our experiments suggested the intronic antisense-orientation partial *Alu* sequence (just a few dozen nucleotides) triggered a downstream exon skipping.

## PgmNr 3344: An epigenome-wide association study revealed novel epigenetic factors for childhood asthma in East Asians.

### Authors:

D. Yoon <sup>1</sup>; J. Kim <sup>1</sup>; W.S. Chang <sup>1</sup>; E.J. Kim <sup>1</sup>; M.J. Kang <sup>2</sup>; S.J. Hong <sup>2</sup>; J.K. Lee <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Division of allergy and chronic respiratory diseases, Korea National Institute of Health, Korea, Republic of; 2) Department of Pediatrics, University of Ulsan College of Medicine, Seoul, Korea

---

Asthma is a heterogeneous disease arise from complex interaction between genetic and environmental factors. To date, numerous genetic factors responsible for asthma have been discovered. However, identifying epigenetic changes, reflecting variation of environmental factors, of asthma is relatively in the early stage compared to those of genetic factors. Moreover, majority of previously reported epigenome-wide association study (EWAS) for asthma were conducted in populations with European ancestry. In this study, we performed EWAS on childhood asthma comprising 41 asthma cases and 29 controls of East Asians (Korean). Asthma cases were diagnosed by physician and controls were selected if no history of asthma. Genome-wide DNA methylation were profiled by using Illumina Infinium HumanMethylation 450K BeadChip. Raw data were quality controlled and quantile normalized prior to further analysis. A two-sample t-test was used to determine epigenetic difference between cases and controls. As a result, four differentially methylated regions (DMR) discovered at  $P < 1 \times 10^{-4}$  near *CDV3*, *ABCB9*, *RNF144B*, and *FOXC1*. Of note, *RNF144B* is known to be involved in Class I MHC mediated antigen processing and presentation. We further assessed DMRs at  $P < 5 \times 10^{-2}$  and performed functional annotation analysis to understand possible biological role of genes closely located near DMRs. Pathway analysis revealed top associated KEGG pathway 'Natural killer cell mediated cytotoxicity' consisting of seven genes: *MICA*, *VAV3*, *PIK3CB*, *FASLG*, *HLA-C*, *PAK1*, and *HLA-E*. Collectively, this study provides valuable knowledge of epigenetic factors for asthma and its possible role in the pathogenesis of asthma.



## PgmNr 3345: RegulomeDB 2.0.

**Authors:**

B. Hitz<sup>1</sup>; A. Diehl<sup>2</sup>; S. Dong<sup>2</sup>; Y. Luo<sup>1</sup>; E. O'Neill<sup>2</sup>; N. Ouyang<sup>2</sup>; A. Boyle<sup>2,3</sup>; J.M. Cherry<sup>1</sup>

[View Session](#) [Add to Schedule](#)

**Affiliations:**

1) ENCODE Data Coordination Center, Department of Genetics, Stanford Univ, Palo Alto, California; 2) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI

---

The rapid growth of human genome sequencing has identified a large number of genetic variants present in the human genome relevant to human disease and other phenotypes. The majority of sequence variation is found within the noncoding regions of the genome where it is presumed to affect gene regulatory elements. We have developed a unique web-accessible database, RegulomeDB, which provides integrated knowledge of existing information describing regulatory elements that lie within noncoding regions. The RegulomeDB v1.1 (<http://www.regulomedb.org/>) was originally built from ENCODE and other community data, and has proven to be a useful tool for noncoding variant annotation. Various genomic information, including eQTL, transcription factor binding, chromatin modification and DNA methylation have been integrated into a common database and displayed at nucleotide resolution. Researchers are able to compare variants of their interest to the wealth of information in RegulomeDB, and rapidly gain insight into possible mechanism of gene regulation. In order to keep up with emerging genome sequencing projects and provide an up to date regulome annotation, we have integrated RegulomeDB into the ENCODE portal, as RegulomeDB v2.0 (<https://regulome-master.demo.encodedcc.org/regulome-search/>). We are also investigating improved annotation with new data types and integration of novel, statistical scoring methods. Overall, we expect this unique resource to have wide impact in the biomedical community.

## PgmNr 3346: Heritability of common complex traits is enriched at topologically associating domain boundary regions.

### Authors:

E. McArthur <sup>1</sup>; J.A. Capra <sup>1,2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Department of Biological Sciences, Vanderbilt University, Nashville, TN

---

**Background:** The 3D conformation of the genome plays an integral role in regulating gene expression. However, the effects of the interaction between genetic variation and 3D architecture on phenotypes are incompletely understood. On a sub-chromosomal scale, chromatin folds into topologically associating domains (TADs), regions that commonly self-interact, but rarely contact outside chromatin. TADs, and the boundaries between them, modulate gene regulation by restricting interactions of regulatory sequences, like enhancers, to their target genes. Structural variation that disrupts the insulating effects of TAD boundaries has been implicated in rare developmental phenotypes. Smaller-scale variants, like SNPs, also have the potential to disrupt TAD boundaries, but the overall contribution of such variants to phenotypes is not known.

**Methods:** We hypothesized that common genetic variants modulate TAD boundaries' effects in ways that are relevant to complex disease. To investigate the relationship between genetic variants in TADs and boundaries with disease, we performed partitioned heritability analysis. We integrated TADs identified from Hi-C experiments in 37 different cellular contexts and, using stratified LD Score Regression, estimated patterns of heritability for 28 complex phenotypes.

**Results:** We show that heritability of many traits is enriched at TAD boundaries, while heritability is not enriched nor depleted within TADs themselves. Although TAD boundaries are generally conserved among cell types, some are observed in almost all cellular contexts, while others are unique to a single cell type. We find that boundaries seen in multiple cell contexts have higher heritability enrichment than less conserved boundaries. This suggests that common SNPs in conserved boundaries are enriched for functional effects. Indeed, boundaries present across many cell types also have higher sequence conservation. We also observe different heritability enrichment patterns among uniquely cell-specific boundaries.

**Conclusions:** In summary, our work demonstrates enrichment for effects of common variants in TAD boundaries on complex phenotypes. Our results suggest that these variants likely modulate the ability of boundaries to insulate regulatory elements; however, further molecular investigation of their mechanistic effects is needed. Ultimately, our results inform non-coding variant interpretation by quantifying the functional contributions of variants in different 3D contexts.

## PgmNr 3347: GADB: Large-scale, curated functional genomics annotation database.

### Authors:

P.P. Kuksa; P. Gangadharan; C.-Y. Lee; Y.-F. Chou; E. Greenfest-Allen; H.-J. Lin; Z. Katani?; O. Valladares; Y.Y. Leung; L.-S. Wang

[View Session](#) [Add to Schedule](#)

**Affiliation:** Penn Neurodegeneration Genomics Center, Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

---

Large-scale, systematic annotation of genetic variants and genomic regions (SVs, called peaks, etc) generated by different sequencing experiments becomes increasingly important and is a complex part of sequencing-based genomic/genetic analyses. Functional genomic data necessary for systems biology analyses, causal gene analyses, interpretation of GWAS signals are currently distributed across individual data collections such as ENCODE, GTEx, FANTOM5, with no single resource providing centralized, unified annotation information allowing seamless integration, retrieval, comparison of functional datasets across experimental assay types, data types, file formats, etc.

To address this, we built Genomic Annotation Database (GADB) and collected, consistently cataloged and harmonized all genome-wide experiments (>60,000 experiments) from 12 consortia such as ENCODE, FANTOM5, ROADMAP, GTEx, etc. GADB contains annotated collection of >22 billion genomic intervals in >67,000 genome-wide annotation files across >1,100 tissues and cell types annotated using OBO Foundry ontologies which provide a framework for grouping the functional genomics tracks into broader cell and tissue categories.

GADB provides integrated, harmonized, searchable and extensible human functional genomic annotation data collection for genetic and genomic analyses measured by 17 different kinds of experimental assays and >30 data types including xQTL, Hi-C interaction data, epigenetic marks, regulatory elements (enhancers, promoters). All GADB annotation tracks are available in both GRCh37/hg19 and GRCh38/hg38 reference genomes.

All these genomic annotation tracks were systematically annotated with assay type, cell type, tissue category, sample, and other information (21 features) and indexed using GIGGLE (<https://github.com/ryanlayer/giggle>).

On average, each genomic position is covered by 3,000 genomic tracks (>9.5 Trillion bps covered). Following the FAIR principles, we built Spark-based API for querying GADB database. Researchers can use GADB to a) quickly identify relevant experiments and data types in specific tissue contexts for downstream analyses; b) quickly search and retrieve all genomic intervals within the genomic region of interest (>3Billion genomic intervals/sec).

GADB (<http://gadp.niagads.org/>) provides integrated, harmonized, extensible functional genomic data repository allowing for seamless retrieval, comparison across various data sources, tissue/cell types, data types.

## PgmNr 3348: The relationship between telomere length and genome-wide DNA methylation in a Bangladeshi population.

### Authors:

L.I. Tamayo<sup>1</sup>; K. Demanelis<sup>1</sup>; D. Delgado<sup>1</sup>; F. Jasmine<sup>1</sup>; M.G. Kibriya<sup>1</sup>; J. Shinkle<sup>1</sup>; H. Ahsan<sup>1,2,3,4</sup>; B.L. Pierce<sup>1,2,3</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Public Health Sciences, University of Chicago, Chicago, IL; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Comprehensive Cancer Center, University of Chicago, Chicago, IL; 4) Department of Medicine, University of Chicago, Chicago, IL

Telomeres are DNA-protein structures that protect chromosomal ends. Telomere length (TL) is a proposed biomarker of biologic aging and of risk for various chronic diseases. DNA methylation (DNAm) also changes with age, and aging is generally associated with increased DNAm at CpG islands and decreased DNAm at non-CpG islands. However, the relationship between TL and genome-wide DNAm has not been extensively studied. In this study, we characterized the association of TL with genome-wide DNAm features and epigenetic aging measures in two cohorts of Bangladeshi adults. DNAm in whole blood was measured at 771,192 CpGs using the EPIC array in the Health Effects of Arsenic Longitudinal Study (HEALS) and 390,810 CpGs using the 450k array in the Bangladesh Vitamin E and Selenium Trial (BEST). After correcting for batch effects, we assessed the association between TL and CpG DNAm using linear models adjusted for age, sex, BMI, smoking status, and cell type composition. We meta-analyzed results from HEALS (n=364) and BEST (n=388) using METAL. Using the DNA Methylation Age Calculator (<https://dnamage.genetics.ucla.edu/>), we computed DNAm age, intrinsic epigenetic age acceleration (EAA) (adjusted for cell type composition), and extrinsic EAA (dependent on changes in cell type composition); and assessed their association with TL.

We identified 100 CpGs that were associated with TL at  $p < 10^{-4}$ . Compared to CpGs not associated with TL, these TL-associated CpGs were enriched in islands and depleted in non-CpG islands ( $p = 0.01$ ). The locational distribution of age-associated CpG differed by direction of association with TL ( $p = 2.9 \times 10^{-5}$ ): negatively associated CpGs (n=75) tended to be located in islands (49%) and shores (32%) while positively associated CpGs (n=25) tended to reside in non-CpG islands (60%). Among HEALS participants, TL was inversely correlated with DNAm age ( $r = -0.21$ ,  $p = 5.2 \times 10^{-5}$ ), consistent with the correlation between TL and age. TL was not associated with intrinsic EAA ( $p = 0.22$ ) but was inversely associated with extrinsic EAA ( $p = 0.02$ ), suggesting longer TL decrease epigenetic age acceleration associated cell type composition.

This analysis revealed that TL was weakly associated with genome-wide DNAm, and these associations with TL were in the opposite direction of age-associated DNAm at CpG islands and non-CpG islands. Additionally, longer TL may be associated with less EAA, suggesting consistency between these biomarkers of aging.

# PgmNr 3349: Landscape of inter-omics interaction of genetic variation, DNA methylation, and gene expression in CD4<sup>+</sup>T lymphocytes, monocytes, and neutrophils.

## Authors:

S. Komaki <sup>1</sup>; Y. Shiwa <sup>1,2</sup>; R. Furukawa <sup>1,3</sup>; T. Hachiya <sup>1</sup>; H. Ohmomo <sup>1</sup>; Y. Sutoh <sup>1</sup>; R. Otomo <sup>1</sup>; M. Sasaki <sup>1</sup>; A. Shimizu <sup>1</sup>

[View Session](#) [Add to Schedule](#)

## Affiliations:

1) Iwate Tohoku Medical Megabank Organization, Yahaba, Iwate, Japan; 2) Department of Molecular Microbiology, Tokyo University of Agriculture; 3) Department of Humanities and Social Sciences, Keio University

## Background

Omics, such as genomic variation, DNA methylation (DNAm), and gene expression, are known to interact with each other. eQTL and mQTL (expression-/methylation-quantitative trait locus) are genomic variations that affect gene expression and DNAm, respectively, and eQTM (expression-quantitative trait methylation) is DNAm that affects gene expression. A number of eQTLs, mQTLs, and eQTMs have been reported and information regarding them is available on some databases, which are frequently employed for annotation of signals found through (epi)genome-wide association studies (GWAS/EWAS), with advantages of multi-omics knowledge. Specifically, as profiles of DNAm and gene expression and EWAS signals vary across cell types, inter-omics interactions can also vary across cell types. To address this, we have been working on multi-omics analyses using sorted blood cells with an emphasis on cell type-specificity, and herein, we present results of our multi-omics analyses.

## Methods

CD4<sup>+</sup>T lymphocytes (CD4T) (n = 99), monocytes (n = 99), and neutrophils (n = 94) isolated through fluorescence-activated cell sorting with >95% purity were used in this study. Genotypic data for ~9 million (M) variation sites, DNAm profiles at ~24M CpGs, and gene expression data for ~20 thousand (k) genes were used. To minimize the effect of outliers, we applied filtering strategies on each omics dataset; then, we conducted cis-eQTL, cis-mQTL, and cis-eQTM analyses (cis-proximity: 500kbp). We employed the Bonferroni correction to adjust p-values, and an adjusted p-value of 0.05 was considered statistically significant threshold.

## Results

We analyzed 22M/19M/26M variation-gene pairs, 3 billion/1 billion/3 billion variation-CpG pairs, and 24M/15M/17M CpG-gene pairs, from which we detected 50k/87k/37k eQTLs, 8M/1M/11M mQTLs, and 5k/470/2k eQTMs, respectively (CD4T/monocytes/neutrophils). Among them, 76% of eQTLs, 65% of mQTLs, and 95% of eQTMs were cell type-specific and 3% of eQTLs, 5% of mQTLs, and 0.5% of eQTMs were common across all three cell types studied.

## Discussion

Interactions between omics observed herein differed substantially across cell types suggesting that

multi-omics analyses based on the mixture of heterogeneous cells could lead to important cell type-specific interactions being overlooked. Using the data from our multi-omics interaction analyses, we would be able to infer the functions and consequences of GWAS and EWAS signals more accurately.

## PgmNr 3350: Activity by contact model identifies targets of regulatory variants for smoking and drinking use traits.

**Authors:**

J. Hughey; D. Liu; GWAS and Sequencing Consortia of Alcohol and Nicotine Use (GSCAN)

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Public Health Sciences, Penn State College of Medicine, Hershey, PA

---

A recent genome-wide association study for smoking and alcohol use with 1.2 million individuals have identified 406 associated loci. Many of these variants implicated in GWAS are in non-coding regions, which are difficult to interpret. With the advancements in genomic data generating technologies, integrating analysis of multi-omic data types have become an increasingly popular method to improve our understanding of mechanisms behind genotype-trait association. Here we use the recent activity by contact model (ABC) to pinpoint the target gene of active enhancers that potentially harbor GWAS associated regulatory variants. The basis of this model is to identify enhancers in a given cell type based off chromatin accessibility and histone modifications. This is the activity proxy which will be coupled with Hi-C contact data to find the physical interactions between regulatory elements and gene promoters. We have extended this model to the IMR90 cell line and biologically relevant brain tissues, to further understand gene regulation and its role in drinking and smoking. The preliminary application of the model predicted 110,239 and 77,205 enhancers for the IMR90 and prefrontal cortex, respectively. Furthermore, the ABC predicted enhancers were compared to the ChromHMM software that predicts and characterizes chromatin states. For IMR90 and prefrontal cortex, 76% and 65% of the ABC predicted enhancers have enhancer or promoter chromatin state characterizations in ChromHMM. Furthermore, variants associated with three tobacco and alcohol use phenotypes were studied: cigarettes per day (cpd), smoking initiation (si) and drinks per week (dpw). We calculate the enrichment using the ratio of the fraction of genome-wide significant SNPs in the active enhancer region over the fraction of all SNPs in the active enhancer region. We observed 3.1x, 4.2x and 2.8x enrichment for DPW, CPD and SI associated loci in the predicted pre-frontal cortex active enhancers. Enrichment was not seen in IMR90 cell line suggesting that this enrichment is tissue specific. These variants also have ABC quantities greater than the median and implicate target genes based off contact data. In conclusion, our results represent a multi-omic approach to dissect the potential regulatory mechanisms behind smoking and drinking related traits. This method can lead to more targeted experimentation of regulatory variants and their prioritized target genes.

## PgmNr 3351: Characterizing the regulatory architecture of essential genes in the human genome.

### Authors:

S. Ramdas <sup>1</sup>; M. Trizzino <sup>2</sup>; C.D. Brown <sup>1</sup>; M. Bucan <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics, University of Pennsylvania, Philadelphia PA; 2) Gene Expression and Regulation Program, Wistar Institute, Philadelphia PA

---

The search for essential genes (EG) is the quest to define a minimal set of genes necessary for the survival of a cell or an organism. These genes are known to be more conserved across species, are intolerant to mutations in human populations, are expressed in early development, are enriched for disease genes, and are broadly expressed across tissues. However, the regulatory architecture of this class of genes is not well-understood. In this study, we characterize the regulatory features of genes essential for mouse viability (developmentally lethal, or DL genes), genes essential for viability of human cell lines (CL), and compare these to two control sets of non-essential genes. Specifically, we use publicly available expression and epigenomic datasets to ask if the evolutionary and functional constraint on EGs (DL and CL genes) is also reflected in regulatory constraint. We also try to understand biological differences between cell-line lethality and organismal lethality.

First, we see that DL genes, despite their high levels of expression, are depleted for eQTLs across 49 human tissues. They also have lower variation in gene expression, indicating selection on this set of genes to maintain stability in expression levels. DL genes further show less tissue sharing of eQTLs than CL genes and NEGs. A larger fraction of eQTLs for DLs is found in enhancer and TSS elements compared to other gene classes; this enrichment disappears when we account for the increased enhancer footprint for DL genes. However, we do not see a difference in effect sizes between eQTLs of different gene classes. These results add to the body of knowledge highlighting the important role of enhancers in regulating developmental genes.

Second, using publicly available epigenomic datasets from multiple human tissues, we show that DL genes are enriched for active regulatory elements, and depleted in repressive elements of the genome. In particular, DL genes are most enriched in enhancers, and show significantly less enrichment at TSS elements compared to non-essential genes. In contrast, CL genes show highest enrichment at TSS elements compared to enhancers. CL and DL genes show greater enhancer sharing across tissues, which could explain their breadth of expression.

Through our efforts, we better characterize the regulatory architecture of a set of genes with an outsized role in human disease; these will aid efforts in understanding disease architectures and variant prioritization.



## PgmNr 3352: Contrasting the genetic architecture underlying gene expression levels in European Americans and African Americans: An eQTL mapping study in GENOA.

### Authors:

J. Smith <sup>1</sup>; L. Shang <sup>2</sup>; W. Zhao <sup>1</sup>; S. Turner <sup>3</sup>; T. Mosley <sup>4</sup>; S. Kardia <sup>1</sup>; X. Zhou <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI; 2) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI; 3) Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN; 4) Memory Impairment and Neurodegenerative Dementia (MIND) Center, University of Mississippi Medical Center, Jackson, MS

---

Identifying genetic variants that are associated with gene expression levels, an analysis commonly referred to as expression quantitative trait locus (eQTL) mapping, is an important first step towards understanding the genetic architecture underlying gene expression variation. Most existing eQTL mapping studies have thus far focused on individuals of European ancestry and are noticeably underrepresented in other populations, with a particular absence of large studies in populations with African ancestry. Lack of large-scale, well-powered eQTL mapping studies in populations with African ancestry can both impede the dissemination of eQTL mapping results for benefiting individuals with African ancestry and hinder the comparable analysis for understanding how gene regulations are shaped through evolution. We fill this critical knowledge gap by performing a large-scale in-depth eQTL mapping study on 1,205 African Americans (AA) and 801 European Americans (EA) from the Genetic Epidemiology Network of Arteriopathy (GENOA) study. We identified a total of 422,586 eQTLs in AA and 458,204 eQTLs in EA after controlling for age, gender, population structure and familial relatedness. Among them, only a small percentage of eQTLs (128,915) overlapped between the two populations, highlighting the importance of eQTL mapping in populations with African ancestry. The identified eQTLs from both populations are enriched near transcription start sites and the overlapping eQTLs between populations have consistent effect directions. eQTL harboring genes (eGenes) tend to have higher SNP heritability as well as cis- and trans-SNP heritability compared to non-eGenes. Importantly, eGenes that are common in the two populations tend to be more conserved than eGenes that are unique in one population, which are more conserved than non-eGenes. In addition, through conditional analysis, we also found that eGenes in AA tend to harbor more independent eQTLs than the eGenes in EA, suggesting potentially diverse genetic architecture underlying gene expression variation between the two populations. Overall, our results represent an important step towards revealing the genetic architecture difference underlying gene expression variation between African Americans and European Americans.

## PgmNr 3353: Jointly modeling allele-specific and total expression to identify cis-acting expression outliers in humans.

### Authors:

B. van de Geijn<sup>1</sup>; G. McVicker<sup>2</sup>; J. Blischak<sup>3</sup>; Y. Gilad<sup>3</sup>; A. Price<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts.; 2) Laboratory of Generics, SALK Institute, San Diego, CA; 3) Human Genetics, University of Chicago, Chicago, IL

---

Non-coding genetic variants can lead to large changes in gene expression and may underly instances of genetic disease. Indeed, a recent study found that individuals with outlier values of total expression ( $|z\text{-score}| > 2$ ) for a given gene harbor an excess of rare variants nearby (Li et al. Nature 2017); however, power to detect expression outliers using total expression is limited and it is challenging to determine whether a particular expression outlier is driven by a cis-acting genetic effect.

We developed a rigorous statistical method to identify rare variant driven outliers by jointly modeling total and allele-specific expression, extending a previous method for identifying quantitative trait loci (WASP: van de Geijn\*, McVicker\* et al. 2015 Nat Methods). By including both total read depth and allele-specific information, we boost power to detect outliers and are able to distinguish between different types of outliers based on model likelihood. The first type, which we are most interested in, is cis-acting effects, where an individual has outlier total expression value and allelic imbalance. The second type is trans-acting effects, where only total expression shows an effect, which could be driven by trans-acting variants or non-genetic factors. Finally, we also identify monoallelic or buffered effects where there is allelic imbalance but no change in total expression. The method directly integrates both total and allele specific RNA-seq count data, taking into account technical effects, overdispersion, and miscalled genotypes (which can lead to false positives).

We applied our method to GTEx RNA-sequencing data from subcutaneous adipose tissue in 350 individuals. We identified 80% more significant expression outliers at 5% FDR (average of 36 vs 20 per individual) with our method, compared to the z-score test for total expression. We then tested whether outliers from each test replicated in visceral adipose tissue from the same individuals using the corresponding test and a nominal  $p\text{-value} < 0.05$ . We replicated 67% of outliers identified by our method, compared to only 23% of outliers identified using the z-score test for total expression. These results indicate that our method provides a powerful approach to detect replicable outliers, even in individual tissues.

## **PgmNr 3354: Gene circuitry mapping assigns validated functional annotation to GWAS non-coding hits: from primary human hepatocytes to culprit disease-associated genes.**

### **Authors:**

Y. Liu; Y. Guo; E. Cohick; O. Chung; J. Huang; A. Almutairy; P. Fan; L. An; J. Weinstein; A. Sigova; D. Bumcrot; A. Schachter; I. Grossman

[View Session](#) [Add to Schedule](#)

**Affiliation:** Camp4 Therapeutics, Cambridge, Massachusetts.

---

Nearly 90% of GWAS hits are in non-coding genomic regions with 57% of these lying in open chromatin regions, suggesting roles in transcriptional regulation. For these variants, the specific functional relevance and impacted genes remain elusive, with most studies currently mapping non-coding variants (NCVs) to genes based on genomic proximity. However, gene transcription is often regulated by enhancer-promoter interactions spanning large distances. We have built a Gene Circuitry Map™ of primary human cells, including hepatocytes, which defines functional cis-regulatory elements, identifies the transcription factors (TFs) that interact with these elements, and elucidates functional long-and short-range enhancer-promoter interactions. Combined with transcriptome-wide analysis of the cellular response to signaling perturbagens, the map empowers the prediction of novel connections between signaling pathways and the transcriptional control of any gene of interest. We used the map to annotate the functionality of NCVs identified in GWAS. To achieve this, 46989 of the 83121 (56%) unique variants in the GWAS catalog (as of 5Apr2019) were mapped into hepatocyte Gene Circuitry regulatory regions, including 7950 (17%) in enhancers or promoters, and 771 (5%) in CTCF chromatin insulator regions (Insulated Neighborhood [IN] boundaries). Also, 41016 (87%) unique NCVs were mapped within INs containing known genes. Next, a focused list of quantifiable liver-specific GWAS traits was compiled encompassing liver function tests (AST, ALT, GGT, bilirubin, prothrombin time), iron regulation (ferritin, iron, transferrin, total iron binding capacity, transferrin saturation), and lipid levels (cholesterol, HDL, LDL, VLDL, triglycerides). Of the 46989 functionally annotated NCVs, 2681 (5.7%) are associated with quantifiable liver traits, out of which 102 (4%) are in known TF motifs, map within INs containing known genes, and map to open chromatin regions defined by ATAC-seq (liver function =14, iron=4, lipid = 84 NCVs). Integration of concordant disruption in motif, transcription binding event and eQTL of a gene in the relevant IN is then poised to deconvolute the pathologic NCV within an LD block. Further experiments are planned to confirm the causal relationship with the downstream gene affected. Our map assigns reliable novel functions to NCVs across the genome, facilitating rapid discovery of disease associated genes and novel mechanisms for their therapeutic management.

## **PgmNr 3355: Better IDEAS: Using epigenetic “fingerprints” to systematically identify disease causal cell types for thousands of traits in the GWAS Catalog and UK Biobank.**

### **Authors:**

C. Guo <sup>1</sup>; K.B. Sieber <sup>1</sup>; M.R. Nelson <sup>1</sup>; Y. Zhang <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### **Affiliations:**

1) Human Genetics, GSK, Collegeville, PA.; 2) Statistics, Penn State University, University Park, PA

---

Identifying disease causal cell types underlying a genetic association is a crucial step towards understanding disease biology and executing effective target validation experiments. Recent advancements in epigenetic profiling has resulted in a wealth of data that can be leveraged to gain a deep understanding of the genetics of cell type-specific gene regulation. Here, we predicted causal cell types for 1,918 traits in the UK Biobank and 1,098 traits in the NHGRI-EBI GWAS Catalog by matching genetic association signals with cell-type specific gene regulatory elements. We applied Integrative and Discriminative Epigenome Annotation System (IDEAS), a 2D genome segmentation method to annotate regions of the genome with gene regulatory potential in 150 cell types. To estimate the proportion of genetic signal that can be attributed to each cell type for a given trait at locus-level resolution, we developed a method called IDEASpredict that integrates GWAS posterior probabilities from statistical fine-mapping with regulatory elements defined by IDEAS. In contrast to existing enrichment analysis methods, IDEASpredict accounts for the correlation of regulatory events across cell types and estimate the cell types in which causal variants are most likely to be functional.

For traits with more than 10 genome-wide significant loci, we found 20% to have a single cell type explain > 50% of the associated loci, with most of those signals residing in immune related cells. In addition to observing novel and expected enrichments for specific disease-cell type pairs, we also were able to disentangle disease biology for related phenotypes. To illustrate, Crohn's disease had 2x higher enrichment in immune derived cells than ulcerative colitis, while ulcerative colitis had 4x higher enrichment in epithelium derived cells. A comparison between T-cell subtypes for the two traits revealed that while memory T-cells explained 90% of the T-cell related signal for ulcerative colitis, they only explained 18% of the T-cell signal for Crohn's. In summary, we present a new approach to identify cell type-specific enrichments and provide an encyclopedia of cell type annotations at the locus level for 3,016 traits. We expect this approach to substantially improve our ability to take advantage of GWAS to understand complex disease pathobiology and improve the design of in vitro studies to validate prospective therapeutic targets.

## PgmNr 3356: Expression regulatory quantitative trait locus with various functions.

**Authors:**

J. Han; C. Lee

[View Session](#) [Add to Schedule](#)**Affiliation:** Department of Bioinformatics and Life Science, Soongsil University, Seoul, Korea

---

Expression regulatory quantitative trait locus (erQTL) is of a great concern to understand underlying mechanisms of genetic factors for complex diseases. Such a concern has been intensively extended to erQTLs for various kinds of molecular products of genes. The current genome-wide erQTL analysis with transcriptome-wide data revealed cis-acting erQTLs for HLA-DQA1 and MRPL43 by employing a mixed model ( $P < 1 \times 10^{-5}$ ), showing associations with all of the mRNA expression, ribosome occupancy, and protein abundance. Every erQTL was covered by a wide range in strong linkage. In particular, an erQTL was located in a strong linkage at the gene encoding MRPL43. While its minor haplotype increases mRNA expression and ribosome occupancy, it reduces protein level. Further analysis revealed that erQTL was attributed to independent functions of three sequence variants in the locus. One variant may regulate binding affinity of transcription factors. Another variant may change splice site that affect to isoforms of MRPL43. A variant in the isoform with long 3' UTR may strengthen microRNA binding affinity because of its 7-mer seed sequence, which might contribute to the flipped effect from transcription to translation.

## PgmNr 3357: MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments.

### Authors:

M. Hu <sup>1</sup>; I. Juric <sup>1</sup>; M. Yu <sup>2</sup>; A. Abnoui <sup>1</sup>; R. Raviram <sup>2</sup>; R. Fang <sup>2</sup>; Y. Zhao <sup>2</sup>; Y. Zhang <sup>2</sup>; Y. Qiu <sup>2</sup>; Y. Yang <sup>3</sup>; Y. Li <sup>3</sup>; B. Ren <sup>2</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH.; 2) University of California San Diego; 3) University of North Carolina Chapel Hill

---

Hi-C and chromatin immunoprecipitation (ChIP) have been combined to identify long-range chromatin interactions genome-wide at reduced cost and enhanced resolution, but extracting information from the resulting datasets has been challenging. Here we describe a computational method, MAPS, **M**odel-based **A**nalysis of **PLAC-seq** and **HiChIP**, to process the data from such experiments and identify long-range chromatin interactions. MAPS adopts a zero-truncated Poisson regression framework to explicitly remove systematic biases in the PLAC-seq and HiChIP datasets, and then uses the normalized chromatin contact frequencies to identify significant chromatin interactions anchored at genomic regions bound by the protein of interest. MAPS shows superior performance over existing software tools in the analysis of chromatin interactions from multiple PLAC-seq and HiChIP datasets centered on different transcriptional factors and histone marks. MAPS is freely available at <https://github.com/ijuric/MAPS>.

## PgmNr 3358: Genome-wide DNA methylation and gene expression patterns reflect genetic ancestry and environmental differences in geographically distinct populations across Island Southeast Asia.

### Authors:

H.M. Natri<sup>1,2</sup>; K.S. Bobowik<sup>3,4,5</sup>; P. Kusuma<sup>6,7</sup>; G.S. Jacobs<sup>7</sup>; C.C. Darusallam<sup>6</sup>; J.S. Lansing<sup>7,8,9,10</sup>; H. Sudoyo<sup>6,11,12</sup>; N.E. Banovich<sup>2</sup>; M.P. Cox<sup>13</sup>; I.G. Romero<sup>3,4,5</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe 85281, AZ, USA; 2) The Translational Genomics Research Institute, Phoenix 85004, AZ, USA; 3) Melbourne Integrative Genomics, University of Melbourne, Parkville 3010, Australia; 4) School of BioSciences, University of Melbourne, Parkville 3010, Australia; 5) Centre for Stem Cell Systems, University of Melbourne, Parkville 3010, Australia; 6) Genome Diversity and Diseases Laboratory, Eijkman Institute for Molecular Biology, Jakarta 10430, Indonesia; 7) Complexity Institute, Nanyang Technological University, Singapore 637723, Singapore; 8) Asian School of the Environment, Nanyang Technological University, Singapore 637723, Singapore; 9) Santa Fe Institute, Santa Fe, NM 87501, USA; 10) Stockholm Resilience Center, Kräftriket, Stockholm 10405, Sweden; 11) Department of Medical Biology, Faculty of Medicine, University of Indonesia, Jakarta 10430, Indonesia; 12) Sydney Medical School, University of Sydney, Sydney, NSW 2006, Australia; 13) Statistics and Bioinformatics Group, School of Fundamental Sciences, Massey University, Palmerston North 4410, New Zealand

The Indonesian populations exhibit a striking level of diversity, regional patterns of admixture, and varying degrees of introgression from Neanderthals and Denisovans. Recent studies have characterized the patterns of genetic variation and admixture across the Indonesian islands. Here, we provide a benchmark dataset of molecular phenotypes in Indonesia. We have integrated CpG methylation and RNAseq data from three island populations that capture the major genomic and geographical axes of diversity across the region: Mentawai is representative of the dominant Asian ancestry and Korowai of the regional Papuan ancestry, while Sumba is a near equal mixture of these two ancestries. Moreover, Korowai contains up to 5% of introgressed DNA from archaic Denisovans. Using this approach, we have investigated the between- and within-island variation of two gene regulatory phenotypes: DNA methylation and gene expression. DNA methylation is an epigenetic process that has an essential role in regulating transcription and has been implicated in affecting various complex traits and diseases. Both DNA methylation and gene expression are influenced by environmental and genetic factors, yet little is known about gene regulatory variation at the population level in Indonesia, and how this variation may contribute to important phenotypes, such as immunity.

We report substantial population differences in DNA methylation and gene expression: we identified differentially methylated CpG probes (DMPs) and differentially expressed genes (DEGs) between these genomically distinct populations. We show that the observed variation in gene expression is partially driven by DNA methylation, with the expression levels of 12.6% of the investigated genes

strongly correlating with nearby CpG methylation. These transcriptionally predictive methylated genes partially overlap the DMPs and DEGs between islands. The DMPs, DEGs, and transcriptionally predictive genes were enriched in pathways involved in immunity. Furthermore, we report within-island variation in DNA methylation and gene expression, likely driven by the local environment. Together, these results indicate a causal relationship between genetic variation, DNA methylation, transcription, and immunity, shaped by the environment. These findings have implications to genomic medicine in underrepresented Indonesia, and help us better understand the roles of genomic and environmental effects in shaping molecular and complex phenotypes.



## PgmNr 3359: Novel detection of mitochondrial RNA cleavage sites.

**Authors:**

G. Carbajosa; A. Hodgkinson

[View Session](#) [Add to Schedule](#)

**Affiliation:** Department of Medical and Molecular Genetics, School of Basic and Medical Biosciences, King's College London, London, United Kingdom

---

Human mitochondrial DNA is transcribed as a polycistronic RNA containing multiple genes. To obtain the final gene products, extensive post-transcriptional processing is required, including the cleavage of RNA through the targeting of secondary degree structures by nuclear encoded enzymes. Recent studies have shed some light on the molecular mechanisms underlying these processes, but they are far from completely described, and little is known about whether RNA cleavage sites vary in location and frequency across 'healthy' individuals. We have developed a computational method by which we can infer mitochondrial RNA cleavage sites by assessing the structure of RNA read stacking in standard RNA sequencing libraries, and we apply this to data from over 1000 individuals to infer variation in these processes on the population scale. Using this approach we identify known RNA cleavage sites at gene boundaries, but also a series of events at non-canonical sites that may be driven by post-transcriptional modifications or lead to alternative transcripts that modulate gene expression. By comparing rates of cleavage across individuals with genetic data, we identify common nuclear genetic variation in known RNA processing genes that modulate these processes across individuals (e.g. *MRPP3*), as well as novel roles for genes in mitochondrial RNA cleavage, building a more complete picture of the processes and pathways that ultimately influence mitochondrial function.

## PgmNr 3360: Human enhancers with complex age architectures are more active than enhancers with simple architectures.

### Authors:

S. Fong; J.A. Capra

[View Session](#) [Add to Schedule](#)

**Affiliation:** Vanderbilt University Medical Center, Human Genetics, 2215 Garland Avenue Nashville, TN 37232-0700

---

Enhancers bind transcription factors (TFs) and regulate gene expression. Variation in enhancer regions can give rise to new gene regulatory functions, but also commonly associates with disease. While evolutionary conservation in protein coding genes is a powerful predictor of variant function, low functional conservation across enhancer landscapes complicates variant interpretation. Thus, new methods for interpreting non-coding variants are needed.

We hypothesize that the age of an enhancer sequence holds information about its function.

Enhancers can be “simple” and consist of a sequence from a single age or “complex” and consist of sequences from different ages—older “core” sequences and younger “derived” sequences—but how age architecture relates to function is not understood.

We explored the relationship between enhancer age architecture and function by integrating sequence alignments from 46 vertebrate species with FANTOM tissue enhancers, transcription factor binding sites (TFBS) from ChIP-seq, and massively parallel reporter assays (MPRAs). Genome-wide, enhancers are enriched in old genomic regions. Over half of current human enhancers occur in sequences from the ancestor of placental mammals, validating previous reports in liver tissue. Sixty-three percent of enhancer sequences[TC1] have simple architectures, while the rest are complex. Both architectures are enriched for GWAS SNPs, but simple have stronger enrichment than complex. However, in MPRAs, enhancers with simple architectures have lower activity than complex enhancers. Within complex enhancers, derived regions have higher activity than older core regions. This supports a model in which the presence of derived regions promotes strong enhancer activity. Indeed, we identify 17 TFs enriched for binding in derived regions. However, derived regions have lower overall TFBS density than simple enhancers and complex cores. Together these results suggest that specific TFs in derived regions promote enhancer function.

Human enhancers have two types of age architectures: simple and complex. These enhancer architectures differ in activity, specific TF binding, and TFBS density. Thus, considering enhancer age architecture may aid interpretations of human regulatory regions, functional targets, and disease variants.

## PgmNr 3361: Genetics of gene expression studies identify RNA polymerase pausing as a regulator of gene expression.

### Authors:

J. Watts <sup>1</sup>; J. Burdick <sup>2</sup>; V. Cheung <sup>2,3,4</sup>

[View Session](#)[Add to Schedule](#)

### Affiliations:

1) Internal Medicine, University of Michigan, Ann Arbor, Michigan.; 2) Howard Hughes Medical Institute, Chevy Chase, MD, USA; 3) Life Sciences Institute, University of Michigan, Ann Arbor MI, USA; 4) Department of Pediatrics, Division of Neurology, University of Michigan, Ann Arbor, MI, USA

---

Genetic studies of gene expression identify regulators of gene expression. Here we show that RNA polymerase II pausing affects gene expression and can be manipulated to restore aberrant gene expression. DNA-dependent RNA Polymerase synthesizes RNA with punctuated pauses which are coupled to RNA processing and the regulation of transcriptional output. Initially described at the promoters of a few genes, genome-wide studies have found RNA polymerase pausing to be a common feature of transcription in metazoans. To assess the effect of RNA Pol II pausing on human gene expression, in normal cells from different individuals, we determined where RNA polymerase pauses along human genes and the expression levels in the corresponding genes. Using the Precision nuclear Run-On and Sequencing assay (PRO-seq), we isolated nascent RNA with actively transcribing RNA polymerase, then mapped the locations of the polymerases. From the same cells, we quantified gene expression by mRNA sequencing. We found RNA Pol II pauses in a highly regulated manner in promoters and gene bodies. We uncovered a 9-mer motif that signals where RNA polymerase II pauses. Overall, genes with more paused RNA polymerase have significantly ( $P < 10^{-16}$ ) lower gene expression. We then carried out genetics of gene expression analyses and found that RNA polymerase pausing regulates gene expression. Using data from the GTEx consortium, we confirmed that sequence variants in RNA polymerase pause sites are significantly associated with expression levels. In this presentation, we will show how mutagenesis of the sequence motif allows targeted alteration of gene expression and thus provides a promising tool to restore aberrant gene expression in genetic disorders.

## PgmNr 3362: Open chromatin and transcriptome profiling reveal key transcriptional drivers and biological processes in iPSC differentiation to lung epithelium.

### Authors:

A. Paranjpye <sup>1</sup>; J. Kerschner <sup>1</sup>; S.H. Leir <sup>1</sup>; S. Yin <sup>1</sup>; G. Bebek <sup>2</sup>; A. Harris <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OHIO.; 2) Center for Proteomics and Bioinformatics, Case Western Reserve University, Cleveland, OHIO.

---

Induced pluripotent stem cell (iPSCs) are a valuable tool for studying cell identity, development, and the organization of tissues. Robust protocols are available to differentiate iPSCs into lung epithelial cells at air liquid interface (ALI). This allows for the modeling of human bronchial epithelial (HBE) cell maturation, which is relevant to respiratory diseases such as asthma and cystic fibrosis. The capacity for differentiation of individual iPSC lines and their response to biochemical and/or physical cues influences maturation into functional HBE-ALI cells. Characterization of the developmental program of the stem cells may reveal aspects of critical biological processes and lineage determination. Here we use genome-wide methods to examine the gene regulatory networks that contribute to the differentiation pathways from iPSCs to HBE-ALI cultures. In addition, we compare iPSC-derived ALI cultures with primary HBE-ALI cells from donors to identify functional similarities and divergence between them.

Using cells from six key stages of differentiation we performed ATAC-seq to define the open chromatin landscape and in parallel RNA-seq to document the transcriptome. By intersecting these high-throughput datasets we reveal the dynamic nature of the transcriptional network controlling differentiation. Changes in the accessibility of promoters, intronic, and intergenic *cis*-regulatory elements (CREs) across the genome are apparent from the distribution of ATAC-seq peaks during stage progression. Transcription factor (TF) motif predictions within ATAC-peaks show changes in frequency of the binding sites for key chromatin architectural proteins such as CCCTC-binding factor (CTCF) and of cell-type selective TFs. Among these TFs is ets homologous factor (EHF) that is known to have an essential role in the differentiated functions of the airway epithelium. Gene ontology enrichment analysis of differentially expressed genes (DEGs) indicates upregulation of networks involved in immune function and in organ specification during the progression to HBE-ALI cells. Finally, the comparison with primary HBE-ALI cells illustrates key similarities that are important in airway epithelial cell identity together with significant divergence. In summary we developed a detailed profile of the regulatory landscape governing the differentiation of iPSCs to HBE-ALI that will be of value in optimizing this important human airway model system.

## PgmNr 3363: Incomplete X chromosome inactivation is reflected in complex trait genetics architecture.

### Authors:

Y. Fu <sup>1</sup>; M. Pirinen <sup>1,2,3</sup>; T. Tukiainen <sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) 1)Institute for Molecular Medicine Finland, Helsinki Institute of Life Sciences, University of Helsinki, Helsinki, Finland; 2) 2)Department of Public Health, University of Helsinki, Helsinki, Finland; 3) 3)Helsinki Institute for Information Technology and Department of Mathematics and Statistics, University of Helsinki

X chromosome inactivation (XCI) balances X-chromosomal (chrX) gene dosages between males and females. However, XCI is incomplete: up to 25% genes escape from XCI, i.e., are expressed from both chrX in females. Notably, at escape genes the inactivated X often shows lower expression levels than its active counterpart. While escape from XCI is seen as sex differences in gene expression, its effect on the genetic architecture of complex traits is under debate. Thus far, the statistical models applied to GWAS data to study XCI have not fully incorporated the incompleteness of XCI.

We investigated how XCI is reflected in genetic associations between sexes including models reflecting known XCI biology and using height as a model trait for highly polygenic architecture. To this end, we performed sex-stratified and sex-combined GWAS in the UK Biobank (N=407,343) and compared the male-female effect estimates at the identified 74 lead SNPs on chrX and 2,711 lead SNPs on autosomes.

Unlike the autosomes, where the male and female allelic effects are close to equal on average (coefficient for male-female comparison,  $\beta_{aut}=0.977$ ), chrX is enriched in SNPs with larger female effect ( $\beta_x=0.879$ ,  $P=0.003$  for the difference). Such observation deviates from the full XCI assumption hence suggesting a contribution from ~23% loci that escape from XCI.

To further quantify the degree of escape, we applied a mixture model to cluster the chrX lead SNPs into three groups: full XCI (F-XCI) and no XCI (N-XCI) models and additionally an incomplete XCI model (I-XCI) that accounts for the case where expression from the inactivated X is ~70% lower than active X. In this analysis, less than 1% of the lead SNPs followed N-XCI, i.e. full two chrX dosage in females, an estimate in line with prior expression-based evidence. However, a much larger fraction, 36% (SE=11.4%) of the lead SNPs were explained by I-XCI. This suggests a proportionally larger contribution from loci undergoing escape from XCI (36% vs. 25%) to height genetics. Interestingly, the lead SNP with the highest I-XCI probability is rs1736534 near *ITM2A*, a gene proposed as an XCI escapee.

Our results show that escape from XCI is indeed reflected in the GWAS results of height, and to a larger degree than previously reported. This finding became possible as we explicitly included the biologically motivated I-XCI in our statistical framework. We are expanding these investigations to other complex traits and genetic architectures.

## PgmNr 3364: Investigation of DNA elements regulating the escape of *KDM5C/Kdm5c* from X-chromosome inactivation.

### Authors:

S. Peeters<sup>1</sup>; A. Korecki<sup>1,2</sup>; E. Simpson<sup>1,2</sup>; C. Brown<sup>1</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Department of Medical Genetics, University of British Columbia, Vancouver, Canada; 2) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, Canada

X-chromosome inactivation (XCI) epigenetically silences one X chromosome in every cell in female mammals, but approximately 12-27% of human and 3-7% of mouse X-linked genes escape from XCI and continue to be expressed from both the active and inactive X chromosomes. Mouse escape genes tend to be distributed as singletons while human escape genes often cluster in larger domains, suggesting both promoter-centric and regional mechanisms of escape. We have shown that despite known species differences between escape gene number and distribution on the X chromosome, human escape genes *RPS4X* and *CITED1* stably escape from XCI in multiple tissues of a transgenic mouse model across development. The recognition of human elements necessary for escape from XCI in mouse provides a tractable system to assess additional human genes for intrinsic escape ability, and characterize the critical regions responsible for continued expression from an inactive X.

As the mouse escape gene *Kdm5c* (in a bacterial artificial chromosome, BAC) has also been shown to escape from XCI when integrated at the 5' end of the X-linked *Hprt* gene, we targeted a BAC containing the human homologue *KDM5C* to the same site. *KDM5C* is known to escape from XCI at the endogenous location in humans, and has been implicated in sex differences in intellectual disability and cancer. Surprisingly, RT-qPCR and promoter DNA methylation analyses demonstrated *KDM5C* to be subject to XCI in mouse. To address our primary hypotheses for why *KDM5C* failed to escape, we generated female mouse embryonic stem cell (ESC) models. (1) Escape may require full gene transcription, and the *KDM5C* we integrated in mouse had an artificial transcription stop site; therefore, a BAC without the stop in *KDM5C* was integrated into an ESC line to analyze for escape following differentiation and XCI. (2) The escape element(s) might not have been included in the BAC; therefore a region near the endogenous *Kdm5c*, located in the mouse BAC but not the human BAC, was deleted in mouse ESCs. This region does not impact *Kdm5c* expression in ESCs, so will be assessed for effects on escape upon differentiation and XCI. If neither of these tests reveals an escape mechanism, other BACs containing genes in the human *KDM5C* escape region can be integrated into the ESC model and analysed. Identifying the features permitting escape from inactivation informs our understanding of the regulatory mechanisms of *KDM5C*, as well as broader questions of XCI.

## PgmNr 3365: The people of Khyber Pakhtunkhwa: An ethnogenetic profile of dental anthropology and molecular genetics of northern part of Pakistan.

### Authors:

H. Ahmad <sup>1,4</sup>; B.E. Hemphill <sup>2</sup>; M.S. Nadeem <sup>3</sup>; K. Muhammad <sup>4</sup>

[View Session](#) [Add to Schedule](#)

### Affiliations:

1) Center for Omic Sciences, Islamia College Peshawar, Peshawar, KP, Pakistan; 2) Department of Anthropology, University of Alaska, Fairbanks, USA, Fairbanks, Alaska 99775, USA; 3) Department of Biochemistry, King Abdulaziz University, Jeddah, Saudi Arabia; 4) Department of Genetics Hazara University, Mansehra, Pakistan

---

Occupying an area of 101,741 km<sup>2</sup> in northern Pakistan, Khyber Pakhtunkhwa Province (KP) is inhabited by 35.53 million people whose origins have been broadly attributed to ancient Dards and Kushans as well as to recently arrived populations from the west generally referred to as subjects of the Pathan dynasties. A significant sporadic literature is available on the history and culture but the scientific investigation concerning the precise origin, general biology and genetic affinities of the myriad ethnic groups residing in KP have yet to be established. Here we are provide new original information on dental morphology and molecular genetics of some 27 ethnic groups of KP.

Phenotypes were established through the analysis of the dental traits of 15,600 volunteers. Dental casts of the maxillary and mandibular dentitions of every subject selected for an ethnic group, were obtained for the permanent teeth, from at least 75 males and 75 females, which were analyzed for a set of 17 tooth-trait combinations, according to the Arizona State University Dental Morphology System. The dental phenotypes thus established were compared with the available data of 3,185 ancient and living groups scored by the same researcher (BEH). Genotypes were analyzed in a phylogeographic perspective among 4,500 volunteers from 15 ethnic groups using Y-STRs, Y-SNPs and mtDNA markers to reconstruct population history and dynamics into and out of this region. Results obtained for 1,318 HVS haplotypes were assigned to 292 haplogroups out of which 134 (10%) were represented once, 40 (3%) were represented twice and 30 (2.3%) three times. The sequences were aligned and compared with the revised Cambridge reference sequence (rCRS), which showed that the 1,061 DNA samples analyzed for paternal ancestry encompass 10 haplogroups among which haplogroup R1a was the most frequent. Nine other haplogroups (R1b, R2, Q, O3, L, J2, I2a, H, and G2a) were represented with differing frequencies. Principal coordinate analysis plots generated by combining our newly reported data with already published datasets from other world regions showed that the paternal origin of the ethnic groups of KP and adjoining areas can be traced to South Asia, Western Eurasia and Eastern Eurasia, coupled with a very limited contribution from Sub-Saharan Africa.

## PgmNr 3366: A deep learning-based approach for fast and accurate genotype imputation.

**Authors:**

R. Dias; A. Torkamani

[View Session](#) [Add to Schedule](#)

**Affiliation:** Scripps Research Translational Institute, La Jolla, CA.

---

Genotype imputation is the statistical inference of unobserved genotypes by using haplotypes from whole genome sequencing (WGS) data as a reference. This task is critical to increase power of genome-wide association studies, allowing the identification of causal variants and helping in the interpretation of observed associations. Statistical genotype imputation is a time-consuming task that demands a substantial amount of computational power, large WGS reference panels, and uses stochastic algorithms that do not always converge into the same solution after running multiple replicates. Here we propose a deep learning-based approach that uses a pre-trained inference neural network, allowing the fast imputation of genetic variants with lower computational cost. Our proposed algorithm consists of a sparse denoising autoencoder that automatically learns the linkage disequilibrium structure among genetic variants, identifying latent factors and using them for predicting the unobserved genotypes. Our results show that our imputation autoencoder algorithm has superior computational performance when compared to other state-of-the-art imputation tools such as Minimac4, running genotype imputation more than 12 times faster and maintaining high accuracy across different genotype array platforms (F1 score  $\geq 0.94$ ,  $r^2 \geq 0.92$ ).